Rare variants in non-coding regulatory regions of the genome that affect gene expression in systemic lupus erythematosus

Sarah A. Jones1,7*, Stuart Cantsilieris2,7, Huapeng Fan1, Qiang Cheng1, Brendan E. Russ3, Elena J. Tucker4,5, James Harris1, Ira Rudloff6, Marcel Nold6, Melissa Northcott1, Wendy Dankers1, Andrew E. J. Toh3, Stefan J. White2,7 & Eric F. Morand1,7

Personalized medicine approaches are increasingly sought for diseases with a heritable component. Systemic lupus erythematosus (SLE) is the prototypic autoimmune disease resulting from loss of immunologic tolerance, but the genetic basis of SLE remains incompletely understood. Genome wide association studies (GWAS) identify regions associated with disease, based on common single nucleotide polymorphisms (SNPs) within them, but these SNPs may simply be markers in linkage disequilibrium with other, causative mutations. Here we use an hierarchical screening approach for prediction and testing of true functional variants within regions identified in GWAS; this involved bioinformatic identification of putative regulatory elements within close proximity to SLE SNPs, screening those regions for potentially causative mutations by high resolution melt analysis, and functional validation using reporter assays. Using this approach, we screened 15 SLE associated loci in 143 SLE patients, identifying 7 new variants including 5 SNPs and 2 insertions. Reporter assays revealed that the 5 SNPs were functional, altering enhancer activity. One novel variant was linked to the relatively well characterized rs9888739 SNP at the ITGAM locus, and may explain some of the SLE heritability at this site. Our study demonstrates that non-coding regulatory elements can contain private sequence variants affecting gene expression, which may explain part of the heritability of SLE.

Systemic lupus erythematosus (SLE, or lupus) is the archetypal multisystem autoimmune disease. SLE patients are predominantly young women who suffer a marked loss of life expectancy and severe morbidity. The causes of SLE are heterogeneous and poorly defined, and patients are routinely treated with broad-spectrum immunosuppressive therapies associated with a high risk of infection, cardiovascular disease, osteoporosis, bone marrow suppression and infertility. Due to the heterogeneity of the disease and the limited knowledge of causative factors, a number of high profile clinical trials of targeted SLE therapies have yielded negative results. Better identification of the causative factors of SLE would allow the development of acutely needed biomarkers, targeted therapies, and potentially personalized medicine approaches.

Considerable evidence supports a genetic contribution to the development of SLE. Twin studies indicate 25–40% concordance for SLE in monozygotic twins, versus 2% concordance in dizygotic twins. Significant effects of ethnicity on SLE disease severity have also been reported, for example both Indigenous Australians and patients of Asian ethnicity have markedly increased SLE prevalence and severity. Microarrays and high-density single nucleotide polymorphism (SNP) genotyping allow genome-wide association studies (GWAS) to be performed on thousands of SLE DNA samples and such studies have implicated several dozen loci in SLE susceptibility. Currently, GWAS studies are estimated to explain 50% of the heritability in SLE. Although the...
effect of some disease-associated SNPs can be explained by effects on the coding sequence of a gene, >80% of
SLE-associated SNPs are in non-coding DNA. Interestingly, the non-coding regions containing SLE-associated
SNPs show an enrichment in enhancer-associated histone modifications, suggesting their potential importance
in driving gene expression and SLE pathogenesis. Indeed several non-coding variants have been functionally
validated using techniques such as luciferase reporter assays and transcription factor binding analysis. The
studied SNPs modulate transcription factor binding strength and can thereby affect gene transcription of nearby
genes, but also of genes further away via long-range chromatin interactions. However, these studies only covered
a small proportion of all the SLE-associated SNPs in non-coding regions.

Table 1. Primers for amplification of each locus.

| Common SNP ID | Sequence Forward | Sequence Reverse | Gene | Location GRCh37 | Regulome DB Score |
|---------------|------------------|------------------|------|-----------------|------------------|
| rs13277113    | GAGCTTCAGGCAAGATGTC | CACCTCGCAGATTCACTAG | BLK  | chr1:11349106-11349337 | 5 |
| rs2618476     | CAGTTTGTTCTTGTTGA | BLK  | chr1:11352441-11352669 | 1d |
| rs2736335     | GTGCAATCGTTGTTCCG | BLK  | chr1:11341344-11341669 | 5 |
| rs969985      | CAGCAGCAGGCTACTGA | ACAGGCAACACTGATTGCAC | BLK  | chr1:11341211-11341453 | 2b |
| rs12574073    | GCCCTGTGTTGTTGACTACCT | ATGGCTCTGTTGTTGCTCTIA | ETS-1 | chr1:1128319404-1128319593 | 3a |
| rs11185603    | GCTCAACTGGAACCTGGGAAG | GAGCTCTGTTGTTGTTGTG | IRZFI | chr7:5030678-50306937 | 2b |
| rs3823536     | TGTACAGGGAAACCCCTTCCTGTC | CTTGAGTCCACAGAGCATG | IRF5 | chr7:128579542-128579750 | 4 |
| rs752637      | GAAACTGTAGCCCCTACAGGA | CAAAAGTGCCCAGAAAGAAGA | IRF5 | chr7:128579213-128579449 | 1b |
| rs10488631    | CAGGTACACAAAGGCTCTCTC | TAGAGGACACTGTGTTCTG | IRF5/IRP1 | chr7:128594148-128594325 | 1f |
| rs9888739     | CACCCATATCATGCTCCCTAGA | GAAAGAACCCATGAGTAAGC | ITGAM | chr16:31313154-31313407 | 1f |
| rs9888879     | GGTTCATCTTCCTGCTTGAC | GCTGTAACACATGGCAACAA | ITGAM | chr16:31310286-31310508 | 2b |
| rs3130320     | GGGTGAGTCACAGGAGAAAGA | ACAGAGAGCCACACGGTTT | NOTCH4 | chr3:32223144-32223381 | 3a |
| rs34202539    | CAGCATGGTGTCAGCACACTCTG | GGTACCCACCCAGAGCTT | TNFIP1 | chr5:150458354-150458578 | 4 |
| rs1150754     | ACCTGACACACCCCTCCTCAC | GCGGTTGAGCTTGCTGATT | TNXB | chr3:32050679-32050949 | 4 |
| rs140489      | GCCAAGTCAGTCGCTCCTTCTC | CAAGAGAGCCAATTTGAGGA | UB2L3 | chr22:1921209-21921364 | 5 |
for further analysis. Sanger sequencing of each candidate locus revealed seven rare variants (defined as either previously undescribed, or only found in a single individual), five single nucleotide variants and two insertions (Table 3).

To determine effects of the variants on gene expression, luciferase assays were performed by transfecting constructs containing each of the rare variants into a cell line, and measuring the amount of luciferase produced. Importantly, cloned sequences did not contain the SNP used to identify the region of interest originally, and thus we are able rule out the possibility of that SNP being responsible for any changes in reporter gene expression. All reference sequences were associated with significantly increased luciferase activation relative to a control transfection, consistent with the cloned DNA fragment having regulatory activity in the cell type. All of the rare variants gave a RegulomeDB score at least as likely to impact binding as the corresponding SNP.

A novel variant at the ITGAM locus. Using our targeted sequencing approach, a rare variant was identified in the ITGAM locus, in linkage disequilibrium with rs9888739, found in GWAS studies to associate with SLE susceptibility\textsuperscript{17,18}. The novel variant inhibited ITGAM expression (Fig. 1A; raw data from luciferase assays shown in Supplementary Table 1), matching the reported impairment of ITGAM expression in association with SLE-associated alleles. ITGAM encodes the CD11b chain of the Mac-1 integrin complex (alphaMbeta2; CD11b/CD18; complement receptor-3) and in the context of SLE, ITGAM expression may be protective through mediation of phagocytosis of iC3b-opsonised apoptotic material, inhibition of T cell activation, restriction of toll-like receptor signaling and inhibition of Th17 responses\textsuperscript{19}.

Previous studies of ITGAM variant rs1143679 had found this allele to be associated with increased risk of renal disease, discoid rash, and immunological manifestations\textsuperscript{20,21}. The patient bearing the novel ITGAM variant

![Figure 1](https://www.nature.com/scientificreports/)

**Figure 1.** Luciferase assays showing effects of novel variants on gene expression. Novel variants were cloned into luciferase reporter constructs and assayed for their effects on luciferase activity as an indicator of their effects on expression of their linked gene. Control = no transfection. Assays were repeated four times and representative results are shown. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.**
we identified had a history of proteinuria and pyuria, arthritis but no discoid rash, based on a 5.5 year period of follow up. The patient was B lymphopenic, had anti-dsDNA antibodies and low complement. Further examination of their immunological profile showed some abnormalities when compared with a larger cohort of SLE patients we have described elsewhere. The patient had no significant difference in levels of circulating interleukin 10 (IL-10) or macrophage migration inhibitory factor (MIF; Fig. 2A,B), clinical associations of which we have previously described. However, the patient bearing the novel C>G variant had substantially higher levels of IL-37 in serum than other SLE patients (1421 pg/mL compared with mean ±/− SD of 277 ±/− 464 pg/mL in a group of 127 SLE patients studied, described previously). IL-37 is an anti-inflammatory cytokine strongly up-regulated in monocytes by TLR ligation and positively correlated with SLE disease activity.

When examining cell populations in the circulation of the patient bearing the novel variant in the ITGAM locus, some differences from the SLE cohort (described in) were observed. While naïve CD8 and CD4 T cell frequencies were unaffected, effector memory CD8 T cells were elevated at 0.0837% of PBMC, outside the upper 95% CI (0.0752%) of the mean (0.0536%) of the SLE cohort (Fig. 2D), and the ratio of total CD4 to CD8 T cells in peripheral blood of the patient bearing the variant was substantially higher than all other SLE patients studied (n = 32) (Fig. 2E). The patient also had a greater proportion of classical monocytes (30.9% compared with mean ±/− SD of 10.54 ±/− 4.37% of the cohort), and plasmacytoid dendritic cells (0.42% compared with mean ±/− SD of 0.286 ±/− 0.160), but no difference in inflammatory monocyte proportions (Fig. 2F).

Novel variants in IRF5 locus. Confirming our approach, we identified another novel rare variant in the same predicted TFBS as rs10488631, which is located 3′ of IRF5. At this locus, a G to A substitution was identified in one patient, and another G to A mutation, 89 nucleotides downstream, was identified in a separate patient. Luciferase assays showed both G to A variants to decrease IRF5 gene expression but this effect was not additive if both variants were present (Fig. 1B and Supplementary Table 1). A role for rs10488631 in SLE has been suggested by several studies, and it has also been implicated in other autoimmune conditions such as systemic sclerosis, Sjogren syndrome and rheumatoid arthritis. The first of these rare variants was in a high information nucleotide within the consensus sequence for NANOG (Fig. 3A), in contrast to rs10488631 where the affected nucleotide is less invariant (Fig. 3B). NANOG is a TF involved in stem cells, and plays a role in regulating pluripotency. There was a second putative TFBS listed in this locus, which is predicted to bind EHF. EHF is part of the ETS TF family, several members of which have previously been implicated in SLE. EHF plays a role in dendritic cell differentiation, and a GWAS has previously associated EHF with SLE in Europeans.
Three other variants were within a predicted TFBS close to, but separate from, the TFBS containing the common SNP. One was located 5′ of the IRF5 gene, near rs3823536 (in linkage disequilibrium with rs4728142, which was previously linked to SLE35). The rare variant disrupts a high information nucleotide in a CACD motif, which can also bind the transcription factor SP136, and luciferase assays showed the novel variant to increase IRF5 expression (Fig. 1C). SP1 is particularly interesting in the context of IRF5 and SLE, as a previous study of the upstream region of IRF5 identified a 5 bp indel polymorphism, creating an additional SP1 binding site, that was associated with SLE37. SP1 binding at other loci has also been implicated in SLE38.

Variant at the ETS1 locus. We identified an additional novel variant located near rs12574073, which is in linkage disequilibrium with rs1128334. These variants are at the 3′ end of ETS1 (as with EHF a member of the ETS family). ETS1 is involved in B cell and Th17 cell differentiation, and an association between rs1128334 and SLE has been reported in Asian SLE cohorts39. Another SNP downstream of ETS1, rs6590330, was also implicated in SLE in an independent study40. The rare variant we identified at this locus was located in a lower information nucleotide in a FOXP3 motif and was found to increase ETS1 expression via luciferase assay (Fig. 1D and Supplementary Table 1). T regulatory cells are characterized by FOXP3 expression, and inhibition of FOXP3 leads to induction of the Th17 pathway, which is known to contribute to SLE pathogenesis25. While regulation of FOXP3 by ETS1 is established40, a reciprocal regulatory relationship is not. However, ETS1 and FOXP3 mRNA levels were both reduced and positively correlated with each other in Treg cells from SLE patients41.

The patient bearing the novel variant in the ETS1 locus was diagnosed at age 12. Over an 8-year observation period, the patient experienced arthritis, haematuria and lymphopenia, but had not (yet) displayed anti-cardiolipin antibodies, discoid lesions, vasculitis or thrombocytopenia, disease manifestations that had previously been associated with various ETS1 alleles in SLE patients42.

Variant at the TNIP1 locus. We found a rare variant near rs10036748, within the TNIP1 gene, which impaired TNIP1 expression according to a luciferase reporter assay (Fig. 1E and Supplementary Table 1). An association for rs10036748 with SLE was made in a Chinese cohort, and TNIP1 is thought to play a role in SLE through the NFKB pathway43. The rare variant was located in a Gfi1/Gfi1b motif. The affected nucleotide was low information in the Gfi1 sequence, but high information in the Gfi1b motif. TNIP1 is a negative regulator of NFκB signaling and polymorphisms in its locus have been associated with a large number of autoimmune diseases. A mouse strain bearing mutant TNIP1, unable to bind ubiquitin, developed a lupus-like phenotype44. Moreover, mice lacking Gfi1 have recently been reported to develop a TLR7-dependent lupus-like phenotype, which the authors showed to involve excess NFκB signaling45.

Discussion

Many studies have shown that a significant proportion of common variants associated with disease are located in genomic regions thought to play a role in regulating gene expression46. Less well studied, however, is the impact of rare variants on gene regulation. In this study we identified seven rare variants, including five SNPs and two indels. Four of the five rare SNPs identified in this study were within predicted TFBS. A previous report of common, non-coding variants in autoimmune conditions such as SLE found that most of the candidate variants were positioned outside the TFBS47. It has been shown in the hemoglobin locus that sequence variants outside TFBS are still capable of disrupting TF binding, presumably through an effect on local chromatin structure48. Additionally, high-resolution analyses of TF binding using ChIP-Nexus has shown that the TF footprint extends further than the binding motif, further evidence that sequence variants outside the TFBS can impact on TF binding49. It is plausible that most variants within the TFBS will have a stronger effect on binding, and may therefore be under negative selection.

This study targeted loci previously implicated in disease susceptibility through GWAS analysis of common variants. We screened 16 small (~200 bp) genomic regions, and identified a number of rare variants that were in
each case at least as likely to have an impact on regulatory potential as the previously associated, common variants as predicted by RegulomeDB. We identified novel variants in the loci of IRF5, ETS1, ITGAM1 and TNIP1, each of which caused alterations in the expression levels of the association genes. Alone, these novel variants are unlikely to be a major contributor to SLE susceptibility at a population level. However, combined with the multitude of other (common and rare) variants in relevant regulatory regions, demonstration that they are functional suggests they will likely play a role in the disease. Identifying which of the millions of variants in every human genome are involved in disease expression is a major challenge in human disease genetics.

Although complete genome sequencing is becoming more affordable, sequencing thousands of samples is still out of the reach of most laboratories, and the vast majority of variants will not play a role in a given disease. Analysing cell-specific epigenetic data allows prioritization of non-coding sequences that may have a role in a specific disease. A study using H3K27Ac (a known marker of active enhancers) enriched loci across a range of cell types found that the most important cell types for SLE appear to be T cells and particularly B cells. The importance of B cells is further underlined by a report that integrated gene expression data of different immune cell types found that the most important cell types for SLE appear to be T cells and particularly B cells. Analysing cell-specific epigenetic data allows prioritization of non-coding sequences that may have a role in a specific disease. A study using H3K27Ac (a known marker of active enhancers) enriched loci across a range of cell types found that the most important cell types for SLE appear to be T cells and particularly B cells. The importance of B cells is further underlined by a report that integrated gene expression data of different immune cell types found that the most important cell types for SLE appear to be T cells and particularly B cells. The importance of B cells is further underlined by a report that integrated gene expression data of different immune cell types found that the most important cell types for SLE appear to be T cells and particularly B cells.

Methods
Ethical approval and informed consent. For experiments involving human samples, all samples were collected using protocols approved by the Monash Health Ethics Committee. All experiments were performed in accordance with relevant guidelines and regulations. Informed consent was obtained from all participants.

Selection of loci. A selection of SNPs identified as SLE susceptibility loci in GWAS studies (referred to in Table 2) were chosen for investigation for novel rare variants based on their being in putative regulatory regions as indicated in the RegulomeDB website (http://regulomedb.org/GWAS/index.html). The selected loci were

| Common SNP ID | Gene Implicated | Variant 1 (Var1) | Variant 2 (Var2) |
|---------------|-----------------|-----------------|-----------------|
| rs9888739     | ITGAM           | C               | G               |
| rs10488631    | IFR5            | Var1            | Var2            |
| rs12574073    | IRF5            | GG              | GA              |
| rs3823536     | IRF5            | C               | A               |
| rs12574073    | ETS1            | T               | T               |
| rs10036748    | TNIP1           | C               | G               |
| rs5754217     | UBE2L3          | Insertion       |
| rs969985      | BLK             | Insertion       |

Table 2. Sequences that were cloned for functional validation.

| Common SNP ID | Location GRCh37 | RegulomeDB score | Associated Gene | Rare variant location | RegulomeDB score |
|---------------|-----------------|------------------|-----------------|-----------------------|------------------|
| rs9888739     | chr16:31313253  | 4                | ITGAM           | chr16:31310306       | 4                |
| rs10488631    | chr7:128594183  | 3a               | IFR5            | chr7:128594188       | 3a               |
| rs3823536     | chr7:12857966   | 2b               | IFR5            | chr7:12857967        | 2b               |
| rs12574073    | chr11:128319478 | 3a               | ETS1            | chr11:128319490      | 2b               |
| rs10036748    | chr5:15045816   | 3a               | TNIP1           | chr5:150458365       | 2b               |
| rs5754217     | chr22:21939675  | UBE2L3           |                 |                       |                  |
| rs969985      | chr8:11341870   | BLK              |                 |                       |                  |

Table 3. List of rare SNPs that were identified.
screened by high resolution melt analysis (below) for additional unidentified SNPs (not overlapping, but in linkage disequilibrium, with the GWAS SNP).

**High resolution melt analysis.** Primers for amplification of each locus, corresponding with Human Genome hg19 notation, are listed in Table 1. High resolution melt (HRM) analysis was performed as described previously\(^4\). In brief, PCR amplifications were performed in 10 \(\mu\)L reaction volumes, consisting of HRM Master Mix (Idaho Technologies, USA), 5 \(\mu\)M each of forward and reverse primer, and 25 ng genomic DNA. PCR products were analysed in a 96 well plate in the LightScanner (Idaho Technologies, USA). The HRM settings for the LightScanner were as follows: start temperature of 70 °C, end temperature at 96 °C, with a hold temperature at 67 °C. HRM curves were normalized using GeneMelt software supplied with the instrument. Aberrant curves were identified by visual analysis and selected PCR products underwent Sanger Sequencing.

**Sanger sequencing.** Sanger Sequencing was undertaken using Big Dye Terminator Chemistry version 3.1 (BDTv3.1) on a 3130xL capillary sequencer supplied by Applied Biosystems. The PCR products were purified using the Exo SAP protocol, in which 5 \(\mu\)L of PCR product is combined with 2 \(\mu\)L of EXOSAPIT enzyme mix from GE Healthcare. The reaction consisted of 1 cycle of 37 °C for 15 minutes, the enzyme was then heat inactivated for 15 minutes at 80 °C. The purified PCR products were then amplified in a sequencing reaction using the BDTv3.1 chemistry. The reaction mix consisted of 4.0 \(\mu\)L of 2.5x ready reaction mix, 2.0 \(\mu\)L of 5x Big Dye Sequencing Buffer, 1.0 \(\mu\)L of Forward or Reverse CRP primer at 3.2 pmol/\(\mu\)L, 2 \(\mu\)L of purified PCR product and 11 \(\mu\)L of DH2O. The cycling conditions were as follows: 1 cycle of 96 °C for 1 minute, 25 cycles of 96 °C for 10 seconds, 50 °C for 5 seconds, 60 °C for 4 minutes. Sequencing products were purified using the Ethanol/EDTA/Sodium Acetate Precipitation protocol in which 2 \(\mu\)L of 125 mM EDTA, 2 \(\mu\)L of 3 M sodium acetate and 50 \(\mu\)L of 100% ethanol was added to each sequencing reaction. The reaction was incubated at room temperature for 15 minutes and samples centrifuged for 2000–3000 g for 30 minutes. The supernatant was removed and 70 \(\mu\)L of 70% ethanol added and centrifuged at 1650 g for 15 minutes at 4 °C. The supernatant was removed and the samples were re-suspended in injection buffer and loaded on the 3130 capillary sequencer. The sequencing data was analysed using the software SeqScanner available from Applied Biosystems.

**Cell culture.** Human B-lymphoblastoid cells (Raji cell line) were cultured in T75 flasks and grown overnight in 10 mL RPMI-1640 medium supplemented with 10% heat inactivated fetal bovine serum (FBS) in a tissue culture incubator humidified with a 5% CO\(_2\) at 37 °C.

**Generation of reporter constructs.** The non-coding DNA variant fragments corresponding to the identified SNPs (Table 2) were constructed into the plasmid pGL4.27 [luc2P/minP/Hygro] (#E845A, Promega, Madison, WI).

**Luciferase assays.** Plasmids constructed to bear the identified variants (3 \(\mu\)g/each) were transfected into 1 million Raji cells. After 24 h of transfection, cells were lysed and luciferase activities, as indicated by relative luciferase units (RLU) were determined using the luciferase assay system (#E1501, Promega, Madison, WI) according to the manufacturer's instructions.

**ELISAs.** ELISAs for MIF, IL-10 and IL-37 were performed as previously described\(^23,24,26\).

**Flow cytometry.** Flow cytometric cell subset analysis in PBMC of healthy controls and patients with SLE is described elsewhere\(^22\).

**Data availability**

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 27 October 2017; Accepted: 9 October 2019;
Published online: 28 October 2019

**References**

1. Bernatsky, S. et al. Mortality in systemic lupus erythematosus. *Arthritis Rheum* **54**(8), 2550–7 (2006).
2. Merrill, J. T. Efficacy and safety of rituximab in moderately-to-severely active systemic lupus erythematosus: the randomized, double-blind, phase II/III systemic lupus erythematosus evaluation of rituximab trial. *Arthritis Rheum* **62**(1), 222–33 (2010).
3. Jourde-Chiche, N., Chiche, L. & Chaussabel, D. Introducing a New Dimension to Molecular Disease Classifications. *Trends in Molecular Medicine* **22**(6), 451–453 (2016).
4. Deapen, D. et al. A revised estimate of twin concordance in systemic lupus erythematosus. *Arthritis Rheum* **35**(3), 311–8 (1992).
5. Vincent, F. B. et al. Focus on systemic lupus erythematosus in indigenous Australians: toward a better understanding of autoimmune diseases. *Intern Med J* **43**, 227–234 (2013).
6. Vaughn, S. E. et al. Genetic susceptibility to lupus: the biological basis of genetic risk found in B cell signaling pathways. *J Leukoc Biol* **92**(3), 577–91 (2012).
7. Jarvinen, T. M. et al. Replication of GWAS-identified systemic lupus erythematosus susceptibility genes affirms B-cell receptor pathway signalling and strengthens the role of IRF5 in disease susceptibility in a Northern European population. *Rheumatology (Oxford)* **51**(1), 87–92 (2012).
8. Chen, L., Morris, D. L. & Vyse, T. J. Genetic advances in systemic lupus erythematosus: an update. *Curr Opin Rheumatol* **29**(5), 423–433 (2017).
9. Guthridge, J. M. et al. Two functional lupus-associated BLK promoter variants control cell-type- and developmental-stage-specific transcription. *Am J Hum Genet* **94**(4), 586–98 (2014).
10. Thynn, H. N. et al. An allele-specific functional SNP associated with two systemic autoimmune diseases modulates IRF5 expression by long-range chromatin loop formation. *J Invest Dermatol* (2019).
52. Chung, S. A. 
49. He, Q., Johnston, J. & Zeitlinger, J. ChIP-nexus enables improved detection of in vivo transcription factor binding footprints.
48. Lower, K. M.
47. Farh, K. K.
44. Ramirez, V. P., Gurevich, I. & Aneskievich, B. J. Emerging roles for TNIP1 in regulating post-receptor signaling.
45. Desnues, B.
43. Adrianto, I.
40. Polansky, J. K.
38. Juang, Y. T.
37. Sigurdsson, S.
34. Armstrong, D. L.
28. Zervou, M. I.
20. Yang, W.
27. Nold, M. F.
26. Foote, A.
23. Godsell, J.
22. Jones, S. A.
21. Kim-Howard, X.
20. Yang, W.
19. de Boer, C. M. et al. DICER1 RNase IIIb domain mutations are infrequent in testicular germ cell tumours. BMC Research Notes 5, 569 (2012).
18. Chung, S. A. et al. Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production. PLoS Genet 7(3), e1001323 (2011).
17. Liu, O. et al. e54419 (2013).
16. Hu, X. et al. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. Am J Hum Genet 92(4), 496–506 (2013).
15. Harismendy, O. et al. 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. Nature 470(7333), 264–8 (2011).
14. Boyle, A. P. et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 22(9), 1790–7 (2012).
13. Myouzen, K. et al. Regulatory polymorphisms in EGR2 are associated with susceptibility to systemic lupus erythematosus. Hum Mol Genet 19(11), 2313–20 (2010).
12. Fagerholm, S. C. et al. The CD11b-integrin (ITGAM) and systemic lupus erythematosus. Lupus 22(7), 657–63 (2013).
11. Yang, W. et al. ITGAM is associated with disease susceptibility and renal nephritis of systemic lupus erythematosus in Hong Kong Chinese and Thai. Hum Mol Genet 18(11), 2063–70 (2009).
10. Kim-Howard, X. et al. ITGAM coding variant (rs1143679) influences the risk of renal disease, discoid rash and immunological manifestations in patients with systemic lupus erythematosus with European ancestry. Ann Rheum Dis 69(7), 1329–32 (2010).
9. Jones, S. A. et al. Glucocorticoid-induced leucine zipper (GILZ) inhibits B cell activation in systemic lupus erythematosus. Ann Rheum Dis 75(4), 739–47 (2015).
8. Nold, M. F. et al. Association of ITGAM polymorphism with systemic lupus erythematosus: a meta-analysis. J Eur Acad Dermatol Venereol 25(3), 271–5 (2011).
7. Nold, M. F. et al. IL-1β is a fundamental inhibitor of innate immunity. Nat Immunol 11(11), 1014–22 (2010).
6. Zervou, M. L. et al. Association of IRF5 polymorphisms with increased risk for systemic lupus erythematosus in population of Crete, a southern-eastern European Greek island. Gene 610, 9–14 (2017).
5. Alarcon-Riquelme, M. E. et al. Genome-Wide Association Study in an American Ancestry Population Reveals Novel Systemic Lupus Erythematosus Risk Loci and the Role of European Admixture. Arthritis Rheum 68(4), 932–43 (2016).
4. Carmona, F. D. et al. The systemic lupus erythematosus IRF5 risk haplotype is associated with systemic sclerosis. PLoS One 8(1), e54419 (2013).
3. Perreiro-Neira, I. et al. Opposed independent effects and epistasis in the complex association of IRF5 to SLE. Genes Immun 8(5), 429–38 (2007).
2. Bass, E. et al. Methylation matters: binding of Ets-1 to the demethylated Foxp3 gene contributes to the stabilization of Foxp3 expression in regulatory T cells. J Mol Med (Berl) 88(10), 1029–40 (2010).
1. Yang, W. et al. Genome-wide association study in Asian populations identifies variants in ET31 and WDFY4 associated with systemic lupus erythematosus. Nat Genet 46(2), e1008841 (2010).
10. Yang, W. et al. Genome-wide association study in Asian populations identifies variants in ET31 and WDFY4 associated with systemic lupus erythematosus. Nat Genet 41(11), 1234–7 (2009).
9. Hartzog, G. A. & Myers, R. M. Discrimination among potential activators of the beta-globin CACCC element by correlation of binding and transcriptional properties. Mol Cell Biol 13(1), 44–56 (1993).
8. Sigurdsson, S. et al. Comprehensive evaluation of the genetic variants of interferon regulatory factor 5 (IRF5) reveals a novel 5 bp length polymorphism as strong risk factor for systemic lupus erythematosus. Hum Mol Genet 17(6), 872–81 (2008).
7. Huang, Y. T. et al. Transcriptional activation of the CAMP-responsive promoter in human T cells is regulated by protein phosphatase 2A-mediated dephosphorylation of SP-1 and reflects disease activity in patients with systemic lupus erythematosus. J Biol Chem 286(3), 1795–801 (2011).
6. Yang, W. et al. Genome-wide association study in Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat Genet 41(11), 1234–7 (2009).
5. Bartek, J. & Lukas, J. Glucocorticoid-induced leucine zipper (GILZ) inhibits B cell activation in systemic lupus erythematosus. Ann Rheum Dis 69(7), 1329–32 (2010).
4. van der Meer, R. W. et al. Methylation matters: binding of Ets-1 to the demethylated Fox3 gene contributes to the stabilization of Foxp3 expression in regulatory T cells. J Mol Med (Berl) 88(10), 1029–40 (2010).
3. He, Q. et al. Expression of Ets-1 and FOXP3 mRNA in CD4(+)CD25 (+) T regulatory cells from patients with systemic lupus erythematosus. Clin Exp Med 14(4), 375–81 (2014).
2. Sullivan, K. E. et al. 3′ polymorphisms of ET31 are associated with different clinical phenotypes in SLE. Hum Mutat 16(1), 49–53 (2000).
1. Adrianto, I. et al. Association of two independent functional risk haplotypes in TNIP1 with systemic lupus erythematosus. Arthritis Rheum 64(11), 3695–705 (2012).
10. Ramírez, V. P., Gurevich, I. & Aneskievich, B. J. Emerging roles for TNIP1 in regulating post-receptor signaling. Cytokine Growth Factor Rev 23(3), 109–18 (2012).
9. Desnues, B. et al. The transcriptional repressor Gfi1 prevents lupus autoimmunity by restraining TLR7 signaling. Eur J Immunol 46(12), 2801–2811 (2016).
8. Maurano, M. T. et al. Systematic localisation of common disease-associated variation in regulatory DNA. Science 337(6099), 1190–5 (2012).
7. Yilmaz, B. K. & et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature 516(7539), 337–43 (2015).
6. Lower, K. M. et al. Analysis of sequence variation underlying tissue-specific transcription factor binding and gene expression. Hum Mol Genet 34(8), 1140–8 (2013).
5. He, Q., Johnston, J. & Zeitlinger, J. ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. Nat Biotechnol 33(4), 395–401 (2015).
4. Hu, X. et al. Integrating autoimmunity risk loci with gene-expression data identifies specific pathogenic immune cell subsets. Am J Hum Genet 89(4), 496–506 (2011).
3. de Boer, C. M. et al. DICER1 RNase IIIb domain mutations are infrequent in testicular germ cell tumours. BMC Research Notes 5, 569 (2012).
2. Chung, S. A. et al. Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production. PLoS Genet 7(3), e1001323 (2011).
1. Yang, W. et al. Meta-analysis followed by replication identifies loci in or near CDKN1B, TET3, CD80, DRAM1, and ARIDSB as associated with systemic lupus erythematosus in Asians. Am J Hum Genet 92(1), 41–51 (2013).
Author contributions
S.J. and S.C. contributed to the study design, wrote the main manuscript text and conducted experiments, H.F., Q.C., B.R., E.T., J.H., I.R., M.N., M.N., W.D. and A.T. conducted analyses and experiments, S.W and E.M contributed to the study design and reviewed the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-51864-9.

Correspondence and requests for materials should be addressed to S.A.J.

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

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

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license 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 license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019