nature portfolio

Peer Review File

High-throughput identification of functional regulatory SNPs in systemic lupus erythematosus

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This manuscript described a strategy for systematic identification of plausible regulatory non-coding SLE risk variants using a combined experimental and bioinfomatic approach with a focus on genetic signals from European derived populations and their effects on transformed and primary B cells. Starting from 87 lead SNPs including 2180 SNPs fulfilling the search criteria (Figure 1), the combined SNP-seq and bioinformatics filtering resulted in 52 SNPs (Figure 2), which was further reduced to 5 functional SNPs by EMSA and luciferase assay (Figure 3C). The authors worked up one of the 5 SNPs, rs22797550, that the SLE risk G allele affected the interaction with Ikaros (encoded by IKZF1, a SLE risk locus), resulting in modulated IKKε expression. The strength of this manuscript includes the combined experimental and bioinformatic approach and the identification of Ikaros's role in modulating IKKε expression. Among the 5 SNPs, rs2297550 is reported to have regulatory effect on IKBKE expression by an eQTL study (reference 32), the authors extended the previous findings by identifying interaction between the risk allele and Ikaros. The results would have been more interesting and novel if they worked up the less well-known loci (WBP2/TRIM47 or BLTP3A/ TAF11, Figure 3C). This manuscript might be improved as follows:
1. Please extend Figure 1 to include the overall experimental designs, decision trees and filtering at each step, which will be helpful to the readers.
2. The T allele-imbalanced binding of rs2814955 appeared to occur in the absence of nuclear extract in Figure 3A. Please double check the lane legend above.
3. It appears that the same subjects were used for IKBKE mRNA (sFig. 7) and IKKε protein (Fig. 5) measurement in PBMCs stratified by genotypes. Do levels of transcripts correlate with IKKε proteins in these individuals in the presence or absence of LPS stimulation? Could the authors include PBMCs without LPS stimulation data in sFig 7 and comment if their results are similar or different from those described in reference 32?
4. sFig 6 showed 4 SNPs with differential binding in pulldown / silver staining of two biological replicates indicated by red circles, which looked quite different in the two replicates. Could the authors comment on their criteria for either positive or negative inclusion and the robustness of this assay?
5. Could the authors repeat experiments in Figure 5 in SLE patients to help elucidate if the interaction of Ikaros to risk allele carriers suppress or upregulate IKKε expression in B, CD8 and NK cells? There are limitations in drawing conclusions using Daudi cell line and LPS-stimulated cells from healthy subjects.
6. Given that IKZF1 is a SLE risk locus, could the authors comment on eQTL analysis of IKZF1 risk variants in the literature (if any) and if the co-inheritance of IKBKE and IKZF1 risk variants amplifies the genetic risk for lupus?

Reviewer #2 (Remarks to the Author):

The manuscript by Wang et al., identifies a novel regulatory pathway involving a SLE GWAS variant, rs2297550, Ikaros and IKKe. In doing so the authors present a thorough and systematic approach resulting in a mechanistic explanation for the genetic association with SLE. This should be a prime objective of human genetics in the post-GWAS era. However, through no fault of their own, the manuscript displays the challenges and limitations of current technologies that do not lend themselves to biochemical characterization of large numbers of SNPs as well as biological differences between cell line models and primary cells. Overall the work presented in this manuscript is well done, well controlled, properly analyzed and interpreted. It seems clear that Ikaros binds with allele specific affinity to rs2297550 and that binding modulates rRNA and protein of IKBKE. However, the context of the effect, being a repressor or activator, is still in question. Extending these findings to a more definitive conclusion about that, would improve the overall manuscript. In addition, there are some other concerns that need to be addressed.

1. The authors state that the focus is on B cells but then use PBMC extracts for some of the experiments. Since nuclear protein composition is likely to vary by cell type, why did the authors not isolate be cells from the PBMC? Was there concern about the yield of protein from B cells needed for EMSA? The apheresis leukoreduction chambers used for isolation should have provided a large number of B cells.

2. In Figure 1D there is strong correlation between PBMC and BL2 recovered allele counts after using SNP-seq. Is this expected given a more diverse nucleoprotein extract from PBMC? Shouldn't some PBMC or BL2 specific snp enrichments have been found?

3. There are discordant effects on mRNA and protein expression between Daudi and primary PBMCs. Stimulation produced a wider dynamic range in genotype expression in PBMC. Why not stimulate the Daudi cells and see if the worsens or improves the discordance?

4. A table should be provided that informs the reader where the 52 plausible functional variant came from with respect to the previous experimental and bioinformatic approaches. For example, how many of the 52 snps were identified by SNP-seq and how many were identified only from bioinformatic analyses? Overall a better job of walking the reader through the various filtering steps and how the retained snps pass through would be helpful.

5. Of the 9 EMSA validated variants, how many of these were identified by SNP-seq? The expectation would be that all of them should have been captured in that assay.

6. Supp Table 6 is confusing as the reader can't easily determine the enrichment of IKZF1 with rs2297550. Column names are not intuitive. In general, the authors should go over all the supp tables and ask the question as to how an uninitiated reader of this paper can interpret the data in the columns.

7. Figure 8 would be more accurate if reflected the cell type variability in protein vs. mRNA expression in IKBKE.

8. Modify figure 4B such that the blots are divided by the genotypes (homo ref, het, homo alt) then by the clone numbers. The string of CC and GG at the top is difficult to read.

9. A fundamental frustration with the overall approach is that it requires significant filtering of variants at multiple steps. Even if SNP-seq is a more rapid method for putative functional variant identification, biochemical follow-up still necessitates filtering to reasonable numbers of variants, thus leaving a lot still on the table.

Reviewer #3 (Remarks to the Author):

The manuscript entitled, "High-throughput identification of functional regulatory SNPs in systemic lupus erythematosus", is seeking to find functional risk alleles and their regulatory mechanisms contributing lupus development/ pathogenesis. Since findings of many lupus associated risk alleles after GWAS, many researchers have tried to find their functional roles in lupus pathogenesis, but most turned out little meaningful results. To overcome narrow spectrum or biased approaches, authors attempted to find causal SNPs by using high throughput approaches and combinatorial approaches using experimental and informatics. The study found several meaningful SNPs and suggest one potential pathways. Overall study design is well considered and has a strong scientific rigor including key controls. However, hew concerns should be addressed for publication.

1. Major concern: contradictable data and biased explanation in role of Ikros binding and Ikke expression. FIg. 8 suggests that an increased binding of Ikaros to G/G allele of rs2297550 works as a positive regulator of Ikke expression, subsequently increasing risk for lupus development. However, Fig. 4 showed an opposite result (G/G allele has less transcripts of Ikke in Daudi B cel line).
2. Authors needs to clarify whether Ikke expression and Ikaros binding to alleles were conducted in resting or stimulated conditions in Fig. 5 and Fig. 6.
3. Since Ikaros functions as both positive and negative regulator depending on the recruiting partners, author may consider this to explain the data in different experiments.
4. All the study was done with nucleic extracts from healthy PBMCs or B cells line. What about nucleic extracts from SLE patients?

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

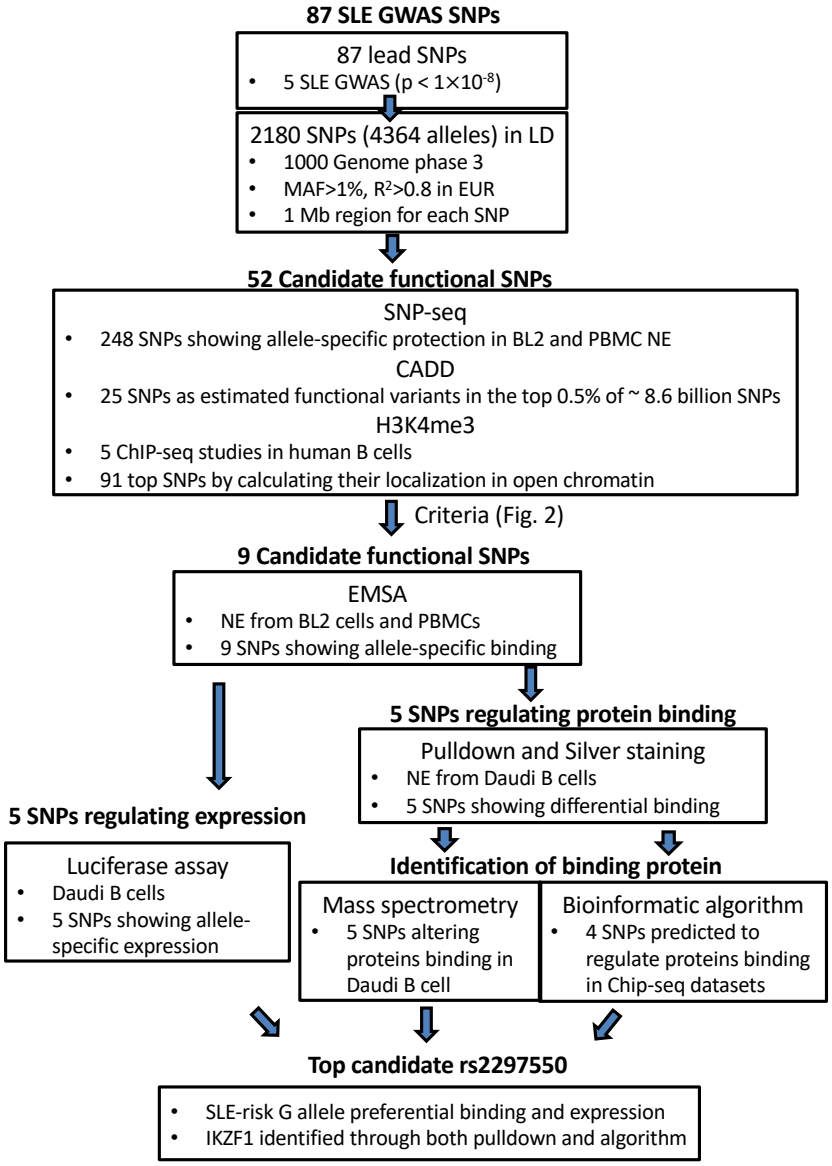
This manuscript described a strategy for systematic identification of plausible regulatory non-coding SLE risk variants using a combined experimental and bioinfomatic approach with a focus on genetic signals from European derived populations and their effects on transformed and primary B cells. Starting from 87 lead SNPs including 2180 SNPs fulfilling the search criteria (Figure 1), the combined SNP-seq and bioinformatics filtering resulted in 52 SNPs (Figure 2), which was further reduced to 5 functional SNPs by EMSA and luciferase assay (Figure 3C). The authors worked up one of the 5 SNPs, rs22797550, that the SLE risk G allele affected the interaction with Ikaros (encoded by IKZF1, a SLE risk locus), resulting in modulated IKKε expression. The strength of this manuscript includes the combined experimental and bioinformatic approach and the identification of Ikaros's role in modulating IKKε expression. Among the 5 SNPs, rs2297550 is reported to have regulatory effect on IKBKE expression by an eQTL study (reference 32), the authors extended the previous findings by identifying interaction between the risk allele and Ikaros. The results would have been more interesting and novel if they worked up the less well-known loci (WBP2/TRIM47 or BLTP3A/ TAF11, Figure 3C).

RESPONSE: We thank the Reviewer for this summary. As the Reviewer is aware, testing each variant is an involved process (e.g. for *IKBKE*, all panels of Figures 4-7 and S7). We agree that each locus identified is interesting. The success in identification of the novel Ikaros/IKKe connection gives us confidence that the others will similarly be fruitful to investigate in coming years.

This manuscript might be improved as follows:

1. Please extend Figure 1 to include the overall experimental designs, decision trees and filtering at each step, which will be helpful to the readers.

RESPONSE: As suggested, we have markedly expanded Figure 1, panel A, which we agree helps to clarify our approach. This is attached here for the Reviewer's convenience.

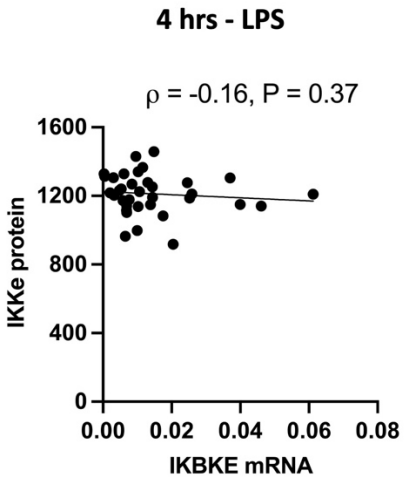


2. The T allele-imbalanced binding of rs2814955 appeared to occur in the absence of nuclear extract in Figure 3A. Please double check the lane legend above.

RESPONSE: Indeed, the lane circled contains nuclear extract but no competitor, parallel with same lane in the other gels shown. We have corrected the lane legend. We thank the Reviewer for catching this error.

3. It appears that the same subjects were used for IKBKE mRNA (sFig. 7) and IKKε protein (Fig. 5) measurement in PBMCs stratified by genotypes. Do levels of transcripts correlate with IKKε proteins in these individuals in the presence or absence of LPS stimulation? Could the authors include PBMCs without LPS stimulation data in sFig 7 and comment if their results are similar or different from those described in reference 32?

RESPONSE: This figure is now Figure S8 because of the addition of a new Figure S7. There is no correlation between IKBKE mRNA and IKKe protein levels in PBMCs stimulated with LPS for 4 hours (see below, now included in revised Figure S8).



Unfortunately, we have no further cells left from these donors to analyze the IKBKE mRNA levels in resting PBMCs. We agree that replication of data from ref 32 could be interesting, but this is not an essential step for our conclusions, and so we hope the Reviewer were understand our decision not to initiate another round of donor recruitment, which we estimate may have required 4-6 months.

4. sFig 6 showed 4 SNPs with differential binding in pulldown / silver staining of two biological replicates indicated by red circles, which looked quite different in the two replicates. Could the authors comment on their criteria for either positive or negative inclusion and the robustness of this assay?

RESPONSE: Bands highlighted represent differences between alleles that resolve in the presence of competitor. We recognize that this judgment is ultimately subjective, which is why we include here all primary data generated. Differences in overall staining intensity between replicates likely reflect incubation time during silver staining. To confirm the silver staining, we utilized a quantitative approach involving pulldown assays and mass spectrometry. We identified nuclear proteins that bound differentially to five SNPs that demonstrated variable binding in at least one replicate of silver staining (**Supplementary Table 6**).

5. Could the authors repeat experiments in Figure 5 in SLE patients to help elucidate if the interaction of Ikaros to risk allele carriers suppress or upregulate IKKε expression in B, CD8 and NK cells? There are limitations in drawing conclusions using Daudi cell line and LPS-stimulated cells from healthy subjects.

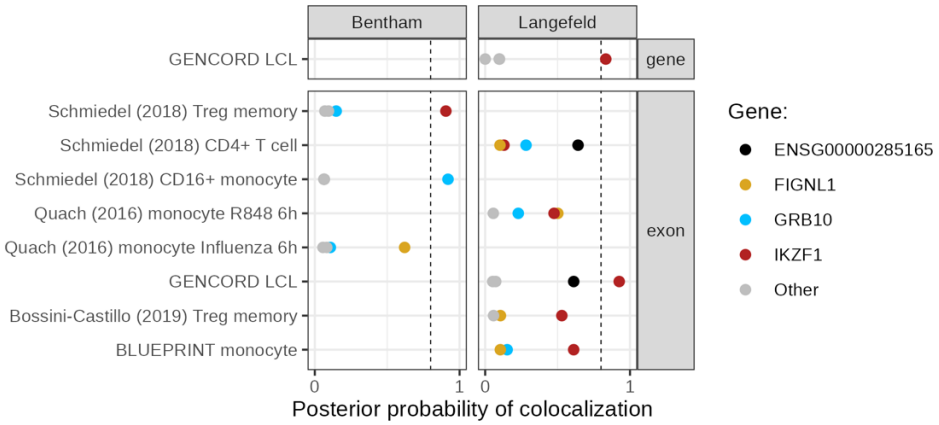
RESPONSE: Our starting point in these studies was GWAS, a method to identify alleles within the genome that are associated with prevalence of SLE – i.e., alleles in altered frequency in patients vs. healthy controls. For a variant to be a GWAS "hit", it must alter risk for disease development. In combination with other polymorphisms and in the appropriate environment, the functional polymorphism must be operative before a patient actually has disease. Variants that influence biology once disease would be detected instead in a "disease progression" or "disease severity" GWAS, where different types of patients are being compared (e.g. severe vs. non-severe patients); such data were not

available to us and were not studied in this project. Thus, the experiment posed by the Reviewer is interesting but not directly material to the question we investigated, which is to identify non-coding variants that drive the incidence of lupus. We have added further text in the Discussion to underscore why, in a study devoted to lupus, it is healthy donors rather than lupus patients who are the relevant research subjects.

We note as an aside that the studies proposed by the Reviewer would also be extraordinarily difficult. We recruited research subjects based on genotype from among >65,000 genotyped donors available through the Mass General Brigham Biobank. We have no access to a comparable bank of genotyped SLE donors. Even were such a bank available, we would face potential confounding on immune phenotypes from therapies such as steroids, hydroxychloroquine, mycophenolate, and rituximab. We are therefore fortunate that the interpretability of our findings in no way depends on the impact of the variants in SLE patients.

6. Given that IKZF1 is a SLE risk locus, could the authors comment on eQTL analysis of IKZF1 risk variants in the literature (if any) and if the co-inheritance of IKBKE and IKZF1 risk variants amplifies the genetic risk for lupus?

RESPONSE: We have performed colocalization analyses with coloc for the *IKZF1* SLE GWAS locus and QTLs reported by the eQTL Catalogue. We tested eQTL and exon-QTL data for immune tissues and cell types, including all genes or exons within ±500kb from the risk variant and with at least one QTL variant passing FDR 5%. Using a posterior probability >0.8 for the hypothesis that there is a shared causal variant between the GWAS and the eQTL signals, we found evidence of colocalization for *IKZF1* eQTLs in one dataset of lymphoblastoid cell lines (LCLs). We also found evidence of colocalization for exon-level QTLs in LCLs and T cells. However, these QTL datasets would not pass a stringent significance threshold, so they should be considered as "suggestive QTLs". For example, the LCL eQTL dataset has no p-value lower than $1x10^{-5}$. Regarding the exon-QTL datasets, if we apply a gene-level FDR threshold to account for the high number of molecular traits tested per gene, no colocalization would be found. Previous colocalization studies that used a stringent significance threshold have not found evidence of colocalization between SLE GWAS and *IKZF1* QTL data (PMID: 33926512). We have added this colocalization analysis as new Supplementary Figure 9 and we include the figure below, acknowledging the limitations of the analysis in the figure legend.



Open Targets Genetics (https://genetics.opentargets.org/) identified *IKZF1* as the likely target gene in this risk locus with their locus-to-gene (L2G) algorithm for the following GWAS: Bentham et al. 2015,

Langefeld et al. 2017, Yin et al. 2020. This algorithm considers distance to lead variant, QTL colocalization evidence, chromatin interaction, and variant pathogenicity. Overall, these findings suggest that the likely target gene for the risk variant in the locus is indeed *IKZF1*.

We investigated the possibility of performing an interaction analysis between the lead variant at the *IKBKE* locus and the lead variant at the *IKZF1* locus. To this end, we enlisted new collaborators (and new co-authors), Miranda Manion and Carl Langefeld. Dr. Langefeld's team attempted this analysis using the data from his GWAS (Langefeld et al. *Nat Commun* 2017 Jul 17:8:16021). He found that the Immunochip did not have sufficient coverage for the *IKZF1* locus to render an informative analysis. In a second alternative, the Japanese SLE cohort from Dr. Chikashi Terao, *IKZF1* did not reach genome-wide significance, rendering this alternative also impossible. We note in the Limitations text of the revised manuscript that we sought opportunities to perform an assessment of epistasis, but that this analysis proved impossible for technical reasons and so the hypothesis remains to be tested. Fortunately, such an analysis, while interesting, is not material to the core conclusions of our study.

Reviewer #2 (Remarks to the Author):

The manuscript by Wang et al., identifies a novel regulatory pathway involving a SLE GWAS variant, rs2297550, Ikaros and IKKe. In doing so the authors present a thorough and systematic approach resulting in a mechanistic explanation for the genetic association with SLE. This should be a prime objective of human genetics in the post-GWAS era. However, through no fault of their own, the manuscript displays the challenges and limitations of current technologies that do not lend themselves to biochemical characterization of large numbers of SNPs as well as biological differences between cell line models and primary cells. Overall the work presented in this manuscript is well done, well controlled, properly analyzed and interpreted. It seems clear that Ikaros binds with allele specific affinity to rs2297550 and that binding modulates rRNA and protein of IKBKE. However, the context of the effect, being a repressor or activator, is still in question. Extending these findings to a more definitive conclusion about that, would improve the overall manuscript.

RESPONSE: We thank the Reviewer for highlighting the importance and also the challenges intrinsic to the study of non-coding variants. Since many transcription factors can either work to amplify or suppress gene expression, depending on the cellular context and the presence of co-factors, it is common to find that an interaction between a transcription factor and its DNA motif has impact that varies with study conditions. We highlighted just such a difference in the effect of rs2297550 between Daudi cells and primary healthy donor cells. To overcome this hurdle and provide confidence in our findings, we employed distinct and orthogonal approaches, including (1) a high-throughput SNP binding screen, SNP-seq; (2) SNP-specific assays including Western, luciferase, and CRISPR genome editing; (3) identification of the transcription factor and demonstration of its functional effect; and (4) recruitment of donors bearing each genotype to confirm the effect in primary cells.

In such cases, the "definitive conclusion" is not that the binding of Ikaros to rs2297550 is either repressive or activating but that it can be either, depending on context. As highlighted in the Discussion, we favor the suggestion that the interaction drives SLE risk by activating IKBKE expression, since this is what we see in primary healthy donors, and since it makes biological sense in the context of incipient SLE that an increase in the product of an interferon-generating gene is likely to be pathogenic.

In the Discussion, we are careful to claim that we have solved only a part of the puzzle, working out the effect of this variant in one cellular system but leaving open the possibility that the Ikaros-rs2297550

interaction has effects in lineages and activation states not studied here. We do not however view this caveat as diminishing the interest or value in our work, since mechanistic questions remain around essentially all causal genetic variants, even coding variants affecting e.g. *PTPN22* and the HLA locus.

In addition, there are some other concerns that need to be addressed.

1. The authors state that the focus is on B cells but then use PBMC extracts for some of the experiments. Since nuclear protein composition is likely to vary by cell type, why did the authors not isolate be cells from the PBMC? Was there concern about the yield of protein from B cells needed for EMSA? The apheresis leukoreduction chambers used for isolation should have provided a large number of B cells.

RESPONSE: For this study, we decided to cast a broad net, focusing on B cells but not necessarily limiting ourselves to this lineage. For SNP-seq, using PBMC alone, we were able to complete the study (in triplicate). We then sought to triangulate to B cells through parallel study of BL2 B cells. Fortunately, this effort proved fruitful, leading to conclusions that we could then validate in primary B cells from genotyped human donors.

2. In Figure 1D there is strong correlation between PBMC and BL2 recovered allele counts after using SNP-seq. Is this expected given a more diverse nucleoprotein extract from PBMC? Shouldn't some PBMC or BL2 specific snp enrichments have been found?

RESPONSE: The SNPs studied were selected because they were GWAS hits in SLE, where the epigenetic studies suggest that most of the risk lies in B cells. Thus, most of the variation observed was expected to reflect B cells, including among PBMC.

3. There are discordant effects on mRNA and protein expression between Daudi and primary PBMCs. Stimulation produced a wider dynamic range in genotype expression in PBMC. Why not stimulate the Daudi cells and see if the worsens or improves the discordance?

RESPONSE: To address this point, we conducted the experiment suggested by the Reviewer. Daudi cells base-edited to each genotype were stimulated with LPS for 2, 4, and 8 hours. We found again that expression of *IKBKE* mRNA and IKKε protein both trended lower in G/G cells, retaining the contrast with primary human cells. These data are now included as new Supplementary Figure 7.

4. A table should be provided that informs the reader where the 52 plausible functional variant came from with respect to the previous experimental and bioinformatic approaches. For example, how many of the 52 snps were identified by SNP-seq and how many were identified only from bioinformatic analyses? Overall a better job of walking the reader through the various filtering steps and how the retained snps pass through would be helpful.

RESPONSE: We thank the Reviewer for helping us to clarify the study. As per Reviewer 1, we expanded a new Figure 1A to highlight the study flow and added new Supplementary Table 6 to define how each variant as prioritized for further study.

5. Of the 9 EMSA validated variants, how many of these were identified by SNP-seq? The expectation would be that all of them should have been captured in that assay.

RESPONSE: Of these 9 variants, 1 (rs9907966) was identified by SNP-seq. We note that the experimental conditions difference between SNP-seq and EMSA, likely accounting for the discordance.

6. Supp Table 6 is confusing as the reader can't easily determine the enrichment of IKZF1 with rs2297550. Column names are not intuitive. In general, the authors should go over all the supp tables and ask the question as to how an uninitiated reader of this paper can interpret the data in the columns.

RESPONSE: We have reviewed each Supplementary Table to ensure that each is now more easily interpretable. For all supplemental tables that are not straightforward to understand, we added one "read_me" sheet to explain each column.

7. Figure 8 would be more accurate if reflected the cell type variability in protein vs. mRNA expression in IKBKE.

RESPONSE: We believe that the effect of the interaction between Ikaros and rs2297550 with the highest biological plausibility is the one shown, because this is what we observe in healthy donors, as noted above. However, we agree with the Reviewer that it is important to highlight for readers that this direction of effect may not be uniform across all cell types and have edited the legend accordingly.

8. Modify figure 4B such that the blots are divided by the genotypes (homo ref, het, homo alt) then by the clone numbers. The string of CC and GG at the top is difficult to read.

RESPONSE: We have adjusted the labels accordingly and agree that this is clearer.

9. A fundamental frustration with the overall approach is that it requires significant filtering of variants at multiple steps. Even if SNP-seq is a more rapid method for putative functional variant identification, biochemical follow-up still necessitates filtering to reasonable numbers of variants, thus leaving a lot still on the table.

RESPONSE: We share the Reviewer's frustration. GWAS implicates >100 loci in SLE, mostly non-coding. Each of these variants represents genetic proof that the pathway implicated is directly relevant to human polygenic SLE. Even identifying them all is difficult, and yet this identification is of interest only if downstream mechanism can be identified also. To decide where to begin, we sought some set of filters to help prioritize variants to go after first. Our study of rs2297550 leads us to believe that the filters used were informative, but we recognize also how many causal mechanisms were likely bypassed in the process.

Reviewer #3 (Remarks to the Author):

The manuscript entitled, "High-throughput identification of functional regulatory SNPs in systemic lupus erythematosus", is seeking to find functional risk alleles and their regulatory mechanisms contributing lupus development/ pathogenesis. Since findings of many lupus associated risk alleles after GWAS, many researchers have tried to find their functional roles in lupus pathogenesis, but most turned out little meaningful results. To overcome narrow spectrum or biased approaches, authors attempted to find causal SNPs by using high throughput approaches and combinatorial approaches using experimental and informatics. The study found several meaningful SNPs and suggest one potential pathways. Overall study design is well considered and has a strong scientific rigor including key controls.

RESPONSE: We are grateful to the Reviewer for this summary of the state of the field and the positive assessment of our results.

However, hew concerns should be addressed for publication.

1. Major concern: contradictable data and biased explanation in role of Ikros binding and Ikke expression. FIg. 8 suggests that an increased binding of Ikaros to G/G allele of rs2297550 works as a positive regulator of Ikke expression, subsequently increasing risk for lupus development. However, Fig. 4 showed an opposite result (G/G allele has less transcripts of Ikke in Daudi B cel line).

RESPONSE: We agree with this point, and in the submitted version of our manuscript, we called attention to this discrepancy:

> "Interestingly, the relationship between genotype and IKKε varied with context. In Daudi B lymphoblasts, a line derived from a patient with Burkitt lymphoma, introduction of the risk GG genotype decreased IKKε expression and deletion of Ikaros enhanced it, defining Ikaros as an inhibitor of IKBKE. By contrast, in healthy donors, circulating B cells showed an increase in IKKε with risk allele carriage, including a gene dose effect, but only after LPS stimulation. Similar regulation was observed in CD8+ T cells, NK cells, and monocytes. The dependence of the impact of genetic variants on cell type and developmental/activation state is well recognized [40]. Since Ikaros can toggle between activator and repressor depending on associated cofactors, it may be that the variation observed here reflects differential expression of these proteins, or on other genetic/epigenetic differences between the lineages studied [41]. Under conditions other than those tested here, it is possible that the risk allele at rs2297550 could suppress IKKε in primary cells, as it does in Daudi cells."

In our introductory response to Reviewer 2, as provided below, we further discuss why this kind of "contradiction" is expected in transcription factor research, since many (including Ikaros) can be either positive or negative regulators of gene expression. Thus, our findings reflect the irreducible complexity of the system studied.

> RESPONSE TO REVIEWER 2 COPIED FROM ABOVE: We thank the Reviewer for highlighting the importance and also the challenges intrinsic to the study of non-coding variants. Since many transcription factors can either work to amplify or suppress gene expression, depending on the cellular context and the presence of co-factors, it is common to find that an interaction between a transcription factor and its DNA motif has impact that varies with study conditions. We highlighted just such a difference in the effect of rs2297550 between Daudi cells and primary healthy donor cells. To overcome this hurdle and provide confidence in our findings, we employed distinct and orthogonal approaches, including (1) a high-throughput SNP binding screen, SNP-seq; (2) SNP-specific assays including Western, luciferase, and CRISPR genome editing; (3) identification of the transcription factor and demonstration of its functional effect; and (4) recruitment of donors bearing each genotype to confirm the effect in primary cells.
>
> In such cases, the "definitive conclusion" is not that the binding of Ikaros to rs2297550 is either repressive or activating but that it can be either, depending on context. As highlighted in the Discussion, we favor the suggestion that the interaction drives SLE risk by activating IKBKE expression, since this is what we see in primary healthy donors, and since it makes biological

sense in the context of incipient SLE that an increase in the product of an interferon-generating gene is likely to be pathogenic.

In the Discussion, we are careful to claim that we have solved only a part of the puzzle, working out the effect of this variant in one cellular system but leaving open the possibility that the Ikaros-rs2297550 interaction has effects in lineages and activation states not studied here. We do not however view this caveat as diminishing the interest or value in our work, since mechanistic questions remain around essentially all causal genetic variants, even coding variants affecting e.g. *PTPN22* and the HLA locus.

2. Authors needs to clarify whether Ikke expression and Ikaros binding to alleles were conducted in resting or stimulated conditions in Fig. 5 and Fig. 6.

RESPONSE: In Figure 5, the top row (A-F) reflects resting conditions and the second row (G-L) represents 4h LPS stimulation. In Figure 6, we added to the legend that these studies were done in unstimulated cells.

3. Since Ikaros functions as both positive and negative regulator depending on the recruiting partners, author may consider this to explain the data in different experiments.

RESPONSE: We very much agree with the Reviewer on this point, as noted above.

4. All the study was done with nucleic extracts from healthy PBMCs or B cells line. What about nucleic extracts from SLE patients?

RESPONSE: We agree that may appear unusual that a study of SLE risk variants should not study lupus patients. However, as detailed in our response to Reviewer 1 point 5, as provided below, we detail that healthy donors are in fact the proper study subjects, since GWAS "hits" drive transition from healthy to disease and therefore must be active in the healthy state to be GWAS hits at all. We have clarified this point by adding the following paragraph to the discussion.

Our studies in genotyped donors employed healthy donors rather than SLE patients. This choice was deliberate. We derived our initial candidate SNPs from GWAS that compared patients with SLE to patients lacking SLE. This design identifies variants modulating the risk that a healthy person will develop disease, and therefore that must be operative before SLE appears – that is, in healthy individuals. Such variants may also be relevant for patients with SLE, modulating disease phenotype, severity, or response to therapy, but such effects remain conceptually distinct from how they predispose to incident disease, as studied here.

RESPONSE TO REVIEWER 1 COPIED FROM ABOVE: Our starting point in these studies was GWAS, a method to identify alleles within the genome that are associated with prevalence of SLE – i.e. alleles in altered frequency in patients vs. healthy controls. For a variant to be a GWAS "hit", it must alter risk for disease development, and therefore be operative before a patient actually has a disease. Variants that influence biology once disease would be detected instead in a "disease progression" or "disease severity" GWAS, where different types of patients are being compared (e.g. severe vs. non-severe patients); such data were not available to us and were not studied in this project. Thus, the experiment posed by the Reviewer is interesting but not directly material to the question we investigated, which is to identify non-coding variants that

drive the incidence of lupus. We have added further text in the Discussion to underscore why, in a study devoted to lupus, it is healthy donors rather than lupus patients who are the relevant research subjects.

We note as an aside that the studies proposed by the Reviewer would also be extraordinarily difficult. We recruited research subjects based on genotype from among >65,000 genotyped donors available through the Mass General Brigham Biobank. We have no access to a comparable bank of genotyped SLE donors. Even were such a bank available, we would face potential confounding on immune phenotypes from therapies such as steroids, hydroxychloroquine, mycophenolate, and rituximab. We are therefore fortunate that the interpretability of our findings in no way depends on the impact of the variants in SLE patients.

We thank the Reviewers for their attention to our manuscript and hope that the revised manuscript has now addressed all concerns and is found worthy for *Nature Communications*.

Peter A. Nigrovic, MD, for the authors

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have successfully addressed my previous comments. I have no further concerns.

Reviewer #2 (Remarks to the Author):

The authors have adequately addressed my concerns as best they can. I have no further comments to offer.

Reviewer #3 (Remarks to the Author):

Authors responded well to the critiques and the revised manuscript is much improved. Overall, the manuscript is appropriate for publication.