



Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells

Mark N. Lee *et al.*

Science **343**, (2014);

DOI: 10.1126/science.1246980

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of March 12, 2014):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/343/6175/1246980.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2014/03/05/343.6175.1246980.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/343/6175/1246980.full.html#related>

This article **cites 70 articles**, 13 of which can be accessed free:

<http://www.sciencemag.org/content/343/6175/1246980.full.html#ref-list-1>

This article has been **cited by** 1 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/343/6175/1246980.full.html#related-urls>

This article appears in the following **subject collections**:

Genetics

<http://www.sciencemag.org/cgi/collection/genetics>

Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells

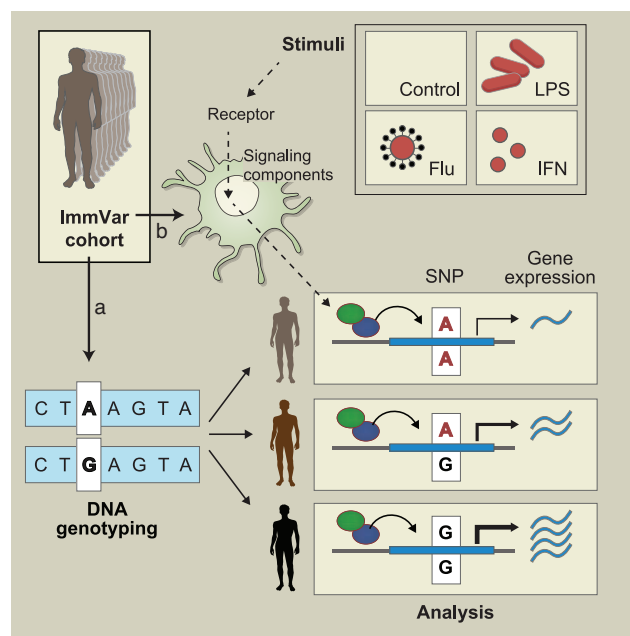
Mark N. Lee,* Chun Ye,* Alexandra-Chloé Villani, Towfique Raj, Weibo Li, Thomas M. Eisenhaure, Selina H. Imboywa, Portia I. Chipendo, F. Ann Ran, Kamil Slowikowski, Lucas D. Ward, Khadir Raddassi, Cristin McCabe, Michelle H. Lee, Irene Y. Frohlich, David A. Hafler, Manolis Kellis, Soumya Raychaudhuri, Feng Zhang, Barbara E. Stranger, Christophe O. Benoist, Philip L. De Jager, Aviv Regev,† Nir Hacohen†

Introduction: Variation in an individual's response to environmental factors is likely to influence susceptibility to complex human diseases. The genetic basis of such variation is poorly understood. Here, we identify natural genetic variants that underlie variation in the host innate immune response to infection and analyze the mechanisms by which such variants alter these responses.

Methods: We derived dendritic cells (DCs) from peripheral blood monocytes of healthy individuals (295 Caucasians, 122 African Americans, 117 East Asians) and stimulated them with *Escherichia coli* lipopolysaccharide (LPS), influenza virus, or the cytokine interferon- β (IFN- β) to generate 1598 transcriptional profiles. We genotyped each of these individuals at sites of common genetic variation and identified the genetic variants that best explain variation in gene expression and gene induction between individuals. We then tested mechanistic predictions from these associations using synthetic promoter constructs and genome engineering.

Results: We identified 264 loci containing genetic variants associated with variation in absolute gene expression in human DCs, of which 121 loci were associated with variation in the induction of gene expression by one or more stimuli. Fine-mapping identified candidate causal single-nucleotide polymorphisms (SNPs) associated with expression variance, and deeper functional experiments localized three of these SNPs to the binding sites of stimulus-activated transcription factors. We also identified a cis variant in the transcription factor, *IRF7*, associated in trans with the induction of a module of antiviral genes in response to influenza infection. Of the identified genetic variants, 35 were also associated with autoimmune or infectious disease loci found by genome-wide association studies.

Discussion: The genetic variants we uncover and the molecular basis for their action provide mechanistic explanations and principles for how the innate immune response to pathogens and cytokines varies across individuals. Our results also link disease-associated variants to specific immune pathways in DCs, which provides greater insight into mechanisms underlying complex human phenotypes. Extending our approach to many immune cell types and pathways will provide a global map linking human genetic variants to specific immunological processes.



READ THE FULL ARTICLE ONLINE

<http://dx.doi.org/10.1126/science.1246980>



Cite this article as M. N. Lee *et al.*, *Science* **343**, 1246980 (2014). DOI: 10.1126/science.1246980

FIGURES IN THE FULL ARTICLE

Fig. 1. A strategy to identify gene-by-environment interactions in the innate immune responses of primary human DCs.

Fig. 2. Genome-wide expression profiles in MoDCs reveal response phenotypes.

Fig. 3. Association analysis reveals cis-eQTLs and cis-reQTLs.

Fig. 4. Functional fine-mapping and mechanism of cis-reQTLs.

Fig. 5. Trans-reQTL association at the *IRF7* cis-regulatory locus.

Fig. 6. Autoimmune and infectious disease-associated SNPs are cis-eQTLs and cis-reQTLs.

SUPPLEMENTARY MATERIALS

Materials and Methods
Supplementary Text
Figs. S1 to S5
Tables S1 to S11
References

RELATED ITEMS IN SCIENCE

Research Article
B. P. Fairfax *et al.*, *Science* **343**, 1246949 (2014).
Perspective
P. K. Gregersen, *Science* **343**, 1087 (2014).

Identifying the genetic basis of variability in the host response to pathogens. A cohort of 534 individuals donated blood for (a) genotyping of common DNA variants and (b) isolation of immune DCs. DCs were stimulated with viral and bacterial components, and the variability in individuals' gene expression responses was mapped to specific DNA variants, which were then shown to affect binding of particular transcription factors.

The list of author affiliations is available in the full article online.

*These authors contributed equally to this work.

†Corresponding author. E-mail: aregev@broad.mit.edu (A.R.); nhacohen@partners.org (N.H.)

Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells

Mark N. Lee,^{1,2,3*} Chun Ye,^{1*} Alexandra-Chloé Villani,^{1,2} Towfique Raj,^{1,2,4} Weibo Li,^{1,3} Thomas M. Eisenhaure,^{1,3} Selina H. Imboywa,² Portia I. Chipendo,² F. Ann Ran,^{1,5,6,7,8} Kamil Slowikowski,⁹ Lucas D. Ward,^{1,10} Khadir Raddassi,¹¹ Cristin McCabe,^{1,4} Michelle H. Lee,² Irene Y. Frohlich,² David A. Hafler,⁸ Manolis Kellis,^{1,10} Soumya Raychaudhuri,^{1,2,12,13} Feng Zhang,^{6,7,8} Barbara E. Stranger,^{14,15} Christophe O. Benoist,² Philip L. De Jager,^{1,2,4} Aviv Regev,^{1,16,17†} Nir Hacohen^{1,2,3†}

Little is known about how human genetic variation affects the responses to environmental stimuli in the context of complex diseases. Experimental and computational approaches were applied to determine the effects of genetic variation on the induction of pathogen-responsive genes in human dendritic cells. We identified 121 common genetic variants associated in cis with variation in expression responses to *Escherichia coli* lipopolysaccharide, influenza, or interferon- β (IFN- β). We localized and validated causal variants to binding sites of pathogen-activated STAT (signal transducer and activator of transcription) and IRF (IFN-regulatory factor) transcription factors. We also identified a common variant in *IRF7* that is associated in trans with type I IFN induction in response to influenza infection. Our results reveal common alleles that explain interindividual variation in pathogen sensing and provide functional annotation for genetic variants that alter susceptibility to inflammatory diseases.

Susceptibility to complex diseases depends on both genetic predisposition and exposure to environmental factors, with interactions between the two (G \times E interactions) likely contributing substantially to disease risk (1, 2). However, the extent and mechanisms by which common human genetic variants interact

with the environment remain poorly explored and have been difficult to detect in clinical studies (1, 3). Genetic analysis of molecular traits, such as gene expression profiles, offers a promising way to dissect the molecular mechanisms underlying G \times E interactions. Expression quantitative trait locus (eQTL) studies have been used to map genetic variants contributing to variation in gene expression, but have largely focused on steady-state expression in humans (4, 5) and thus have excluded G \times E interactions. In model organisms, differences in growth conditions or treatment with various stimuli have revealed the existence of response eQTLs (reQTLs) (6–9), defined as QTLs associated with the change in expression after stimulation. Here, we sought to identify reQTLs in humans, to explain the mechanism by which the environment interacts with these variants, and to determine whether these variants are associated with immune-mediated diseases.

We used dendritic cells (DCs) of the innate immune system as a model system for reQTL studies, with physiological and clinical relevance. DCs play a direct role in the host recognition of pathogens using specialized sensors that engage well-characterized signaling and transcriptional networks. For example, bacterial lipopolysaccharide (LPS) activates two distinct arms of the Toll-like receptor 4 (TLR4) pathway, whereas influenza infection primarily activates the RNA-sensing TLRs (for example, TLR3) and the RIG-I-like receptors (for example, RIG-I) (10). These, in turn, lead to the translocation of transcription factors (TFs) from the cytoplasm into the nu-

cleus to induce the expression of immune genes, including interferon- β (IFN- β) secretion that engages the type I IFN response pathway to induce the expression of hundreds of antiviral effectors. Genetic studies have associated common variants near many genes in these pathways with risk of different inflammatory diseases (11, 12). DCs also play a critical role in the pathologic immune responses underlying inflammatory diseases (11–13), also reflected in recent genome-wide association studies (GWAS) of several diseases (14–17), especially inflammatory bowel disease (14).

Results

Assessing the Impact of Genetic Variation on Pathogen Sensing in Primary Human DCs

We developed an integrated experimental and computational pipeline (Fig. 1A) to identify variability in human DC responses and associate this variability with common genetic variants. First, we optimized a high-throughput protocol to isolate primary CD14⁺CD16^{lo} monocytes from human blood samples (fig. S1, A to F), differentiate them into monocyte-derived DCs (MoDCs), and stimulate them with three immunostimulatory agents (Fig. 1B): *Escherichia coli* LPS, influenza virus, or IFN- β (a cytokine induced by LPS and influenza). Second, we collected genome-wide transcript profiles from resting and stimulated DCs from 30 healthy individuals and computationally identified a “signature” of 415 genes that would be informative of variation in the response in a larger cohort. Third, we generated 1598 transcriptional profiles [using an amplification-free platform suitable for small cell numbers (18)] from DCs isolated from 534 healthy individuals (295 Caucasians, 122 African Americans, and 117 East Asians) (table S1) in four states: resting, LPS-stimulated, influenza-infected, and IFN- β -stimulated. Finally, we associated expression variation with genetic variation to identify cis- and trans-eQTLs and reQTLs, and we investigated candidate reQTLs for their mechanism of action.

Variability of Pathogen-Sensing Responses Between Individuals

To assess the interindividual variability of DC responses, we first profiled genome-wide expression using microarrays in resting, LPS-stimulated, and influenza-infected MoDCs isolated from each of 30 healthy individuals (18 Caucasians, 6 Asians, and 6 African Americans). We found 1413 genes that were regulated in LPS- or influenza-treated cells [\log_2 (fold change) >0.75 or ≤ -1.5 ; false discovery rate (FDR) <0.01 ; fig. S2A and table S2A; see table S2, B to I, for enriched biological processes] at 5 and 10 hours, respectively (time points selected to have maximal expression of induced clusters) (fig. S1, B and C).

We quantified reproducibility in these responses by recalling 12 of the 30 donors 2 to 9 months after the first collection for MoDC

¹Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard, Cambridge, MA 02142, USA. ²Harvard Medical School, Boston, MA 02115, USA. ³Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, MA 02129, USA. ⁴Program in Translational NeuroPsychiatric Genomics, Department of Neurology, Brigham and Women’s Hospital, Boston, MA 02115, USA. ⁵Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA. ⁶McGovern Institute for Brain Research, MIT, Cambridge, MA 02139, USA. ⁷Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139, USA. ⁸Department of Biological Engineering, MIT, Cambridge, MA 02139, USA. ⁹Bioinformatics and Integrative Genomics, Graduate School of Arts and Sciences, Harvard University, Cambridge, MA 02138, USA. ¹⁰Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA 02139, USA. ¹¹Department of Neurology, Yale School of Medicine, New Haven, CT 06511, USA. ¹²Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women’s Hospital, Boston, MA 02115, USA. ¹³Arthritis Research UK Epidemiology Unit, Musculoskeletal Research Group, University of Manchester, Manchester Academic Health Sciences Centre, Manchester M13 9NT, UK. ¹⁴Section of Genetic Medicine, The University of Chicago, Chicago, IL 60637, USA. ¹⁵Institute for Genomics and Systems Biology, The University of Chicago, Chicago, IL 60637, USA. ¹⁶Department of Biology, MIT, Cambridge, MA 02139, USA. ¹⁷Howard Hughes Medical Institute.

*These authors contributed equally to this work.

†Corresponding author. E-mail: aregev@broad.mit.edu (A.R.); nhacohen@partners.org (N.H.)

isolations and profiling (table S1); we identified genes whose expression interindividual variance is significantly higher than their intraindividual variance based on the serial replicates [taking into account known covariates including gender, age, and population and unknown factors using surrogate variable analysis (19)]. Two hundred twenty-two of the 1413 regulated genes (16%) showed significantly higher ($FDR < 0.1$) inter- than intraindividual variability, either in their absolute expression or in the differential expression (stimulated/baseline level) to at least one stimulus (Fig. 2A and table S3). These results suggest that there is consistent variation in these traits that may have a genetic basis.

Expression Profiling of a Pathogen Response Signature

Mapping the genetic basis of interindividual variation in pathogen responses requires profiling of DC gene expression in a larger cohort; this poses a substantial technical challenge given the limited numbers of primary cells and multiple stimuli. Furthermore, the responses of DCs to virus, bacterial ligands, and IFN are limited in scope and do not encompass the entire genome. We therefore defined a 415-gene signature set (Fig. 2B and table S3) that could be monitored in small numbers of cells using a sensitive multiplex RNA detection system (18), allowing us to scale up our study. The signature consisted of (i) all 222 of 1413 regulated genes with greater inter- than intraindividual variability in the microarray study; (ii) all 24 regulated genes exhibiting greater inter- than intrapopulation

variability ($FDR < 0.1$) (table S3) in the microarray study; (iii) 76 genes comprising the known components of the TLR4, TLR3, RIG-I, and IFNAR pathways (Fig. 1B) (10, 20–22); (iv) 61 regulated genes that play key roles in the DC response (for example, *IFNB1* and *IFITM3*) (22–24); (v) 28 regulated genes residing in loci previously associated with autoimmune or infectious diseases (25); and (vi) 35 control genes including those that had among the lowest interindividual variance in the microarray profiles. This representative signature allowed us to monitor genes with high interindividual variability, as well as key components and responses of the pathogen-sensing pathway.

We then generated 1598 transcriptional profiles, using the 415-gene signature, from MoDCs isolated from 534 healthy individuals (and 37 serially collected replicates) in up to four conditions: resting (528 individuals), LPS-stimulated (356), influenza-infected (342), and IFN- β -stimulated (284) (Fig. 2C; fig. S1, E and F; and table S1). The IFN- β stimulus allowed us to partition genes that were induced by both LPS and influenza (cluster III) into those induced secondary to type I IFN signaling (cluster IIIa) or by other mechanisms shared between these bacterial and viral sensing pathways (cluster IIIb), such as nuclear factor κ B (NF- κ B) or activating protein 1 (AP-1) activation (Figs. 1B and 2C). The signature expression profiles clustered similarly to those from the microarray analysis (Fig. 2C and fig. S2C) and captured most of the variance in the genome-wide profile [cross-validation, >0.99 ground-truth Pearson correlation coefficient; fig. S2D (26)],

demonstrating the utility of the signature to capture the genome-wide response.

reQTLs That Modulate Cellular Responses to Pathogens

We mapped cis-eQTLs and cis-reQTLs by testing for association between common single-nucleotide polymorphisms (SNPs) (minor allele frequency $>5\%$; genotyped with Illumina HumanOmniExpress BeadChip) and variation in either absolute transcript abundance (eQTL) or stimulation-induced change in transcript abundance (reQTL) in a nearby gene (in which the transcriptional start site or stop codon is within 1 Mb of the SNP). We pooled individuals from the three human populations together and included covariates and principal components to increase statistical power while avoiding systematic confounders such as population structure and expression heterogeneity. We identified 264 genes with cis-eQTLs in at least one condition (permutation $FDR < 0.05$) (Fig. 3A; fig. S3, A to C; and table S4, A to D) (27). Notably, 22 of 264 genes also associated with additional independent cis-eQTLs after conditioning on the top five most significant SNPs in each cis-eQTL region (table S6 and fig. S3D) (27), which reflects the complexity of the regulatory landscape around each gene.

We detected 121 cis-reQTLs (Fig. 3B and table S4, E to G; permutation $FDR < 0.05$; 91% internal reproducibility in at least two human populations, table S5), the subset of genetic variants that affect the induction of gene expression by LPS, influenza, or IFN- β . Of the 121 genes with cis-reQTLs, 7 associations were found

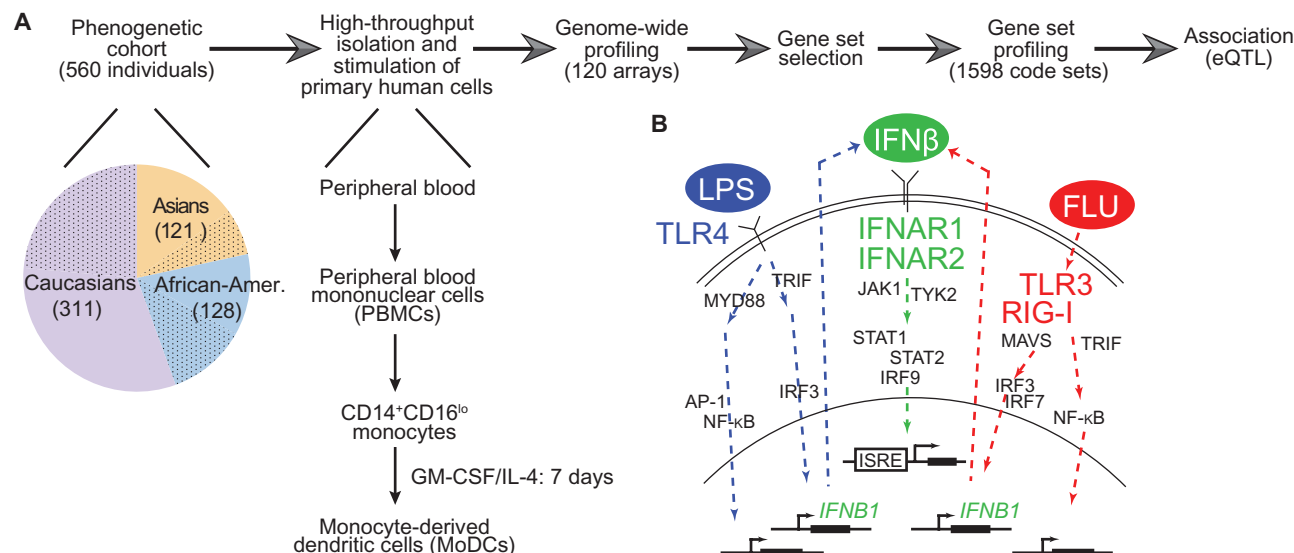


Fig. 1. A strategy to identify gene-by-environment interactions in the innate immune responses of primary human DCs. (A) Strategy used to identify baseline and response eQTLs and reQTLs, consisting of five steps: (i) high-throughput isolation and stimulation of primary human MoDCs from 560 healthy individuals (dotted slices, male; solid-colored slices, female) collected as part of the PhenoGenetic cohort; (ii) whole-genome gene expression measurements in a subset of the cohort; (iii) selection of signature gene set, consisting of regulators and regulated genes; (iv) digital multiplex gene expression measurements of signature genes in the entire cohort; and (v) mapping of genetic variation to

expression variation. GM-CSF, granulocyte-macrophage colony-stimulating factor; IL-4, interleukin-4. **(B)** Model of innate immune pathways activated by three stimuli demonstrating their downstream relationships. LPS from *E. coli* engages the TLR4 receptor; IFN- β engages the heterodimeric IFNAR receptor; influenza A/PR8 (Δ NS1) ("FLU") engages the cytoplasmic TLR3 and RIG-I receptors. Receptor engagement activates signal transduction cascades that regulate expression of inflammatory genes, IFNs, and IFN-stimulated genes. IFNAR activation also occurs during LPS and FLU stimulations because LPS and FLU both induce IFN production, leading to activation of ISREs. JAK1, Janus kinase 1.

only in the influenza condition (for example, *IFNA21*; Fig. 3C and fig. S3B; meta-analysis, $P < 1 \times 10^{-5}$; permutation FDR = 0.02 to 0.03); 15 in both the LPS and influenza conditions (for example, *TEC*; Fig. 3C and fig. S3B); and 57 in all three stimulation conditions (for example, *ARL5B*, *SLFN5*, and *CLEC4F*; Fig. 3, C to E,

and fig. S3B), likely reflecting activation of the shared IFN- β pathway. We hypothesized that causal variants driving cis-reQTLs alter the sequence of genomic elements that respond to TFs downstream of the pathogen-sensing receptors (Fig. 3F), and represent gene-by-environment interactions.

To validate cis-reQTLs, we quantified allele-specific expression for pan-stimulation-specific associations in heterozygote individuals (Fig. 3, D and E). This was feasible for a gene with an exonic SNP (*CLEC4F* rs2075221) that was the most significant reQTL SNP, and for a gene with an exonic SNP (*SLFN5* rs11651240) in linkage

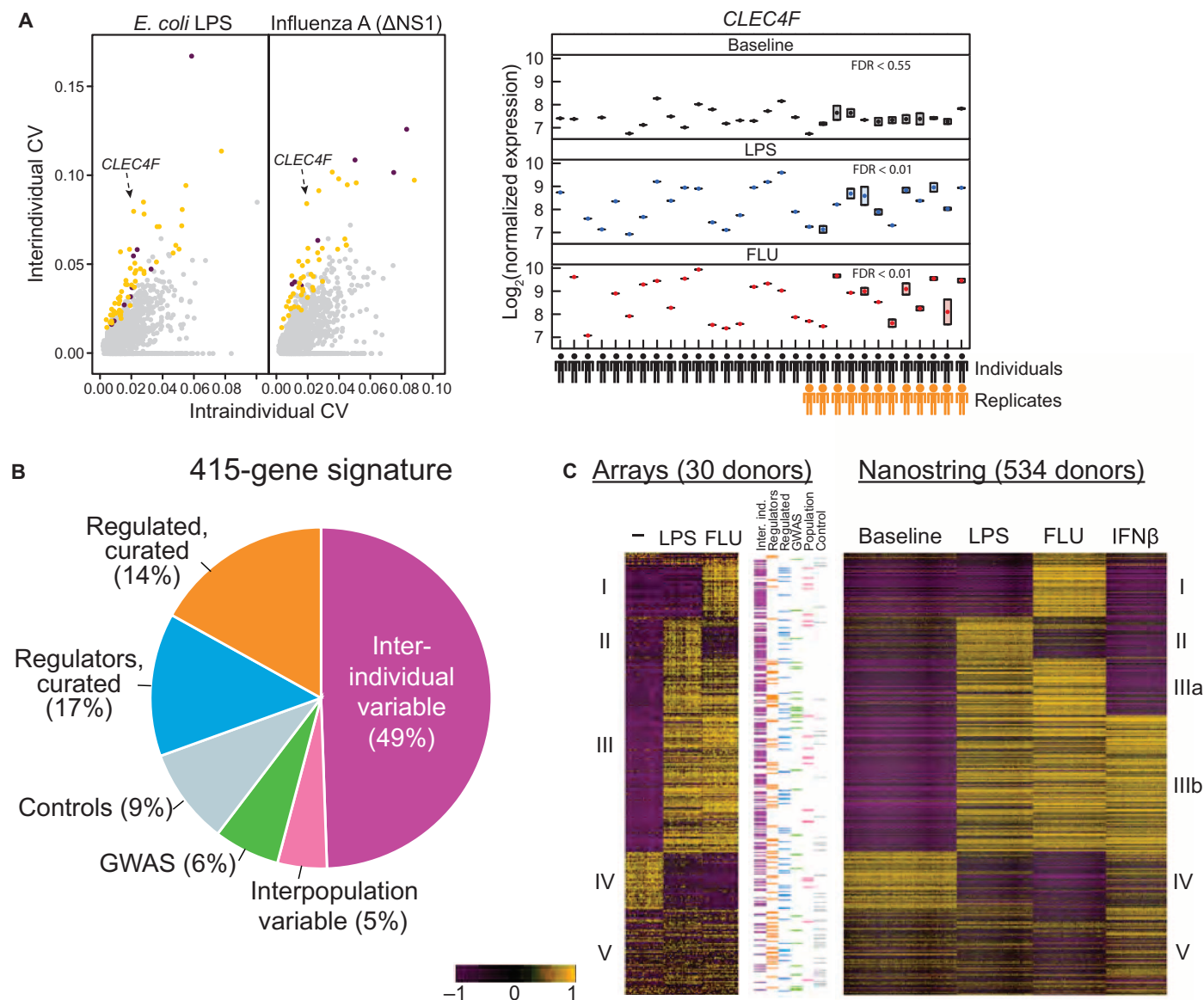


Fig. 2. Genome-wide expression profiles in MoDCs reveal response phenotypes. (A) Coefficient of variation (CV) of gene expression between 30 different donors ("Interindividual CV") plotted against CV of expression within 12 serial replicate samples ("Intraindividual CV") for each differentially regulated (fold change >0.75 or <-1.5) gene after LPS or FLU stimulation. Yellow (up-regulated) and purple (down-regulated) circles represent genes with significant (moderated t test, FDR < 0.1) inter- versus intraindividual variation. (Right) log₂(expression, microarray data) of *CLEC4F* in baseline, LPS-stimulated, and FLU-infected MoDCs from 30 donors and 12 replicates, demonstrating example of a gene that shows significant (FDR < 0.01) inter- versus intraindividual variation after LPS and FLU stimulations but not at baseline (fig. S2B). Standard error of replicate samples ($n = 12$) is shown for each sample. (B) Pie chart of 415 signature genes selected for Nanostring code set: 222 (49%) are regulated genes that showed significant (mixed model variance components test, permutation FDR < 0.1)

inter- versus intraindividual variability; 61 (14%) are curated, regulated genes with a known function in the innate immune response; 76 (17%) are curated regulators in the TLR4, TLR3, RIG-I, and IFNAR pathways; 41 (9%) are control genes including low-variance genes, sex-specific genes, and nonexpressed genes; 28 (6%) are regulated genes that were reported in the regions of autoimmune and infectious disease GWAS; and 21 (5%) are regulated genes that showed significant inter- versus intrapopulation variability. (C) Gene expression heat map of the 415-gene signature in MoDCs from the microarray study (30 individuals) and the Nanostring study (534 individuals). Each row represents a gene; each column represents a donor sample at baseline, stimulated with LPS, infected with FLU, or stimulated with IFN- β . Rows were clustered by k -means clustering of Nanostring data set with major clusters (I, II, IIIa, IIIb, and IV) labeled. Between the two heat maps, each row was labeled with colored dashes corresponding to one of the six categories described in (B).

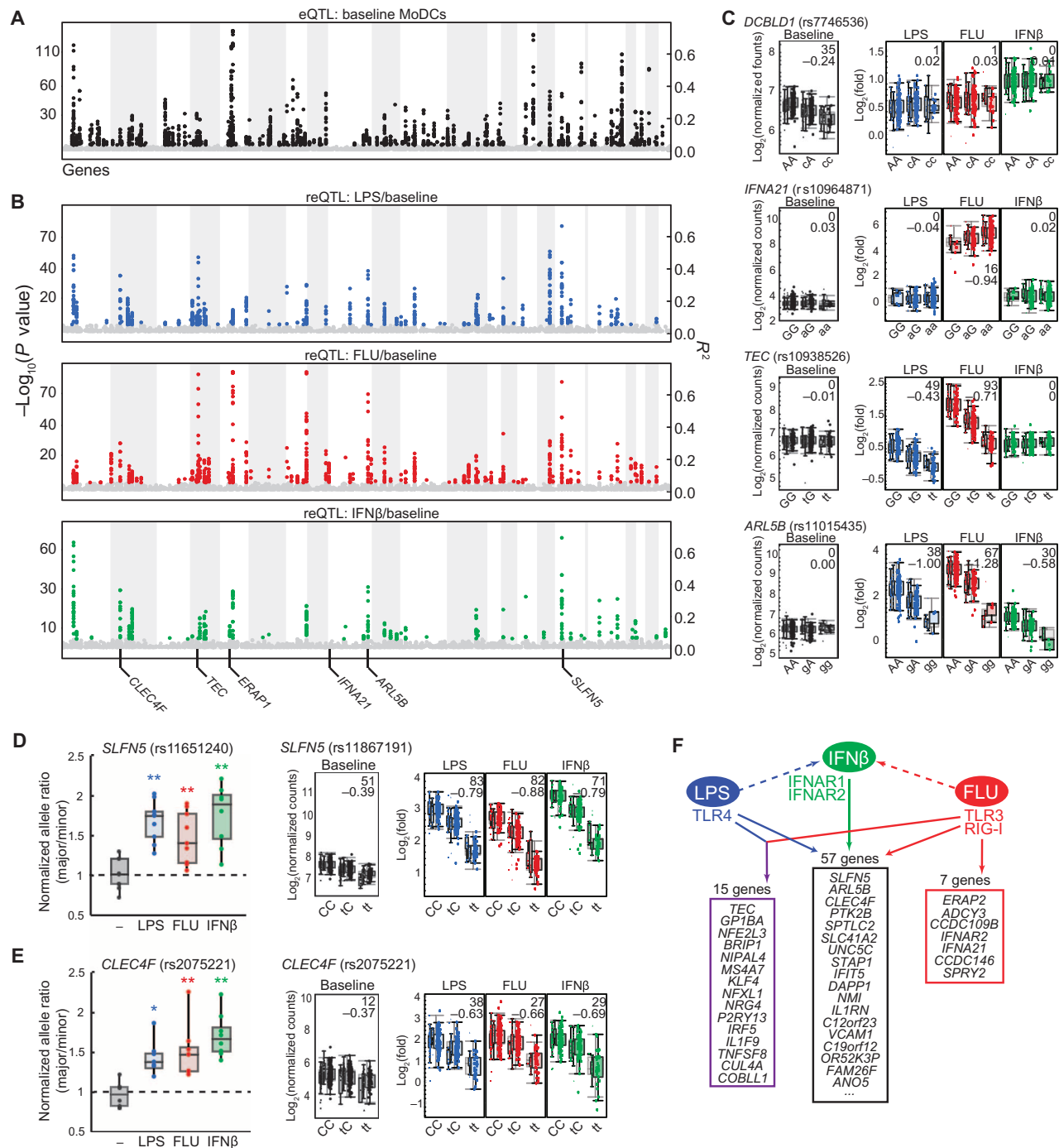


Fig. 3. Association analysis reveals cis-eQTLs and cis-reQTLs. (A and B) Manhattan plot of cis-eQTLs (A) (baseline expression) and cis-reQTLs (B) (LPS-, FLU-, and IFN- β -stimulated fold changes relative to baseline) showing $-\log_{10}(P$ values) (left y axis) and R^2 values (right y axis) for all cis-SNPs, displayed on the x axis with associated genes ordered by chromosomal location. (C) Box-whisker plots showing expression [left; $\log_2(\text{nCounts})$, y axis] or fold change [right; $\log_2(\text{fold})$, y axis] of *DCBLD1*, *IFNA21*, *TEC*, and *ARL5B* in resting, LPS-stimulated, FLU-infected, and IFN- β -stimulated MoDCs as a function of genotype of the respective cis-SNPs (x axis: rs27434, rs10964871, rs10938526, and rs11015435). African Americans, Asians, and Europeans in this order are displayed as separate box-whisker plots adjacent to each other in each condition. $-\log_{10}(P$ values) and β statistics are displayed in top right corners. (D and E) Allelic imbalance analysis of *SLFN5* (D) and *CLEC4F* (E) in resting, LPS-stimulated, FLU-infected, and IFN- β -stimulated MoDCs, showing the ratio of gene expression between the major and

minor alleles in heterozygote (rs11651240 for *SLFN5*, rs2075221 for *CLEC4F*) cDNA samples [$n = 8$ in (D); $n = 9$ in (E)] normalized to the ratio in the corresponding genomic DNA samples; significant deviation from 1.0 (dashed line) is consistent with allelic imbalance. Data are from one experiment representative of three (mean and SD shown). $*P < 0.01$, $**P < 0.001$, compared to unstimulated cells (Student's t test). On the right panels, box-whisker plots showing expression [left; $\log_2(\text{nCounts})$, y axis] or fold change [right; $\log_2(\text{fold})$, y axis] of *SLFN5* (D) and *CLEC4F* (E) in resting, LPS-stimulated, FLU-infected, and IFN- β -stimulated MoDCs as a function of genotype of the respective cis-SNPs: rs11867191 and rs2075221. (F) Schematic showing the different combinations of stimuli leading to significant cis-reQTLs, with the most significant examples listed. Specificity to conditions was defined with M value >0.9 taken as the inclusion criteria and M value <0.1 taken as the exclusion criteria for each condition.

disequilibrium with the most significant reQTL SNP (rs11867191, $R^2 = 0.501$ CEPH, CEU). As predicted, transcripts derived from the major and minor alleles differed in expression after stimulation with LPS (*SLFN5*: $P < 0.001$, t test; *CLEC4F*: $P < 0.01$); influenza (*SLFN5*: $P < 0.001$; *CLEC4F*: $P < 0.01$); or IFN- β (*SLFN5*: $P < 0.001$; *CLEC4F*: $P < 0.001$) but not at baseline (Fig. 3, D and E).

Cis-reQTLs That Alter the Sequence of TF Binding Sites

We hypothesized that the causal variants underlying cis-reQTLs functionally alter the chromosomal binding sites of TFs activated by one or more stimuli. To fine map reQTLs, we performed a trans-ethnic meta-analysis (28) of imputed variants in each population (~10 M). We then examined whether the most highly associated meta-reQTLs overlap with TF binding sites identified in high-throughput human chromatin immunoprecipitation with massively parallel DNA sequencing (ChIP-seq) data sets (for example, ENCODE) or

computationally predicted conserved regulatory elements (29–31). We found substantial enrichment (table S7) of known binding sites for TFs from the signal transducer and activator of transcription (STAT) family (TFs that are activated downstream of type I IFN signaling). The most significant enrichment over background was found for STAT2 (116-fold, binomial $P < 2.55 \times 10^{-21}$) and STAT1 (126-fold, binomial $P < 2.98 \times 10^{-13}$) binding sites (derived from ChIP-seq in IFN- α -stimulated K562 cells) within cis-eQTLs after IFN- β -stimulation.

Among the 57 genes (for example, *SLFN5*, *CLEC4F*, and *ARL5B*; Fig. 3F) with cis-reQTLs in all three stimuli, we observed that the most significant cis-reQTL in the (i) *SLFN5* locus (rs11080327) lies in an ENCODE ChIP-seq signal (29) for STAT1 (Fig. 4, A to C); (ii) *CLEC4F* locus (rs35856355) alters a commonly occurring cytosine in a canonical IFN-stimulated response element (ISRE) that is a target of IFN-activated TFs (Fig. 4, B and C) (22, 30–32); and (iii)

ARL5B locus (rs2130531) changes a guanine in the canonical ISRE motif (Fig. 4, B and C). Additional cis-reQTLs that alter putative STAT binding sites based on ChIP-seq data or predicted ISRE motifs include rs10086852 (*PTK2B*), rs1981760 (*NOD2*), rs73023464 (*C19ORF12*), rs1331717 (*IFI44*), and rs12064196 (*IFI44*) (Fig. 4A). Because STAT TFs are activated downstream of the type I IFN receptor (IFNAR) and bind to ISRE motifs (Fig. 4A) (22), we hypothesized that the SNPs in these sites are likely causal SNPs within their respective cis-reQTL regions.

Cis-reQTLs That Affect Differential Binding of Stimulus-Activated TFs

To experimentally validate our predicted regulatory mechanisms, we determined whether IFN- β -activated TFs bind differentially at these SNPs. Using radiolabeled 24- to 26-bp (base pair) double-stranded DNA (dsDNA) probes in electrophoretic mobility shift assays (EMSAs), we found

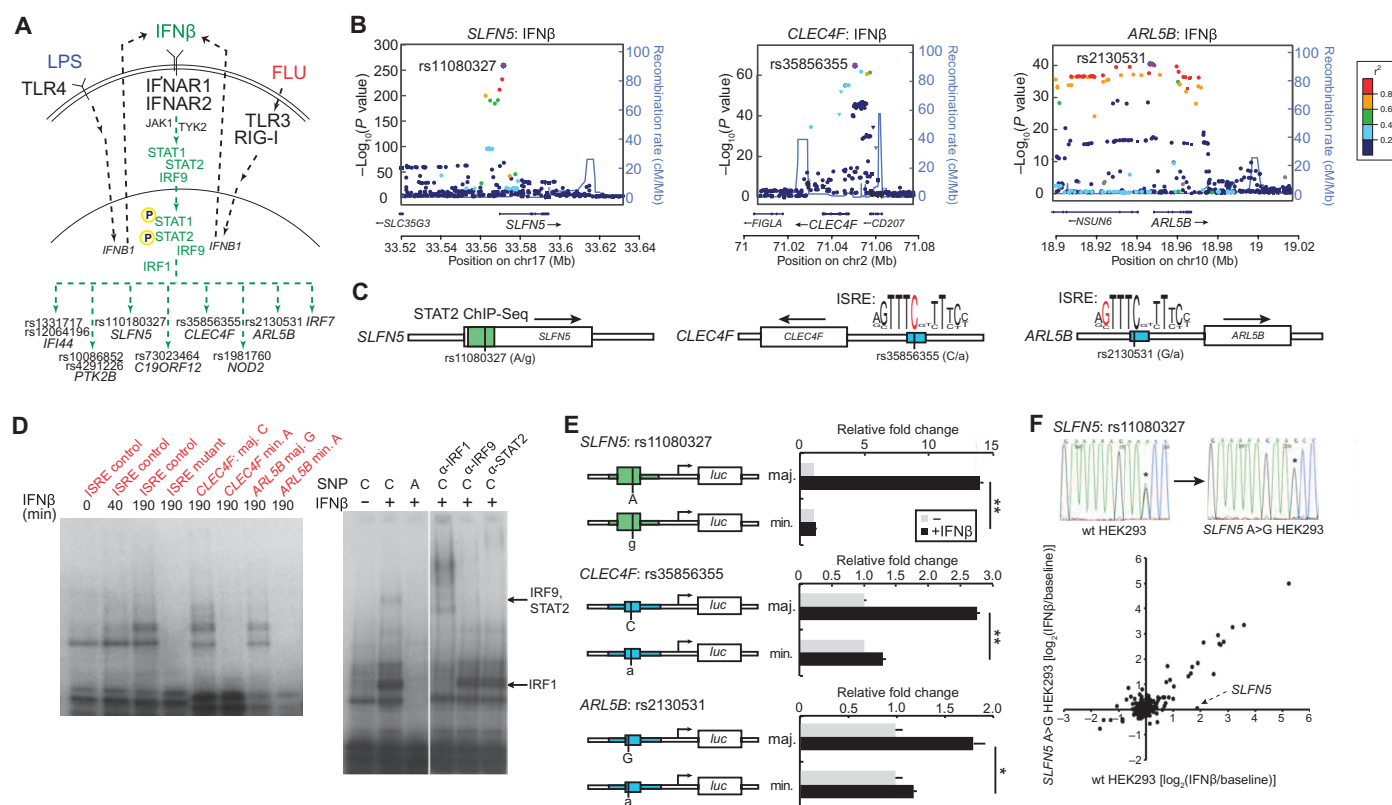


Fig. 4. Functional fine-mapping and mechanism of cis-reQTLs. (A) Pathway diagram of signal transduction cascade downstream of IFNAR activation. Activation of receptor leads to downstream activation of JAK-STAT cascade, leading to posttranslational activation of STAT and IRF TFs. (B) LocusZoom plots showing the $-\log_{10}(P)$ values of imputed cis-eQTLs (y axis) in the chromosomal regions (x axis) of *SLFN5*, *CLEC4F*, and *ARL5B*. The most significant imputed SNPs in each locus are labeled. (C) Schematic representation of alleles in the regions near the *SLFN5*, *CLEC4F*, and *ARL5B* genes that are in STAT2 ChIP-seq binding sites or that perturb ISRE motifs (SNPs are shown as vertical bars and in red letters). (D) EMSAs with 24- to 26-bp radiolabeled dsDNA probes—containing a known ISRE motif control, the *CLEC4F* rs35856355 major (C) or minor allele (A) sequences, or the *ARL5B* rs2130531 major (G) or minor allele (A) sequences—incubated with

nuclear lysates from IFN- β -stimulated MoDCs. On the right, supershift assays with or without antibodies against IRF1, IRF9, and STAT2 (designated α -IRF-1, and so on) with the *CLEC4F* rs35856355 major (C) probe are shown. (E) Luciferase expression from reporter constructs transfected into HEK293 cells that were left unstimulated or were stimulated with IFN- β (1000 U/ml) for 21 hours. Sequences (150 to 200 bp) from the major and minor haplotypes of the *SLFN5*, *CLEC4F*, and *ARL5B* regions were subcloned 5' of a minimal promoter and firefly luciferase gene. Firefly luciferase expression was normalized to *Renilla* luciferase expression expressed from cotransfected plasmid. (F) Fold change $\log_2(\text{IFN-}\beta \text{ stim/unstim})$ of signature genes in wild-type (wt) HEK293 cells (rs11080327^{wt}), plotted against fold change in CRISPR-converted (rs11080327^{ΔG}) HEK293 cells. Data are from one experiment representative of three [mean and SD shown in (E)]. * $P < 0.05$, ** $P < 0.01$, compared to unstimulated cells (Student's t test).

that probes encompassing the major alleles (rs35856355^C and rs2130531^G, which are associated with increased expression in *CLEC4F* and *ARL5B*), but not the minor alleles, shifted after incubation with IFN- β -stimulated MoDC nuclear lysates (Fig. 4D). Consistently, nonradiolabeled major, but not minor, allele probes competed for binding to the IFN- β -stimulated nuclear factors (fig. S4A). By incubating with antibodies to TFs known to bind the ISRE element, we found that IRF1 (IFN-regulatory factor 1) supershifted the low-molecular weight band and STAT2 and IRF9 supershifted the higher-molecular weight band, consistent with the known complex of these latter two proteins (32) (Fig. 4D). These results suggested that the SNPs in *CLEC4F* and *ARL5B* alter the binding of IFN-activated IRF1, STAT2, and IRF9 TFs.

To directly test the functional effects of the SNPs, we cloned 150- to 200-bp genomic regions surrounding rs11080327 (*SLFN5*), rs35856355 (*CLEC4F*), and rs2130531 (*ARL5B*) upstream of

a luciferase reporter driven by a minimal promoter (Fig. 4E). For each region, we created two constructs that differ only at the respective SNPs. Because almost all cell types respond to IFN- β stimulation, we transfected the reporter constructs into human embryonic kidney 293 (HEK293) cells and stimulated the cells with IFN- β . Consistent with our hypothesis, the constructs containing the major alleles of rs11080327 (*SLFN5*), rs35856355 (*CLEC4F*), and rs2130531 (*ARL5B*) induced 11.6-, 2.1-, and 1.5-fold, respectively, more luciferase production than the cognate constructs containing the minor alleles (*t* test, $P < 0.001$, $P < 0.001$, and $P < 0.01$, respectively) (Fig. 4E). Furthermore, mutation of motif-conserving nucleotides across the *CLEC4F* ISRE motif reduced induction by IFN- β , whereas mutation of nonconserved nucleotides did not (fig. S4B), consistent with the ISRE consensus motif.

Finally, to further test (33) the functionality of rs11080327—the *SLFN5* variant with the strongest functional effect—in its native chromo-

some, we directly edited the genome using the CRISPR system (34, 35), converting the heterozygous rs11080327^{A/G} in HEK293 cells to a homozygous rs11080327^{G/G} (Fig. 4F and fig. S4C). We then stimulated both native and CRISPR-converted cells with IFN- β . The fold induction of *SLFN5* decreased from 3.64 in heterozygous cells to 1.03 in the converted rs11080327 (G) homozygotes, whereas the induction of other genes was unaffected (Fig. 4F and fig. S4C, independent CRISPR construct).

The enrichment, EMSA, reporter assays, and directed mutagenesis results suggest that we have identified causal variants in some of these cis-eQTL regions and demonstrate that the reQTLs occur because of differential binding of stimulus-activated TFs.

A Trans-eQTL That Also Associates with IRF7 as a Cis-eQTL

We next attempted to detect cases where variation in gene expression is explained by distant

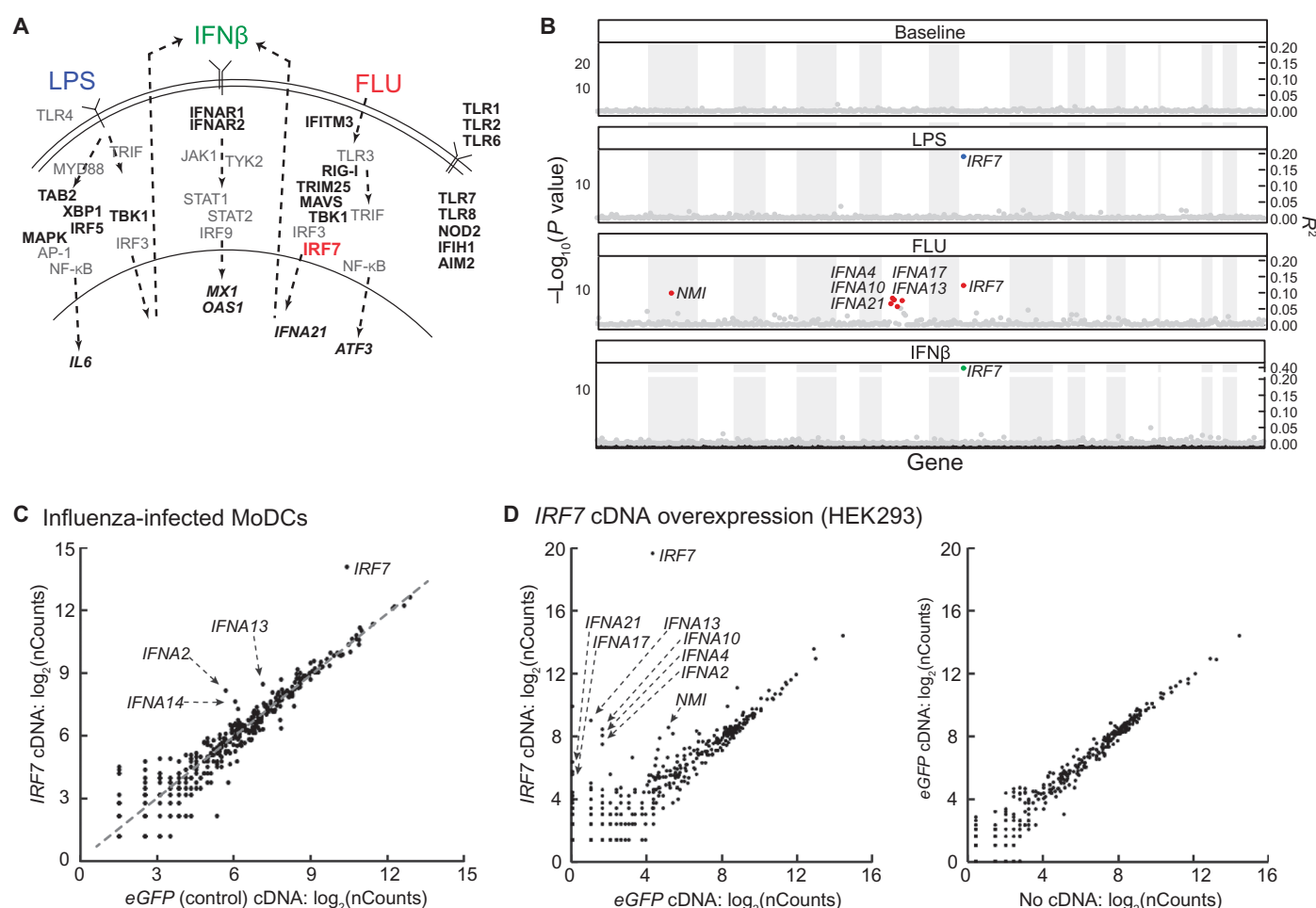


Fig. 5. Trans-reQTL association at the *IRF7* cis-regulatory locus. (A) Diagram showing selected components of TLR4, TLR3, RIG-I, and IFNAR pathways. Components with significant cis-eQTLs (permutation FDR < 0.05) are shown in black (or red if they also have a trans-eQTL); components that do not have significant cis-eQTLs are gray. (B) Manhattan plot showing the trans-association of rs12805435 to all 415 genes on signature gene set in baseline, LPS-stimulated, FLU-infected, and IFN- β -stimulated conditions. Trans-reQTL

to *NMI* and cluster of *IFN- α* genes (*IFNA4*, *IFNA10*, *IFNA13*, *IFNA17*, and *IFNA21*) are labeled. (C) Expression [$\log_2(\text{nCounts})$] of 415 signature genes in FLU-infected MoDCs overexpressing *IRF7*, plotted against expression in FLU-infected MoDCs overexpressing *eGFP* control. (D) Expression [$\log_2(\text{nCounts})$] of 415 signature genes in HEK293 cells overexpressing *IRF7*, plotted against expression in HEK293 cells overexpressing *eGFP* control. Right, expression of genes in cells with *eGFP* overexpression versus cells without cDNA.

genetic variants acting in trans. We noted that cis-eQTLs for regulator genes tend to have smaller effect sizes and less significant P values than those for regulated genes (fig. S3E). To decrease the multiple testing burden of detecting trans-associations, we restricted testing SNPs local to genes on our gene signature set. A number of these are cis-eQTLs associated with the expression of genes encoding regulators (for example, TFs) in pathways that we stimulated (Fig. 5A). Of these, rs12805435 was the most significant, associating with the expression of the master antiviral TF *IRF7* in cis only after influenza, LPS, or IFN- β stimulation (table S4, A to D). This SNP associated with both the expression and induction of seven additional genes in trans (that is, the transcriptional start site or stop codon is located >1 Mb from the SNP) only after influenza infection: *NMI*, *IFNA4*, *IFNA10*, *IFNA13*, *IFNA21*, *IFNA17*, and *IFNA5* ($P < 6.26 \times 10^{-5}$ in expression, $P < 1.68 \times 10^{-5}$ in differential expression) (Fig. 5B and tables S8 to 10).

To test whether the trans-associated genes are indeed targets of IRF7, we infected MoDCs

that overexpressed *IRF7* or a control gene with influenza virus, and found that *IRF7* overexpression induced the expression of *IFNA2*, *IFNA13*, and *IFNA14* in influenza-infected cells (Fig. 5C). In addition, in HEK293 cells, which normally express very low levels of *IRF7*, we overexpressed *IRF7* and observed induction of *NMI*, *IFNA4*, *IFNA10*, *IFNA13*, *IFNA14*, and *IFNA21* (Fig. 5D), suggesting that IRF7 is sufficient to drive downstream expression. We have thus identified a stimulus-specific cis-eQTL associated with *IRF7* expression, which is also a trans-reQTL that underlies the variability of IFN induction in response to influenza infection in humans.

DC reQTLs That Overlap with Autoimmune and Infectious Disease SNPs from GWAS

We examined the subset of loci associated in GWAS with inflammatory disorders. We first extended an analysis of Crohn's disease in which loci are enriched for genes specifically expressed in DCs (12, 14) to multiple sclerosis, celiac disease, psoriasis, and leprosy (fig. S5A). We found that genes nearest to the susceptibility loci for

these diseases were enriched not only in DC-specific genes but also in genes induced by LPS and/or influenza (fig. S5B), which suggests that some of the GWAS loci modulate the expression of the corresponding genes in activated DCs.

Supporting this hypothesis, 15 cis-reQTLs and 23 cis-eQTLs that were not significant at baseline but only became significant after stimulation are the same SNPs previously identified in GWAS of autoimmune and infectious diseases, including Crohn's disease, multiple sclerosis, celiac disease, psoriasis, and leprosy (25) (Fig. 6, A and B, and table S11). These include *NOD2* with leprosy (rs9302752), *IRF7* with systemic lupus erythematosus (SLE) (rs4963128), *TRAF1* with rheumatoid arthritis (rs881375), and *CREM* with Crohn's disease (rs12242110) and ulcerative colitis (rs4246905). For example, rs9302752—a SNP previously associated with susceptibility to leprosy (17)—was associated in our study with the absolute and fold expression of *NOD2* under IFN- β stimulation ($P = 3.49 \times 10^{-25}$) but not at baseline (Fig. 6A and table S11); *NOD2* plays a known role in pathogen sensing and possibly mycobacterial

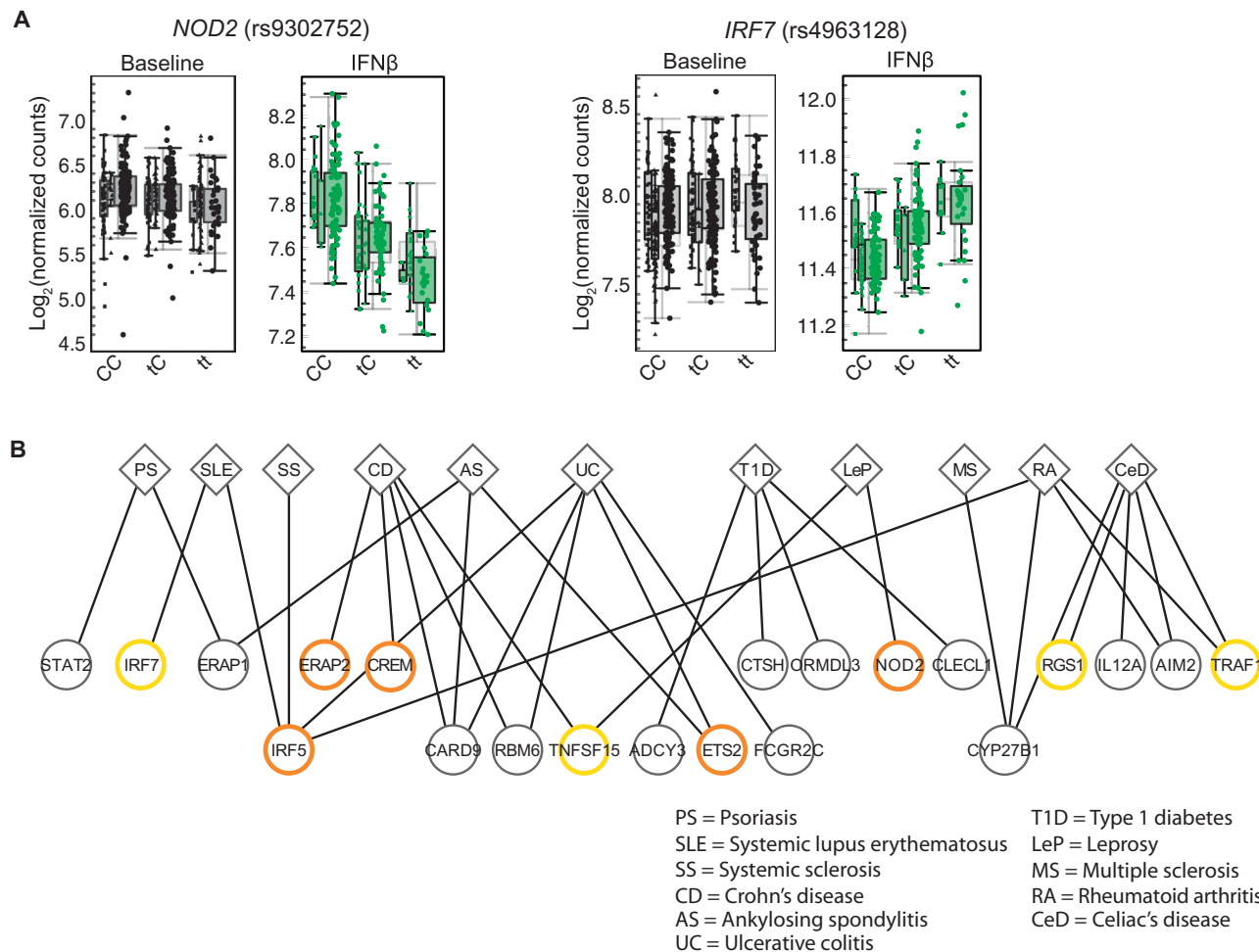


Fig. 6. Autoimmune and infectious disease-associated SNPs are cis-eQTLs and cis-reQTLs. (A) Expression [$\log_2(\text{nCounts})$] of *NOD2* in resting and IFN- β -stimulated MoDCs from 184 Caucasians as a function of genotype of the leprosy GWAS SNP, rs9302752 (left). (Right) expression [$\log_2(\text{nCounts})$] of *IRF7* in resting and IFN- β -stimulated MoDCs from 184 Caucasians as a

function of genotype of the SLE GWAS SNP, rs4963128. (B) Plot showing overlap of genome-wide significant ($P < 5 \times 10^{-8}$) GWAS SNPs with cis-eQTLs and reQTLs in MoDCs, with clinical phenotypes connected to corresponding gene expression phenotypes by lines. Orange circles represent cis-reQTLs ($P < 10^{-7}$); yellow circles represent stimulus-specific cis-eQTLs ($P < 10^{-7}$).

immunity (36). Similarly, rs4963128—a variant associated with SLE (37)—was associated in our study with the expression of *IRF7* after IFN- β stimulation ($P = 1.10 \times 10^{-16}$) but not at baseline (Fig. 6A), in line with the importance of type I IFN responses in SLE pathogenesis (38). We note that rs4963128 is on the same haplotype ($R^2 = 0.69$, $D' = 0.94$) as the *IRF7* SNP rs12805435 that is associated with the trans-eQTL effect described above. These results suggest a role for innate immune pathogen-sensing pathways in the pathogenesis of these inflammatory disorders.

Discussion

Although genetic association studies have identified alleles that confer disease risk, little is known about how these genetic variants contribute to disease through their effects on specific biological processes and their interaction with environmental stimuli. We addressed this question by quantifying the DC response to pathogens in a set of genotyped individuals, and then leveraged our understanding of these pathways to explain the mechanisms underlying the observed genotype-environment-phenotype interactions.

Many of the QTL associations we identified are only detectable in the presence of specific stimuli, which underscores the need to activate cells to capture additional genotype-phenotype relations (39–41). Furthermore, few eQTL studies have definitively identified the causal variants underlying the associations. By measuring variability in hundreds of individuals, applying stimuli that partially overlap in their downstream pathways, and leveraging genomic data sets such as the 1000 Genomes Project and ENCODE (29, 42), we (i) pinpointed causal variants in reQTL regions and (ii) identified the stimulus-activated TFs that bind differentially at these SNPs, which explain some of the $G \times E$ interactions. Our data set can thus be used to explore mechanisms of $G \times E$ interactions (43) and are consistent with deoxyribonuclease I sensitivity QTL (44) and ChIP-seq QTL (45) studies that showed that differential TF binding between individuals is pervasive in resting cells.

The reQTLs we identified provide genetic explanations for interindividual variation in innate immune responses. This is best exemplified by the trans-reQTL in the *IRF7* locus that regulates the type I IFN antiviral response. Our study reveals the effects of this trans-reQTL on target genes (antiviral IFN module) in the context of a particular cell type (DCs) and in response to specific ligands (influenza). The changes in this immune response are, in turn, likely to affect organismal phenotypes that are driven by the IFN module, including susceptibility to viral infections and autoimmune diseases like SLE.

Overall, our high-throughput experimental pipeline and integrative analysis of primary human DCs reveals abundant gene-by-environment interactions, points to the effects of disease variants on pathogen detection, and motivates

extending our approach to a wide variety of immune cell types and stimulatory conditions to better explain the impact of human genetic variation on the immune response.

References and Notes

- M. Civelek, A. J. Lusis, Systems genetics approaches to understand complex traits. *Nat. Rev. Genet.* **15**, 34–48 (2014). doi: [10.1038/nrg3575](#); pmid: [24296534](#)
- D. Thomas, Gene-environment-wide association studies: Emerging approaches. *Nat. Rev. Genet.* **11**, 259–272 (2010). doi: [10.1038/nrg2764](#); pmid: [20212493](#)
- E. E. Eichler *et al.*, Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* **11**, 446–450 (2010). doi: [10.1038/nrg2809](#); pmid: [20479774](#)
- B. E. Stranger *et al.*, Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224 (2007). doi: [10.1038/ng2142](#); pmid: [17873874](#)
- E. E. Schadt *et al.*, Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297–302 (2003). doi: [10.1038/nature01434](#); pmid: [12646919](#)
- I. Gat-Viks *et al.*, Deciphering molecular circuits from genetic variation underlying transcriptional responsiveness to stimuli. *Nat. Biotechnol.* **31**, 342–349 (2013). doi: [10.1038/nbt.2519](#); pmid: [23503680](#)
- L. D. Orozco *et al.*, Unraveling inflammatory responses using systems genetics and gene-environment interactions in macrophages. *Cell* **151**, 658–670 (2012). doi: [10.1016/j.cell.2012.08.043](#); pmid: [23101632](#)
- E. N. Smith, L. Kruglyak, Gene-environment interaction in yeast gene expression. *PLOS Biol.* **6**, e83 (2008). doi: [10.1371/journal.pbio.0060083](#); pmid: [18416601](#)
- Y. Li *et al.*, Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLOS Genet.* **2**, e222 (2006). doi: [10.1371/journal.pgen.0020222](#); pmid: [17196041](#)
- O. Takeuchi, S. Akira, Pattern recognition receptors and inflammation. *Cell* **140**, 805–820 (2010). doi: [10.1016/j.cell.2010.01.022](#); pmid: [20303872](#)
- E. Rai, E. K. Wakeland, Genetic predisposition to autoimmunity—What have we learned? *Sem. Immunol.* **23**, 67–83 (2011). doi: [10.1016/j.smim.2011.01.015](#); pmid: [21288738](#)
- X. Hu *et al.*, Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am. J. Hum. Genet.* **89**, 496–506 (2011). doi: [10.1016/j.ajhg.2011.09.002](#); pmid: [21963258](#)
- M. Merad, P. Sathe, J. Helft, J. Miller, A. Mortha, The dendritic cell lineage: Ontogeny and function of dendritic cells and their subsets in the steady state and the inflamed setting. *Annu. Rev. Immunol.* **31**, 563–604 (2013). doi: [10.1146/annurev-immunol-020711-074950](#); pmid: [23516985](#)
- L. Jostins *et al.*, Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012). doi: [10.1038/nature11582](#); pmid: [23128233](#)
- L. C. Tsoi *et al.*, Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat. Genet.* **44**, 1341–1348 (2012). doi: [10.1038/ng.2467](#); pmid: [23143594](#)
- D. L. Thomas *et al.*, Genetic variation in *IL28B* and spontaneous clearance of hepatitis C virus. *Nature* **461**, 798–801 (2009). doi: [10.1038/nature08463](#); pmid: [19759533](#)
- F. R. Zhang *et al.*, Genomewide association study of leprosy. *N. Engl. J. Med.* **361**, 2609–2618 (2009). doi: [10.1056/NEJMoa0903753](#); pmid: [20018961](#)
- G. K. Geiss *et al.*, Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26**, 317–325 (2008). doi: [10.1038/nbt1385](#); pmid: [18278033](#)
- J. T. Leek, J. D. Storey, Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLOS Genet.* **3**, 1724–1735 (2007). doi: [10.1371/journal.pgen.0030161](#); pmid: [17907809](#)
- T. Kawai, S. Akira, The role of pattern-recognition receptors in innate immunity: Update on Toll-like receptors. *Nat. Immunol.* **11**, 373–384 (2010). doi: [10.1038/ni.1863](#); pmid: [20404851](#)
- R. Barbalat, S. E. Ewald, M. L. Mouchess, G. M. Barton, Nucleic acid recognition by the innate immune system. *Annu. Rev. Immunol.* **29**, 185–214 (2011). doi: [10.1146/annurev-immunol-031210-101340](#); pmid: [21219183](#)
- G. R. Stark, I. M. Kerr, B. R. Williams, R. H. Silverman, R. D. Schreiber, How cells respond to interferons. *Annu. Rev. Biochem.* **67**, 227–264 (1998). doi: [10.1146/annurev.biochem.67.1.227](#); pmid: [9759489](#)
- A. L. Brass *et al.*, The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. *Cell* **139**, 1243–1254 (2009). doi: [10.1016/j.cell.2009.12.017](#); pmid: [20064371](#)
- A. R. Everitt *et al.*, IFITM3 restricts the morbidity and mortality associated with influenza. *Nature* **484**, 519–523 (2012). doi: [10.1038/nature10921](#); pmid: [22446628](#)
- L. A. Hindorff, J. MacArthur, J. Morales, H. A. Junkins, P. N. Hall, A. K. Klemm, T. A. Manolio, A catalog of published genome-wide association studies; <http://www.genome.gov/gwastudies> [accessed 1 June 2013].
- Y. Donner, T. Feng, C. Benoist, D. Koller, Imputing gene expression from selectively reduced probe sets. *Nat. Methods* **9**, 1120–1125 (2012). doi: [10.1038/nmeth.2207](#); pmid: [23064520](#)
- Supplementary text is available on Science Online.
- B. Han, E. Eskin, Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011). doi: [10.1016/j.ajhg.2011.04.014](#); pmid: [21565292](#)
- ENCODE Project Consortium *et al.*, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi: [10.1038/nature11247](#); pmid: [22955616](#)
- L. D. Ward, M. Kellis, HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012). doi: [10.1093/nar/gkr917](#); pmid: [22064851](#)
- A. P. Boyle *et al.*, Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012). doi: [10.1101/gr.137323.112](#); pmid: [22955989](#)
- K. Honda, T. Taniguchi, IRFs: Master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. *Nat. Rev. Immunol.* **6**, 644–658 (2006). doi: [10.1038/nri1900](#); pmid: [16932750](#)
- T. F. Mackay, E. A. Stone, J. F. Ayroles, The genetics of quantitative traits: Challenges and prospects. *Nat. Rev. Genet.* **10**, 565–577 (2009). doi: [10.1038/nrg2612](#); pmid: [19584810](#)
- L. Cong *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013). doi: [10.1126/science.1231143](#); pmid: [23287718](#)
- P. Mali *et al.*, RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013). doi: [10.1126/science.1232033](#); pmid: [23287722](#)
- M. Divangahi *et al.*, NOD2-deficient mice have impaired resistance to *Mycobacterium tuberculosis* infection through defective innate and adaptive immunity. *J. Immunol.* **181**, 7157–7165 (2008). pmid: [18981137](#)
- International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN) *et al.*, Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in *ITGAM*, *PXK*, *KIAA1542* and other loci. *Nat. Genet.* **40**, 204–210 (2008). doi: [10.1038/ng.81](#); pmid: [18204446](#)
- V. Pascual, D. Chaussabel, J. Banchereau, A genomic approach to human autoimmune diseases. *Annu. Rev. Immunol.* **28**, 535–571 (2010). doi: [10.1146/annurev-immunol-030409-101221](#); pmid: [20192809](#)
- D. A. Smirnov, M. Morley, E. Shin, R. S. Spielman, V. G. Cheung, Genetic analysis of radiation-induced changes in human gene expression. *Nature* **459**, 587–591 (2009). pmid: [19349959](#)

40. C. E. Romanoski *et al.*, Systems genetics analysis of gene-by-environment interactions in human cells. *Am. J. Hum. Genet.* **86**, 399–410 (2010). doi: [10.1016/j.ajhg.2010.02.002](https://doi.org/10.1016/j.ajhg.2010.02.002); pmid: [20170901](https://pubmed.ncbi.nlm.nih.gov/20170901/)
41. L. B. Barreiro *et al.*, Deciphering the genetic architecture of variation in the immune response to *Mycobacterium tuberculosis* infection. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1204–1209 (2012). doi: [10.1073/pnas.1115761109](https://doi.org/10.1073/pnas.1115761109); pmid: [22233810](https://pubmed.ncbi.nlm.nih.gov/22233810/)
42. 1000 Genomes Project Consortium *et al.*, An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012). doi: [10.1038/nature11632](https://doi.org/10.1038/nature11632); pmid: [23128226](https://pubmed.ncbi.nlm.nih.gov/23128226/)
43. H. Stower, Gene regulation: From genetic variation to phenotype via chromatin. *Nat. Rev. Genet.* **14**, 824 (2013). doi: [10.1038/nrg3622](https://doi.org/10.1038/nrg3622); pmid: [24166029](https://pubmed.ncbi.nlm.nih.gov/24166029/)
44. J. F. Degner *et al.*, DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**, 390 (2012). doi: [10.1038/nature10808](https://doi.org/10.1038/nature10808); pmid: [22307276](https://pubmed.ncbi.nlm.nih.gov/22307276/)
45. M. Kasowski *et al.*, Variation in transcription factor binding among humans. *Science* **328**, 232–235 (2010). doi: [10.1126/science.1183621](https://doi.org/10.1126/science.1183621); pmid: [20299548](https://pubmed.ncbi.nlm.nih.gov/20299548/)

Acknowledgments: We thank all study participants in the PhenoGenetic project for their contributions. We are grateful to Q. Sievers and C. Wu for advice about preparing primary cells and V. Agarwala and D. A. Landau for valuable discussions. This work was supported by the National Institute of General Medical Sciences, NIH, grant RC2 GM093080 (C.O.B. and N.H.); the National Institute of Allergy and Infectious Diseases grant U19 AI082630 and NIH Director's New Innovator Award DP2 OD002230 (N.H.); the National Human Genome Research Institute grant P50 HG006193, NIH Director's Pioneer Award DP1 CA174427, and Howard Hughes Medical Institute (A.R.); and NIH grant R01 HG004037 (M.K.). F.Z. is supported by an NIH Director's Pioneer Award (DP1 MH100706), an NIH Transformative R01 grant (R01 DK097768), the Keck, McKnight, Merkin, Vallee, Damon Runyon, Searle Scholars,

Klingenstein, and Simons Foundations, Bob Metcalfe, and Jane Pauley. M.N.L. is supported by the NIH Medical Scientist Training Program fellowship (T32 GM007753). T.R. is supported by NIH fellowship F32 AG043267. K.S. is supported by NIH grant T32 HG002295. Gene expression data are deposited in National Center for Biotechnology Information's Gene Expression Omnibus under accession no. GSE53166.

Supplementary Materials

www.sciencemag.org/content/343/6175/1246980/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S5
Tables S1 to S11
References (46–72)

8 October 2013; accepted 10 January 2014
[10.1126/science.1246980](https://doi.org/10.1126/science.1246980)