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

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1126/science.1246980">http://dx.doi.org/10.1126/science.1246980</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>American Association for the Advancement of Science (AAAS)</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Mon Jan 07 09:00:22 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/96695">http://hdl.handle.net/1721.1/96695</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a></td>
</tr>
</tbody>
</table>
Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells

Mark N. Lee1,2,3,†, Chun Ye1,†, Alexandra-Chloé Villani1,2, Towfique Raj1,2,4, Weibo Li1,3, Thomas M. Eisenhaure1,3, Selina H. Imboywa2, Portia I. Chipendo2, F. Ann Ran1,5, Kamil Slowikowski6, Lucas D. Ward1,7, Khadir Raddassi8, Cristin McCabe1,2, Michelle H. Lee2, Irene Y. Frohlich2, David A. Hafler8, Manolis Kellis1,7, Soumya Raychaudhuri1,2,9,10, Feng Zhang11,12,13, Barbara E. Stranger14,15, Christophe O. Benoist2, Philip L. De Jager1,2,4, Aviv Regev1,16,17,* and Nir Hacohen1,2,3,*

1Broad Institute of 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 & Women's Hospital, Boston MA 02115, USA
5Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA
6Bioinformatics and Integrative Genomics, Graduate School of Arts and Sciences, Harvard University, Cambridge, MA 02138, USA
7Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
8Department of Neurology, Yale School of Medicine, CT 06511, USA
9Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA
10Arthritis Research UK Epidemiology Unit, Musculoskeletal Research Group, University of Manchester, Manchester Academic Health Sciences Centre, UK
11McGovern Institute for Brain Research, MIT, Cambridge, MA, 02139, USA
12Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, 02139, USA
13Department of Biological Engineering, MIT, Cambridge, MA, 02139, USA
14Section of Genetic Medicine, The University of Chicago, Chicago, IL 60637, USA

†Correspondence to: aregev@broad.mit.edu, nhacohen@partners.org.
These authors contributed equally to this work.

Supplementary Materials: Materials and Methods
Supplementary Text
Figs. S1 to S5
Tables S1 to S11
References (46-72)
Abstract

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 E. coli lipopolysaccharide, influenza or interferon-β (IFNβ). We localized and validated causal variants to binding sites of pathogen-activated STAT and IRF transcription factors. We also identified a common variant in IRF7 that is associated in trans with type I interferon induction in response to influenza infection. Our results reveal common alleles that explain inter-individual variation in pathogen sensing and provide functional annotation for genetic variants that alter susceptibility to inflammatory diseases.

Introduction

Susceptibility to complex diseases depends on both genetic predisposition and exposure to environmental factors, with interactions between the two (GxE 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 GxE 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), thus excluding GxE 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 Toll-like receptors (e.g., TLR3) and the RIG-I-like receptors (e.g., RIG-I) (10). These, in turn, lead to the translocation of transcription factors from the cytoplasm into the nucleus to induce the expression of immune genes, including IFNβ secretion that engages the type I interferon response pathway to induce the expression...
of hundreds of anti-viral 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+CD16lo monocytes from human blood samples (figs. S1A–S1F), differentiate them into monocyte-derived dendritic cells (MoDCs), and stimulate them with three immunostimulatory agents (Fig. 1B): E. coli lipopolysaccharide (LPS), influenza virus, or IFNβ (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 1,598 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, 117 East Asians; table S1) in four states: resting, LPS-stimulated, influenza-infected and IFNβ-stimulated. Finally, we associated expression variation with genetic variation to identify cis- and trans-eQTLs and reQTLs, and investigated candidate reQTLs for their mechanism of action.

Variability of pathogen-sensing responses between individuals

To assess the inter-individual 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, 6 African-Americans). We found 1,413 genes that were regulated in LPS- or influenza-treated cells [\(\log_2(\text{fold change}) > 0.75 \text{ or } -1.5; FDR < 0.01\); fig. S2A, table S2A; see table S2B–I for enriched biological processes] at 5 hrs and 10 hrs, respectively (time points selected to have maximal expression of induced clusters; figs. S1B and S1C).

We quantified reproducibility in these responses by recalling 12 of the 30 donors 2-9 months after the first collection for MoDC isolations and profiling (table S1); we identified genes whose inter-individual expression variance is significantly higher than their intra-individual variance based on the serial replicates (taking into account known covariates including gender, age and population and unknown ones using Surrogate Variable Analysis (SVA) (19)). 222 of the 1,413 regulated genes (16%) showed significantly higher (FDR < 0.1) inter- than intra-individual 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.

Science. Author manuscript; available in PMC 2014 September 07.
Expression profiling of a pathogen response signature

Mapping the genetic basis of inter-individual 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 interferon 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: (1) all 222 of 1,413 regulated genes with greater inter- than intra-individual variability in the microarray study; (2) all 24 regulated genes exhibiting greater inter- than intra-population variability (FDR < 0.1; table S3) in the microarray study; (3) 76 genes comprising the known components of the TLR4, TLR3, RIG-I and IFNAR pathways (Fig. 1B) (10, 20-22); (4) 61 regulated genes that play key roles in the DC response (e.g. IFNB1 and IFITM3) (22-24); (5) 28 regulated genes residing in loci previously associated with autoimmune or infectious diseases (25); and (6) 35 control genes including those that had among the lowest inter-individual variance in the microarray profiles. This representative signature allowed us to monitor genes with high inter-individual variability, as well as key components and responses of the pathogen-sensing pathway.

We then generated 1,598 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 IFNβ-stimulated (284) (Figs. 2C, S1E, S1F, table S1). The IFNβ 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 NF-κB or AP-1 activation (Figs. 2C and 1B). 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 gtPCC; fig. S2D (26)), demonstrating the utility of the signature to capture the genome-wide response.

reQTLs modulate cellular responses to pathogens

We mapped cis-eQTLs and cis-reQTLs by testing for association between common SNPs (minor allele frequency >5%; genotyped with Illumina Human OmniExpress 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 1Mb 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, figs. S3A–C, table S4A–D) (27). Notably, 22/264 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), reflecting the complexity of the regulatory landscape around each gene.
We detected 121 cis-reQTLs (Fig. 3B and table S4E–G; permutation FDR < 0.05; 91% internal reproducibility in at least 2 human populations, table S5), the subset of genetic variants that affect the induction of gene expression by LPS, influenza or IFNβ. Of the 121 genes with cis-reQTLs, 7 associations were found only in the influenza condition (e.g., IFNA21, Fig. 3C and fig. S3B; meta-analysis, $P < 1 \times 10^{-5}$, permutation FDR 0.02-0.03), 15 in both the LPS and influenza conditions (e.g., TEC; Figs. 3C and S3B), and 57 in all three stimulation conditions (e.g. ARL5B, SLFN5, CLEC4F;Figs. 3C–3E and S3B), likely reflecting activation of the shared IFNβ pathway. We hypothesized that causal variants driving cis-reQTLs alter the sequence of genomic elements that respond to transcription factors 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 (Figs. 3D and 3E). 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 disequilibrium (LD) with the most significant reQTL SNP (rs11867191, $R^2 = 0.501$ 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 IFNβ (SLFN5, $P < 0.001$; CLEC4F, $P < 0.001$) but not at baseline (Figs. 3D and 3E).

**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 transcription factors activated by one or more stimuli. To fine map reQTLs, we performed a trans-ethnic meta-analysis (28) of imputed variants in each population (~10M). We then examined whether the most highly associated meta-reQTLs overlap with transcription factor (TF)binding sites identified in high-throughput human ChIP-Seq datasets (e.g., ENCODE) or computationally-predicted conserved regulatory elements (29-31). We found substantial enrichment (table S7) of known binding sites for TFs from the STAT family (TFs that are known to be 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 IFNα-stimulated K562 cells) within cis-eQTLs after IFNβ stimulation.

Among the 57 genes (e.g., SLFN5, CLEC4F, ARL5B; Fig. 3F) with cis-reQTLs in all 3 stimuli, we observed that the most significant cis-reQTL in the: (1) SLFN5 locus (rs11080327) lies in an ENCODE ChIP-Seq signal (29) for STAT1 (Figs. 4A–C); (2) CLEC4F locus (rs35856355) alters a commonly occurring cytosine in a canonical ISRE (interferon-stimulated response element) that is a target of IFN-activated transcription factors (Figs. 4B and 4C) (22, 30-32); and (3) ARL5B locus (rs2130531) changes a guanine in the canonical ISRE motif (Figs. 4B and 4C). 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 transcription factors are activated downstream of the type
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 affect differential binding of stimulus-activated transcription factors**

To experimentally validate our predicted regulatory mechanisms, we determined whether IFNβ-activated transcription factors bind differentially at these SNPs. Using radiolabeled 24-26 bp dsDNA probes in electrophoretic mobility shift assays (EMSA), we found that probes encompassing the major alleles (rs35856355C and rs2130531G, which are associated with increased expression in *CLEC4F* and *ARL5B*), but not the minor alleles, shifted after incubation with IFNβ-stimulated MoDC nuclear lysates (Fig. 4D). Consistently, non-radiolabeled major but not minor allele probes competed for binding to the IFNβ-stimulated nuclear factors (fig. S4A). By incubating with antibodies to transcription factors known to bind the ISRE element, we found that IRF1 super-shifted the low MW band and STAT2 and IRF9 super-shifted the higher MW 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 transcription factors.

To directly test the functional effects of the SNPs, we cloned 150-200 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. Since almost all cell types respond to IFNβ stimulation, we transfected the reporter constructs into HEK-293 cells and stimulated the cells with IFNβ. Consistent with our hypothesis, the constructs containing the major alleles of rs11080327 (*SLFN5*), rs35856355 (*CLEC4F*) and rs2130531 (*ARL5B*) induced respectively 11.6-fold, 2.1-fold and 1.5-fold 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 IFNβ while mutation of non-conserved 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 chromosome, we directly edited the genome using the CRISPR system (34, 35), converting the heterozygous rs11080327A/G in HEK-293 cells to a homozygous rs11080327G/G (Fig. 4F and fig. S4C). We then stimulated both native and CRISPR-converted cells with IFNβ. The fold induction of *SLFN5* decreased from 3.64 in heterozygous cells to 1.03 in the converted rs11080327 (G) homozygotes, while the induction of other genes was unaffected (Fig. 4F and fig. S4C).

The enrichment, EMSA, reporter assays and directed mutagenesis results suggest that we have identified causal variants in some of these *cis*-reQTL regions and demonstrate that the reQTLs occur due to differential binding of stimulus-activated transcription factors.

**A cis-eQTL associating with IRF7 is a trans-reQTL**

We next attempted to detect cases where variation in gene expression is explained by distant 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 \( \text{trans} \) associations, we restricted testing SNPs local to genes on our gene signature set. A number of these are \( \text{cis} \)-eQTLs associated with the expression of genes encoding regulators (e.g., transcription factors) in pathways that we stimulated (Fig. 5A). Of these, rs12805435 was the most significant, associating with the expression of the master anti-viral transcription factor IRF7 in \( \text{cis} \) only after influenza, LPS or IFN\( \beta \) stimulation (table S4A–D). This SNP associated with both the expression and induction of 7 additional genes in \( \text{trans} \) (i.e. 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 table S8–10).

To test whether the \( \text{trans} \)-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 HEK-293 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 \( \text{cis} \)-eQTL associated with IRF7 expression, which is also a \( \text{trans} \)-reQTL that underlies the variability of IFN induction in response to influenza infection in humans.

**reQTLs in DCs overlap with autoimmune and infectious disease SNPs from GWAS**

We examined the subset of loci associated in genome-wide association studies (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 not only enriched in DC-specific genes but also in genes induced by LPS and/or influenza (fig. S5B), suggesting that some of the GWAS loci modulate the expression of the corresponding genes in activated DCs.

Supporting this hypothesis, 15 \( \text{cis} \)-reQTLs and 23 \( \text{cis} \)-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) (Figs. 6A, 6B and table S11). These include: NOD2 with leprosy (rs9302752), IRF7 with systemic lupus erythematosus (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 expression of NOD2 and its response under IFN\( \beta \) 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 immunity (36). Similarly, rs4963128 – a variant associated with SLE (37) – was associated in our study with the expression of IRF7 after IFN\( \beta \) 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 \textit{IRF7} SNP rs12805435 that is associated with the \textit{trans}-reQTL effect described above. These results suggest a role for innate immune pathogen-sensing pathways in the pathogenesis of these inflammatory disorders.

**Discussion**

While 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 dendritic cell 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, underscoring the need to activate cells to capture additional genotype-phenotype relationships (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 datasets such as the 1000 Genomes Project and ENCODE (29, 42), we (i) pinpointed causal variants in reQTL regions and (ii) identified the stimulus-activated transcription factors that bind differentially at these SNPs, explaining some of the GxE interactions. Our dataset can thus be used to explore mechanisms of GxE interactions (43) and are consistent with DNaseI-sensitivity QTL (44) and ChIP-Seq QTL (45) studies that showed that differential transcription factor binding between individuals is pervasive in resting cells.

The reQTLs we identified provide genetic explanations for inter-individual variation in innate immune responses. This is best exemplified by the \textit{trans}-reQTL in the \textit{IRF7} locus that regulates the type I IFN anti-viral response. Our study reveals the effects of this \textit{trans}-reQTL on target genes (anti-viral 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 impact 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 dendritic cells 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.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**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 NCBI's Gene Expression Omnibus under accession no. GSE57542.

References


Science. Author manuscript; available in PMC 2014 September 07.


27. Supplementary text is available on Science Online.


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

(A) Strategy used to identify baseline and response expression quantitative trait loci (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. Lipopolysaccharide (“LPS”) from E. coli engages the TLR4 receptor; interferon-beta (“IFNβ”) engages the heterodimeric IFNAR receptor; influenza A/PR8 (ΔNS1) (“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.
Fig. 2. Genome-wide expression profiles in MoDCs reveal response phenotypes

(A) Coefficient of variation (CV) of gene expression between 30 different donors (“Inter-individual CV”) plotted against CV of expression within 12 serial replicate samples (“Intra-individual CV”) for each differentially regulated (fold change >0.75 or <–1.5) gene following LPS or FLU stimulation. Yellow (up-regulated) and purple (down-regulated) circles represent genes with significant (moderated t test, FDR < 0.1) inter- vs. intra-individual 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- vs. intra-individual variation following 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 codeset: 222 (49%) are regulated genes that showed significant (mixed model variance components test, permutation FDR < 0.1) inter- vs. intra-individual 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 non-expressed 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 intra-population variability. (C) Gene expression heatmap 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 IFNβ. Rows were clustered by k-means clustering of NanoString dataset 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 6 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 IFNβ–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, which are displayed on the x-axis with associated genes ordered by their chromosomal location. (C) Box-whisker plots showing expression (left; $\log_2$(nCounts), y-axis) or fold change (right; $\log_2$(fold), y-axis) of DCBLD1, IFNA21, TEC and ARL5B in resting, LPS-stimulated, FLU-infected and IFNβ–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 $\beta$ 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 IFNβ–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 standard deviation 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 IFNβ–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.
Fig. 4. Functional fine-mapping and mechanism of cis-eQTLs

(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 transcription factors. (B) LocusZoom plots showing the $-\log_{10}(P\text{-values})$ 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) Electrophoretic mobility shift assays (EMSA) with 24- to 26-bp radiolabeled dsDNA probes—containing a known ISRE motif control, a mutated 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 IFNβ–stimulated MoDCs. On the right, supershift assays with or without antibodies against IRF1, IRF9, and STAT2 (designated $\alpha$-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 1000 U/mL IFNβ for 21 h. 150–200 bp sequences 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 $\log_{2}(\text{IFNβ stim/unstim})$ of signature genes in wild-type HEK-293 cells (rs11080327$^{\text{A/G}}$), plotted against fold change in CRISPR-converted (rs11080327$^{\text{G/G}}$) HEK-293 cells. Data are from one experiment representative of three (mean and standard shown in (E)). *$P < 0.05$, **$P < 0.01$, compared to unstimulated cells (Student’s $t$-test).
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 shown in grey. (B) Manhattan plot showing the trans-association of rs12805435 to all 415 genes on signature gene set in baseline, LPS-stimulated, FLU-infected and IFNβ-stimulated conditions. Trans-reQTL to NMI and cluster of IFNα genes (IFNA4, IFNA10, IFNA13, IFNA17 and IFNA21) are labeled. (C) Expression (log2(nCounts)) of 415 signature genes in FLU-infected MoDCs overexpressing IRF7, plotted against expression in FLU-infected MoDCs overexpressing eGFP control. (D) Expression (log2(nCounts)) of 415 signature genes in HEK-293 cells overexpressing IRF7, plotted against expression in HEK-293 cells overexpressing eGFP control. Right, expression of genes in cells with eGFP overexpression vs. cells without cDNA.
Fig. 6. Autoimmune and infectious disease–associated SNPs are cis-eQTLs and cis-reQTLs
(A) Expression (log2(nCounts)) of NOD2 in resting and IFNβ–stimulated MoDCs from 184
Caucasians as a function of genotype of the leprosy GWAS SNP, rs9302752 (left). Right,
expression (log2(nCounts)) of IRF7 in resting and IFNβ–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 < 5×10⁻⁸) 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).