Human genetic diversity regulating the *TLR10/TLR1/TLR6* locus confers increased cytokines in response to *Chlamydia trachomatis*

Alyson B. Barnes,[1] Rachel M. Keener,[1,2] Benjamin H. Schott,[1,2] Liuyang Wang,[1] Raphael H. Valdivia,[1] and Dennis C. Ko[1,2,3,4,*]

Summary

Human genetic diversity can have profound effects on health outcomes upon exposure to infectious agents. For infections with *Chlamydia trachomatis* (*C. trachomatis*), the wide range of genital and ocular disease manifestations are likely influenced by human genetic differences that regulate interactions between *C. trachomatis* and host cells. We leveraged this diversity in cellular responses to demonstrate the importance of variation at the Toll-like receptor 1 (*TLR1*), *TLR6*, and *TLR10* locus to cytokine production in response to *C. trachomatis*. We determined that a single-nucleotide polymorphism (SNP) (rs1057807), located in a region that forms a loop with the *TLR6* promoter, is associated with increased expression of *TLR1, TLR6,* and *TLR10* and secreted levels of ten *C. trachomatis*-induced cytokines. Production of these *C. trachomatis*-induced cytokines is primarily dependent on MyD88 and TLR6 based on experiments using inhibitors, blocking antibodies, RNAi, and protein overexpression. Population genetic analyses further demonstrated that the mean IL-6 response of cells from two European populations were higher than the mean response of cells from three African populations and that this difference was partially attributable to variation in rs1057807 allele frequency. In contrast, a SNP associated with a different pro-inflammatory cytokine (rs2869462 associated with the chemokine CXCL10) exhibited an opposite response, underscoring the complexity of how different genetic variants contribute to an individual's immune response. This multidisciplinary study has identified a long-range chromatin interaction and genetic variation that regulates *TLR6* to broaden our understanding of how human genetic variation affects the *C. trachomatis*-induced immune response.

Introduction

Infections by the obligate intracellular bacterium *Chlamydia trachomatis* (*C. trachomatis*) are a global health problem. Trachoma is the most common infectious cause of blindness,[1] genital tract *C. trachomatis* infections are a leading cause of infertility,[2] and lympho-granuloma venereum strains cause urogenital or anorectal infections primarily in men who have sex with men (MSM).[3] However, there is evidence that clinical manifestations among affected individuals are highly variable. Up to 80% of *C. trachomatis* genital infections in women are asymptomatic, meaning many infections go undiagnosed and untreated. If untreated, *C. trachomatis* infections can lead to pelvic inflammatory disease (PID), ectopic pregnancies, and infertility.[4–6] This diversity in outcomes reflects variation in exposure route, bacterial load, microbiota, *C. trachomatis* strains, and potentially health status of its hosts. In addition, susceptibility to *C. trachomatis* infections and disease outcomes is partially modulated by human genetic variation, as has been reported in several candidate gene studies.[7–18] Polymorphisms in cytokine genes and loci associated with immune responses have been associated with more severe cases of trachoma[12,13] and enhanced pathology and risk

for PID in genital infections.[15,19,20] The impact of genetic diversity is also observed among mouse strains, which vary in the course and outcome of *C. trachomatis* genital tract infection.[21–23] Thus, previous findings in humans and mice point to host genetic variation regulating the risk and severity of *C. trachomatis* infection. However, aside from a single genome-wide association study (GWAS) of scarring trachoma that found no genome-wide significant associations,[16] there are no other reported GWASs of *C. trachomatis* infection and none of *C. trachomatis* genital tract infection.

Studying human genetic variation of infectious disease in a clinical setting is difficult due to confounding variables such as environmental exposure and comorbidities in individuals. We can minimize these limitations through the application of the GWAS framework in a controlled experimental setting. Therefore, we applied a platform for GWASs of cellular host-pathogen traits called Hi-HOST (high-throughput human *in vitro* susceptibility testing)[24–27] to *C. trachomatis* infections.[27] With this platform, we measured variation in cytokine responses to *C. trachomatis* across genetically diverse human lymphoblastoid cell lines (LCLs; EBV (Epstein-Barr Virus)-immortalized B cells). Hi-HOST uses LCLs as a genetic model for the impact of pathogens

on entry, intracellular replication, and immune response. LCLs are transcriptionally similar to antigen-activated B cells[28] and have long been used as a rich, standardized resource for studying human genetic diversity[27,29–36] (allowing for leveraging and comparing published work). While the most relevant cell types in studying *C. trachomatis* are likely the epithelial and immune cells the pathogen interacts with during genital and ocular infection, it is important to note that some *C. trachomatis* serovars (including the L2 strain used in our study) can cause enlargement and inflammation of lymph nodes and have been shown to bind B cells and stimulate proliferation.[37,38] Further, while expression quantitative trait loci (eQTLs) can be cell-type specific and display differences in effect size and even directionality across cell types,[39] many eQTLs are shared across cell types,[40] and findings in LCLs can still be informative for understanding human disease.[41] By applying the Hi-HOST approach to *C. trachomatis* phenotypes in H2P2 (Hi-HOST Phenome Project), we identified ten loci associated with *C. trachomatis*-induced cellular traits at genome-wide significance ($p < 5 \times 10^{-8}$), including four single-nucleotide polymorphisms (SNPs) associated with *C. trachomatis*-induced cytokine levels.[27]

Because a fine-tuned balance of cytokine production is necessary for effective pathogen clearance without excessive inflammation and damage to host tissues, levels of cytokines are a particular aspect of host response that may be crucial to susceptibility and severity. *C. trachomatis* induces the production of pro-inflammatory cytokines including IL-1β, IL-6, IL-8, IFNγ, and TNFα.[42–45] While these cytokines help eradicate infection, a prolonged cytokine response may promote tissue damage.[42] Notably, mouse studies have revealed that delayed or decreased production of innate immune cytokines correlate with a longer course of infection and severe hydrosalpinx, a condition in which the oviduct fills with fluid following infection.[46] These results demonstrate that regulation of inflammation can have dramatic impacts on *C. trachomatis* infection progression. Thus, there is a critical need for understanding variation in the inflammatory cytokine response to *C. trachomatis*.

Here, we demonstrate that the rs1057807 locus regulates expression of the nearby pattern-recognition receptors, Toll-like receptors 1, 6, and 10 (*TLR10* [MIM: 606270]/ *TLR1* [MIM: 601194]/*TLR6* [MIM: 605403]). Specifically, a region with SNPs in high linkage disequilibrium with rs1057807 located in the Replication Factor C, Subunit 1 gene (*RFC1* [MIM: 102579]) forms a chromatin loop to the *TLR6* promoter. Blocking antibodies, RNAi, and overexpression confirmed *C. trachomatis*-induced cytokines were strongly dependent on TLR6, with less dependence on TLR1. Population genetic analyses with *C. trachomatis*-induced cytokine production demonstrated that variation in allele frequency of rs1057807 contributes to differences in cytokine levels among populations. The A-allele, associated with a more pro-inflammatory response, demonstrated evidence of positive selection in broadly East Asian populations.

Material and methods

QQ plots, LocusZoom, and Manhattan plots
QQ plots were plotted using the quantile-quantile function in R. Regional Manhattan plots (LocusZoom) were made using LocusZoom.[47] Manhattan plots were generated using the qqman R package.[48]

Human cells
LCLs (EBV-immortalized B cells) were from the Coriell Institute. The Duke institutional review board (IRB) evaluated our studies using Coriell LCLs as meeting the definition of research not involving human subjects (Pro00044583). HeLa cells were from the Duke Cell Culture Facility (originally from ATCC). LCLs were maintained at 37°C in a 5% CO_2 atmosphere and were grown in RPMI 1640 media (Invitrogen) supplemented with 10% fetal bovine serum (FBS), 2 mM glutamine, 100 U/mL penicillin-G, and 100 mg/mL streptomycin (pen-strep). HeLa cells were grown in DMEM supplemented with 10% FBS, 100 U/mL penicillin-G, and 100 mg/mL streptomycin. Infection assays were carried out in the same media but without pen-strep and phenol red. Cells were verified as mycoplasma free by the Universal Mycoplasma Detection Kit (ATCC).

C. trachomatis strains and infections
C. trachomatis L2-GFP[49] was a gift from the Derre Lab and the Valdivia lab. Elementary bodies (EBs) were purified on Omnipaque-350 gradients as previously published.[50] Each preparation was titered by counting inclusion formation units by microscopy on monolayers of Vero cells. *C. trachomatis* was diluted in RPMI with 10% FBS and added at either MOI 0.5 for HeLa cells or MOI 5 for LCLs in 100 μL in 96-well plates as indicated in text. Cells were mixed via pipetting and centrifuged onto cells at 1,500 × *g* for 30 min at 4°C. Infected cells were then incubated at 37°C. For suspension cells, at 24, 48, and 72 h, cells were mixed and 15 μL of cells were used for flow cytometry quantification of *C. trachomatis* infection levels. At 72 h, supernatant was collected to measure cytokine production by either Luminex assay for 17 human cytokines (Millipore) or enzyme-linked immunosorbent assay (ELISA) for [IFNα2] or [IL-6] (R&D).

Cytokine measurements
Cytokine concentration was measured by Luminex (Millipore 17-plex) or ELISA (R&D DuoSet ELISA Development Systems DY206 and DY9345) as indicated in the text. The ELISA detection limits were 9.38 pg/mL for IL-6 and 3.13 pg/mL for IFNα2. The custom Luminex assay obtained from Millipore detected 17 cytokines and chemokines: IFNα2, IFNγ, IL10, MDC, IL12p70, IL13, PDGF-AB/BB, IL1Rα, IL4, IL6, IL8, CXCL10, MIP1β, RANTES, TNFα, TNFβ, and VEGF. Statistical tests were performed on $\log_2$-transformed pg/mL values.

3C
Chromosome conformation capture (3C) was performed as described previously.[51] In brief, chromatin from 10,000,000 LCLs was crosslinked with formaldehyde at room temperature for 10 min and quenched with 1 M glycine. Chromatin was washed with PBS to remove residual formaldehyde and lysed. Following lysis, cells were washed with PBS and resuspended in CutSmart buffer. 20% sodium dodecyl sulfate was added to the nuclei and incubated for 1 h at 37°C while shaking at 900 rpm. Following incubation with sodium dodecyl sulfate, 20% Triton

X-100 was added and incubated for 1 h at 37°C while shaking at 900 rpm. Chromatin was digested with the restriction enzyme HindIII-HF (NEB R3104L) at 37°C while shaking at 900 rpm. We extended the digestion to two 2 h digestions followed by an overnight digestion. Restriction enzyme was inactivated with 20% sodium dodecyl sulfate, and digested chromatin was resuspended in T4 DNA ligase buffer with 20% Triton X-100, incubated for 1 h at 37°C, and then ligated with T4 DNA ligase (NEB M0202S). Following ligation, proteinase K was added, and chromatin was de-crosslinked at 65°C overnight. Phenol-chloroform extraction yielded DNA that was assayed by TaqMan qPCR. Digestion efficiency was assayed by SYBR qPCR. Primer sequences are shown in Table S1. Significance was determined by one-way ANOVA with Holm-Šídák's multiple comparisons test against the mean of the peak fragment (474) on log_2-transformed data.

MyD88 inhibitor experiments
50,000 LCLs were plated in a non-TC (tissue culture)-treated 96-well plate in RPMI assay media and recovered for 2 h. MyD88 was inhibited by 2 h pretreatment with 16 μM of MyD88 inhibitor (ST2825, APExBIO #A3840) before *C. trachomatis* L2-GFP infection. Infections were conducted as described above. A panel of cytokines was measured with Luminex. IL-6 cytokine production was measured with ELISA. Significance was determined by t tests performed on log_2-transformed data. Percent MyD88 dependence was calculated as $1 - \left(\frac{MyD88\ infected\ cytokine}{DMSO\ infected\ cytokine} \right) \times 100$.

TLR1, TLR6 blocking experiments
50,000 LCLs were plated in a non-TC-treated 96-well plate in RPMI assay media and recovered for 2 h. TLR blocking experiments were performed by 2 h pretreatment with 10 μg/mL of anti-hTLR1-IgG (InvivoGen mabg-htlr1), anti-hTLR6-IgG (InvivoGen mabg-htlr6), or IgG control (InvivoGen mabg1-ctrlm) before infection with *C. trachomatis* L2-GFP. Infections were conducted as described above. Significance was determined by one-way ANOVA with Dunnett's multiple comparisons test on log_2-transformed data.

HeLa RNAi
HeLa cells at 7,500 in 96-well TC-treated plates were transfected for 48 h with 0.33 μM total small interfering RNA (siRNA) from either non-targeting siGENOME (Horizon) siRNA #5 (NT5; #D-001210-05) or SMARTpool directed against human TLR1 (#M-008086-01), TLR2 (#M-005120-03), TLR6 (#M-005156-01), or TLR10 (#M-008087-01) with Lipofectamine RNAiMax transfection reagent (Thermo Fisher 13778100). Infections were conducted as described above.

For quantifying gene knockdown, HeLa cells at 37,500 in 24-well TC-treated plates were treated for 48 h with 0.33 μM total siRNA as described above. Knockdown levels were measured by Taqman qPCR (Thermo Fisher TLR1 [Hs00413978_m1], TLR2 [Hs00152932_m1], TLR6 [Hs01039989_s1], WDR19 [Hs00228414_m1]).

Each experimental mean was adjusted to the grand mean by multiplying by a constant. Statistical significance was determined by Brown-Forsythe and Welch ANOVA with Dunnett's multiple comparisons test on normalized data comparing all means to the mean of the NT5 group as a control.

TLR ligand experiments
50,000 LCLs were plated in a non-TC-treated 96-well plate in RPMI assay media and recovered for 2 h. Pam$_3$CSK$_4$ (InvivoGen #tlrl-pms) at 1 μg/mL, FSL-1 (InvivoGen #tlrl-fsl) at 1 μg/mL, or *C. trachomatis* L2-GFP was added to the corresponding wells. Plates were spun at 1,500 × *g* for 30 min at 4°C and placed in the incubator for 72 h. When not coupled with a *C. trachomatis* infection, cells were stimulated with 1 μg/mL of the ligands and then incubated for 24 h. IL-6 production was measured by ELISA. IL-6 pg/mL levels were normalized to untreated infected IL-6 production for each cell line. Spearman correlation was performed on log_2-transformed data.

QUANTI-Blue SEAP reporter assay
HEK-Blue hTLR1-TLR2, hTLR2-TLR6, hTLR2, and hTLR2 KO-TLR1-TLR6 cells were obtained from InvivoGen. HEK-Blue cells were maintained in DMEM growth media with 10% FBS and supplemented with HEK-Blue Selection (InvivoGen hb-sel), Normocin (InvivoGen ant-nr-1), and pen-strep. Cells were split with PBS and gentle pipetting. 50,000 cells/well were plated in DMEM assay media in a TC-treated 96-well plate. The following day, cells were infected with *C. trachomatis* L2-GFP at MOI 0.5 as described above. Media was not changed after centrifugation because of the poor cell adherence. Cells were incubated at 37°C for 72 h. NF-κB activity was assayed by QUANTI-Blue solution (InvivoGen rep-qbs). 20 μL cell supernatant was incubated with 180 μL of QUANTI-Blue solution in a non-TC-treated flat-bottom, clear 96-well plate. Stimulation of the present TLRs activates NF-κB and AP-1, which induce the production of SEAP. The SEAP signal was read at 620 nm at 2 h. Statistical significance was determined by ordinary one-way ANOVA with Dunnett's multiple comparisons test on raw data.

Population analyses and linear modeling
All population analyses were performed on the log_2-transformed concentration of IL-6 (pg/mL) measured in culture supernatants. Linear modeling analysis was performed in R version 4.1.1 using the lm function included in the base package. For estimating the effects of *C. trachomatis* burden and SNP on cytokine levels, we modeled:

$$log_2(IL6\ or\ IFNa2) = \beta_0 + \beta_1(70\ hr\ \%\ infected) + \beta_2 rs1057807 + \varepsilon$$

Two models were compared for estimating the effects of SNP and population as follows:

$$log_2(IL6) = \beta_0 + \beta_1 population + \beta_2 rs1057807\ (true) + \varepsilon$$

$$log_2(IL6) = \beta_0 + \beta_1 population + \beta_2 rs1057807\ (random) + \varepsilon$$

rs1057807 genotype (AA, AG, or GG) was modeled as a numerical variable. Population was modeled as a categorical variable with 4 categories: ESN (Esan in Nigeria), GWD (Gambian in Western Divisions in the Gambia), KHV (Kinh in Ho Chi Minh City, Vietnam), and IBS (Iberian in Spain). Coefficients were calculated relative to ESN. Regression coefficients obtained from the model with the true genotype data were compared to a distribution created by randomly permuting the genotype data 1 million times to estimate how rs1057807 changed the effect of each population (relative to ESN). Empirical p values were calculated based on the distribution from permuted genotype data. R scripts are provided in Methods S1. Values for cytokine levels and other phenotypes used in modeling are provided in Data S1.
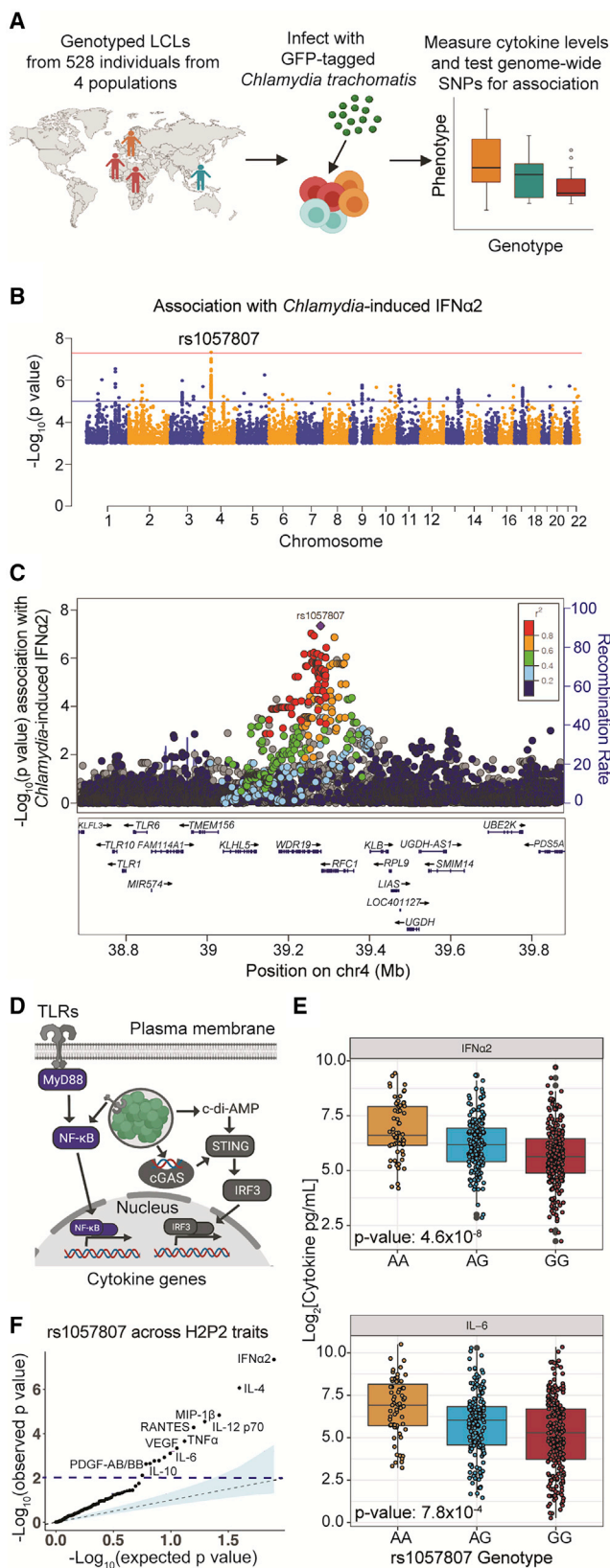
Figure 1. A locus on chromosome 4 is associated with levels of multiple cytokines following *C. trachomatis* infection

(A) Schematic of the H2P2 *C. trachomatis* cellular GWAS approach. *C. trachomatis* infection of LCLs derived from four populations (Esan in Nigeria [ESN], Gambian in Western Divisions of the

Results

Cellular GWAS of cytokine levels upon *C. trachomatis* infection reveals an association with SNPs on human chromosome 4

Previously, we carried out a family-based GWAS on 528 LCLs across 79 traits[27] (Figure 1A). Of those 79 traits, 34 traits were cytokine levels in uninfected cells or cells infected with a pathogen. We identified a SNP, rs1057807, that surpassed genome-wide significance ($p = 4.6 \times 10^{-8}$ from QFAM-Parents in PLINK[52,53]) with production of the cytokine IFNα2 following *C. trachomatis* infection (Figure 1B). A traditional genome-wide significance threshold of $p < 5 \times 10^{-8}$ was used,[54,55] and a QQ plot for $\log_2[\text{IFN}\alpha2]$ with a minor allele frequency cutoff of 5% displayed no inflation of the test statistic ($\lambda = 1.009$) with modest deviation of rs1057807 toward a p value lower than expected by chance (Figure S1). This SNP is located on chromosome 4 in the gene encoding a DNA replication factor (*RFC1*), which is located 500 kb away from the *TLR10/TLR1/TLR6* genes, encoding known pathogen recognition receptors (Figure 1C). *C. trachomatis* stimulates cytokine production through two parallel pathways—a TLR pathway involving the adaptor MyD88[56–59] and a STING (stimulator of interferon genes)-dependent pathway[60,61] (Figure 1D).

The A-allele of rs1057807 is associated with higher IFNα2 production in LCLs upon exposure to

Gambia [GWD], Iberian in Barcelona, Spain [IBS], and Kinh Ho Chi Minh City, Vietnam [KHV]) was monitored at 27, 46, and 70 h post-infection by flow cytometry. At 70 h post-infection, supernatants from cultured cells were collected for subsequent assay by Luminex panel, and $\log_2$-transformed concentrations were used to determine genome-wide association.

(B) Manhattan plot of *C. trachomatis*-induced IFNα2 production shows a group of SNPs in linkage disequilibrium associated with *C. trachomatis*-induced levels of IFNα2 on chromosome 4. rs1057807 is the lead SNP ($p = 4.5 \times 10^{-8}$). Dotted line indicates genome-wide significance (5×10^{-8}). p values were calculated using QFAM-Parents in PLINK.

(C) Regional plot of chromosome 4 reveals *TLR1*, *TLR6*, and *TLR10* are located 500 kb away from rs1057807. All genes located within a 600 kb span are displayed. SNPs are plotted by position on chromosome 4 and by $-\log_{10}(\text{p value})$ of association to *C. trachomatis*-induced IFNα2 production and are color-coded by linkage disequilibrium (r^2) value to rs1057807 from 1000 Genomes European Data.

(D) *C. trachomatis* induces cytokine production in parallel pathways: a TLR-dependent pathway and a STING-dependent pathway.

(E) rs1057807 A-allele is associated with IFNα2 and IL-6 production in LCLs infected with *C. trachomatis*. p values are from family-based association analysis using QFAM-parents in PLINK. Line marks the median and box indicates the first and third quartiles.

(F) Q-Q plot of rs1057807 for the 79 H2P2 phenotypes demonstrates many phenotypes have p values lower than expected by chance (indicated by deviation from the gray dotted line). rs1057807 is associated with *C. trachomatis*-induced production of multiple cytokines. Individual cytokines screened in H2P2 are plotted against their $-\log_{10}(\text{p value})$ relative to the rs1057807 locus. Blue dotted line indicates a p value of 0.01.

C. trachomatis than the G-allele (Figure 1E). While only the association with IFNα2 surpassed genome-wide significance, this SNP was also associated with higher expression levels of additional cytokines. Of 17 cytokines measured, ten cytokines showed an association with rs1057807 of $p < 0.01$ (from QFAM-Parents in PLINK[52,53]) with the same direction of effect, including IL-6 and TNFα (Figure 1F; $p < 0.01$ indicated by blue dotted line), cytokines that are induced during *C. trachomatis* infection of mice and humans.[45,46,62,63] This effect on multiple cytokines was not simply secondary to interindividual variation in burden, as no association was observed between this SNP and levels of *C. trachomatis* infection at multiple time points (Table S2), and linear modeling of $\log_2$[IFNα2] with SNP and 70 h post-infection burden as covariates revealed the SNP effect was much larger (regression coefficient $= -0.623$ versus 0.017 for burden) and was highly significant, with a p value of 1.87×10^{-13} (versus $p = 0.003$ for burden) (Table S3). Previous studies have identified a signal within the *TLR10/TLR1/TLR6* locus that modulated cytokine expression following stimulation of whole blood with the TLR1-TLR2 agonist, Pam$_3$CSK$_4$.[64,65] Notably, rs1057807 is not in linkage disequilibrium with rs4543123, the lead SNP identified by Mikacenic et al.[65] ($r^2 = 0.0$ in European, East Asian, and African populations) or any known SNPs in the *TLR10/TLR1/TLR6* gene region including the *TLR1* missense variant, rs5743618 (I602S), identified by Hawn et al.[66] Additionally, rs4543123 and other SNPs in the *TLR10/TLR1/TLR6* gene locus were not associated with *C. trachomatis*-induced cytokine production (see Figure 1C).

Long-range chromosomal interactions involving the rs1057807 locus likely regulate expression of TLRs

We examined SNP annotation data in Haploreg[67] to determine if rs1057807 or SNPs in linkage disequilibrium ($r2 > 0.6$ in European or African populations) might impact the function of any nearby genes. No nonsynonymous variants were noted, but the data suggested possible association with the expression levels of nearby genes. We evaluated a previously published transcriptomics dataset of LCLs[68] to determine if rs1057807 was an eQTL for any genes in a 700 kb window. At a Bonferroni-corrected significance threshold of $p < 0.005$ (for ten candidate genes), rs1057807 was associated with the expression of *TLR1* ($p = 2.7 \times 10^{-5}$ from linear regression), *TLR6* ($p = 6.1 \times 10^{-5}$), and *TLR10* ($p = 5.9 \times 10^{-5}$) (Figure 2A), providing a plausible explanation for this SNP being associated with levels of multiple cytokines.

Despite the association with *TLR10/TLR1/TLR6* expression, the rs1057807 locus and *TLR10/TLR1/TLR6* locus are 500 kb apart, suggesting that any interaction between these two regions must be through long-range chromatin interactions. Promoter Capture Hi-C, a variation of 3C that reveals long-range interactions with promoters,[69] from an LCL (GM12878) suggested a chromatin loop connecting the rs1057807 locus to the core promoter of *TLR6*

(Figure 2B, top). A possible interaction between the *TLR6* promoter and the rs1057807 locus was further supported by DNase hypersensitivity (DHS) linkage (Figure 2B, bottom),[70] since many known enhancers become DNase hypersensitivity sites synchronously with promoters of their target genes.[70] Together, Promoter Capture Hi-C and DHS linkage are two distinct and complementary approaches that independently suggested an interaction between the rs1057807 region and the *TLR6 p*romoter.

To validate and more finely map the long-range interaction and identify possible regulatory regions leading to variation in TLR expression, we prioritized rs1057807-linked SNPs using three chromatin landscape datasets: chromatin QTLs (caQTLs),[72] DNase-hypersensitive regions,[73] and regions exhibiting enhancer activity.[74] This stratification pointed us to a 100 kb region to test by 3C. We tested eight different fragments in this region for interaction with the *TLR6* promoter and identified a peak localized to the HindIII fragment 474 kb away from the *TLR6* promoter region in LCLs (Figure 2C). While LCLs demonstrate characteristics and transcriptional profiles very similar to antigen-activated B cells,[28] and thus can serve as a model for immune cell interactions with *C. trachomatis*, we also wanted to determine whether this long-range interaction was observed in a cervical epithelial cell line. Using 3C, we observed a similar peak in interaction frequency between the *TLR6* promoter and the HindIII fragment located 474 kb away in HeLa cells (Figure 2D). Thus, we observed a chromatin loop and a long-range interaction between the *TLR6* promoter and a locus nearly 500 kb away.

TLR1 and TLR6 regulate cytokine levels in response to *C. trachomatis* infection

The eQTL and 3C data indicated that the genetic association between the rs1057807 locus and cytokine levels might be mediated by regulation of expression of *TLR1*, *TLR6*, and/or *TLR10*. The TLR1, TLR6, and TLR10 proteins localize to the plasma membrane and signal through the MyD88 adaptor protein.[75–78] Of the cytokines associated with rs1057807, *IL-6*[79] is a direct target of the NF-κB transcription factor, which is activated by MyD88 and cell-surface TLRs.[80] Other cytokines are reported to be primarily STING-dependent, such as type I interferons (IFNα2)[81] (see Figure 1D). To test if cytokines could be stimulated by extracellular *C. trachomatis*, we incubated LCLs with heat-inactivated elementary bodies (Figure 3A). Following 48 h of stimulation, heat-inactivated *C. trachomatis* induced IL-6 production to a similar level as live *C. trachomatis* (Figure 3A).

To broadly assess if the induction of cytokines following *C. trachomatis* infection is dependent on TLRs that function in a MyD88-dependent manner, we infected LCLs with *C. trachomatis* in the presence of a MyD88 inhibitory peptide and measured cytokine production 72 h post-infection. We observed a greater inhibition of *C. trachomatis*-induced IL-6 (85%) compared to IFNα2
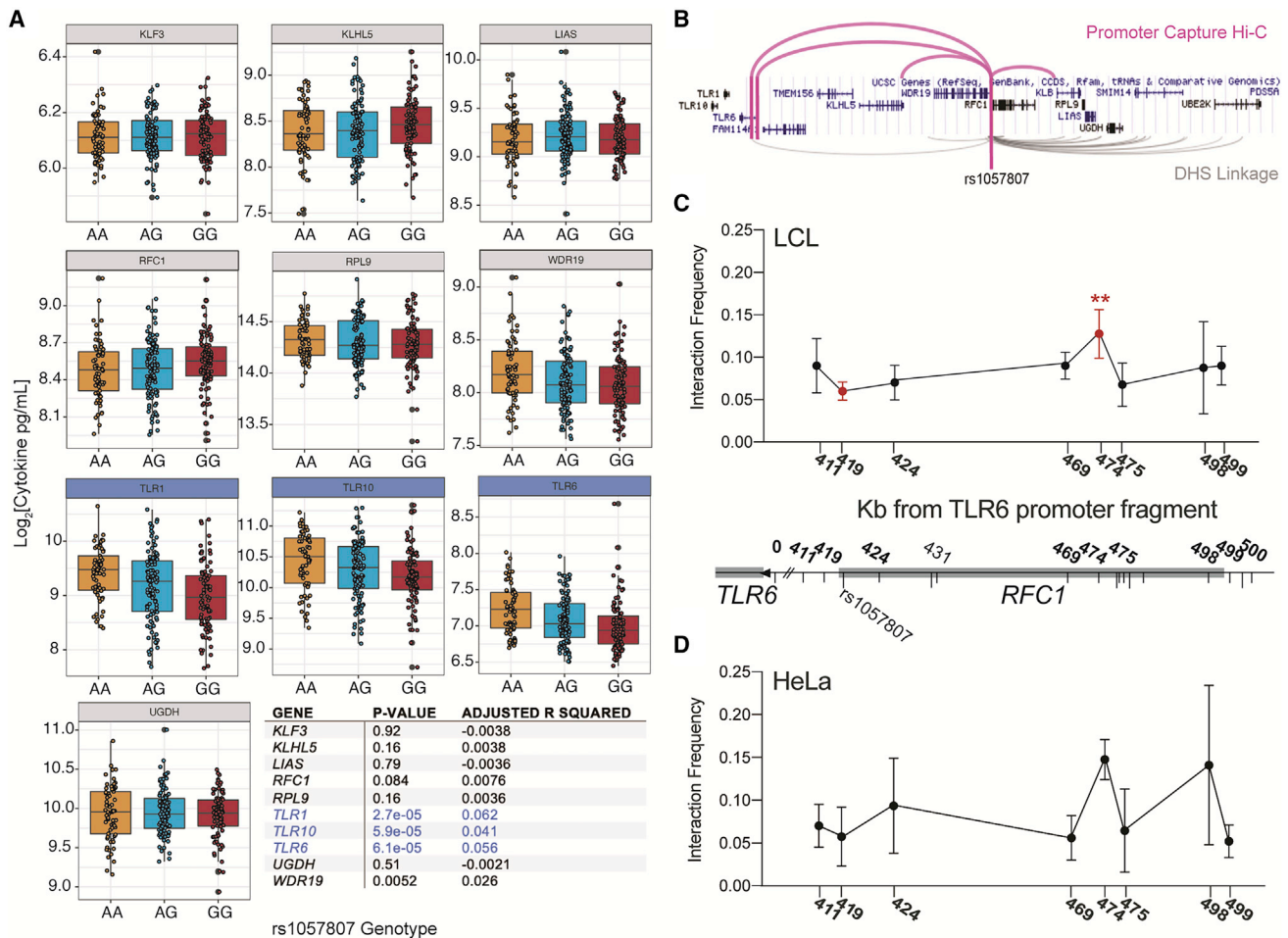
Figure 2. eQTL and long-range chromatin interaction data implicate *TLR6* as regulated by the rs1057807 locus

(A) The rs1057807 A-allele is associated with higher expression of *TLR1*, *TLR6*, and *TLR10* mRNA in HapMap LCLs (n = 270 individuals total). Gene expression values for each LCL are from the Stranger 2007 dataset.[68] p values were determined by linear regression. Significance determined by a Bonferroni-corrected p value of <0.005. Line marks the median and box indicates the first and third quartiles.

(B) Promoter Capture Hi-C[69] (top) and DNase hypersensitivity linkage[70] (bottom) indicate a loop forming between the region containing rs1057807 and the *TLR6* core promoter. Plots of LCL GM12878 Promoter Capture Hi-C and GM12878 DNase hypersensitivity linkage were modified from 3D Genome Browser.[71]

(C) Chromosome conformation capture (3C) in LCL HG01519 suggests the formation of a chromatin loop between the *TLR6* promoter fragment and an *RFC1* region containing SNPs in high linkage disequilibrium with rs1057807 in four experiments. Significance was determined by a one-way ANOVA with Holm-Šídák's multiple comparisons test against the mean of the peak fragment (474) on $\log_2$-transformed data. Data are represented as mean ± SEM.

(D) 3C in HeLa cells further suggests the formation of a chromatin loop between the *TLR6* promoter fragment and an *RFC1* region containing SNPs in high linkage disequilibrium with rs1057807 in two experiments. Data are represented as mean ± SEM.

(37%) as measured by ELISA (Figure 3B). The other ten *C. trachomatis*-induced cytokines associated with rs1057807 also show a partial dependence on MyD88 when measured with Luminex, consistent with the rs1057807 locus regulating expression of *TLR1*, *TLR6*, and *TLR10* to impact the production of multiple cytokines (Figure 3C). While these experiments were done with LCL HG01519, the degree of MyD88 dependence for IL-6 levels varied with ten LCLs from different donors (six from the ESN population and four from the IBS) from 26% to 93% (mean, 68%) (Figure 3D). These data are consistent with activation via a TLR-MyD88 pathway being the primary means of IL-6 activation in LCLs but also demonstrate interindividual variation in MyD88 dependence and sug-

gest that other pathogen recognition receptors such as STING, inflammasomes, and cytosolic NOD1/NOD2 are likely involved as well.

To further test the role of TLR1, TLR6, and TLR10 in *C. trachomatis*-induced cytokine production, we infected cells in the presence of anti-hTLR1 and anti-hTLR6 antibodies that sterically block TLR signaling. Because TLR10 does not have a known function or ligand, there are no known blocking antibodies for TLR10. Blocking TLR1 and TLR6 in LCLs, when compared to an IgG control, both caused modest but statistically significant reductions in cytokine levels following *C. trachomatis* infection (Figure 3E). Thus, blocking signaling by TLR1 and TLR6 further validate the role these TLRs play in
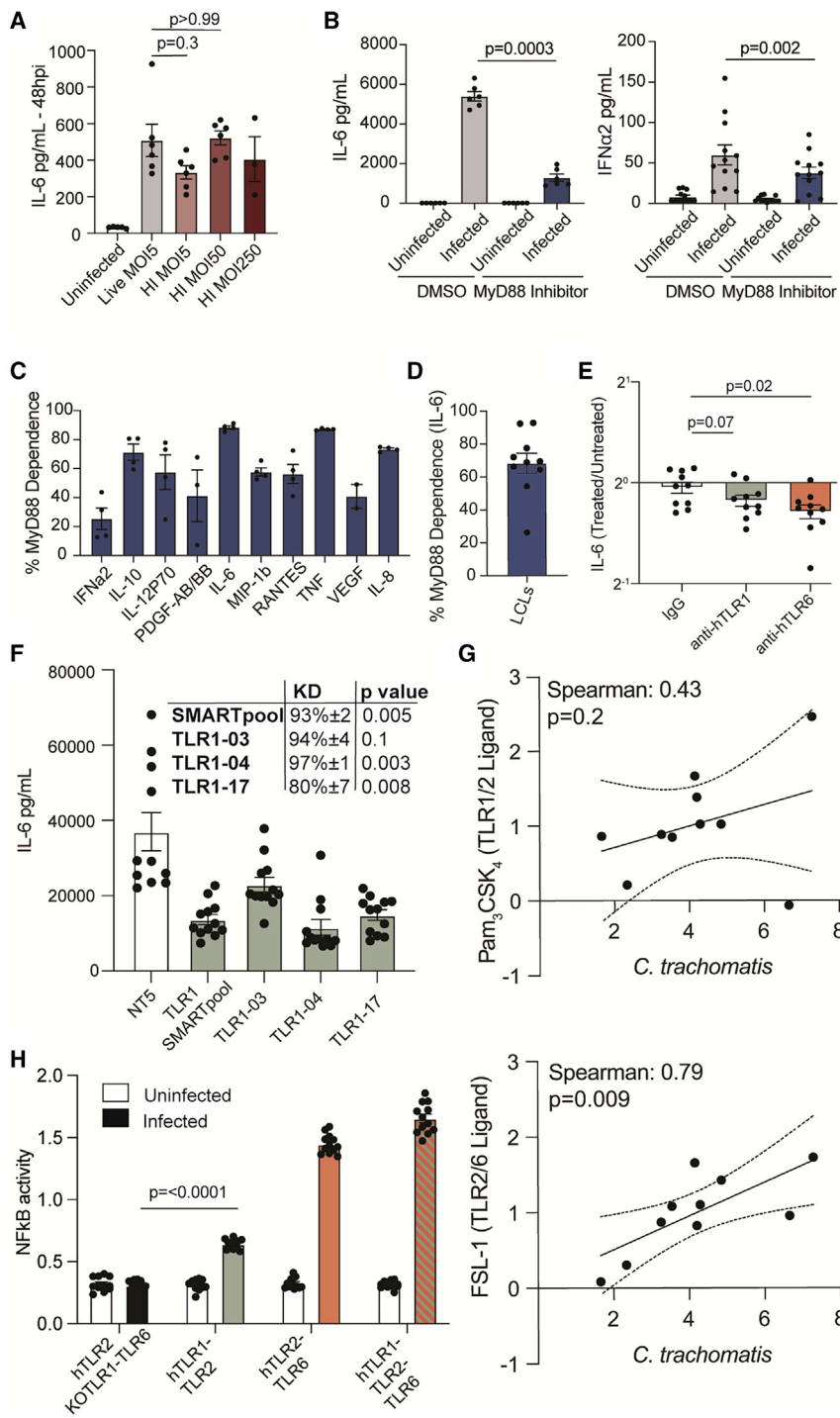
Figure 3. TLR1 and TLR6 induce IL-6 production in response to *C. trachomatis*

(A) Heat-inactivated *C. trachomatis* induces IL-6 production in LCLs. LCL HG01610 was left unexposed or exposed to live *C. trachomatis* or heat-inactivated *C. trachomatis* at the indicated MOI. IL-6 concentration was measured 48 h post-infection by ELISA in two experiments for uninfected, live MOI5, HI MOI5, and HI-MOI50 and a single experiment for HI MOI250. Samples below the limit of detection were assigned 9.4 pg/mL for IL-6. *C. trachomatis* was heat-inactivated at 55°C for 30 h prior to infection. Statistical significance was measured by Brown-Forsythe and Welch ANOVA with Dunnett's multiple comparisons test on log$_2$-transformed data comparing all means to the mean of the live MOI5 group as a control. Data are represented as mean ± SEM.

(B) IL-6 (left) and IFNα2 (right) production is partially dependent on MyD88. LCL HG01519 was pre-treated with 16 μM of a MyD88 inhibitor (ST2825) or DMSO 2 h before *C. trachomatis* infection. IL-6 production from two experiments and IFNα2 production in four experiments were measured by ELISA. Samples below the limit of detection were assigned 9.4 pg/mL for IL-6 and 3.1 pg/mL for IFNα2. Paired t tests were performed on log$_2$-transformed data. Data are represented as mean ± SEM.

(C) *C. trachomatis*-induced cytokines measured in H2P2 and associated with rs1057807 (p < 0.01 from QFAM-Parents in PLINK) are partially MyD88 dependent. LCL HG01519 was pre-treated with 16 μM of ST2825 or DMSO 2 h before *C. trachomatis* infection. A panel of 17 cytokines were measured at 72 h post-infection by Luminex. Only cytokines measured within the limit of detection and associated with rs1057807 in H2P2 with p < 0.01 are plotted. Percent MyD88 dependence was calculated as $1 - \left(\frac{MyD88\ infected\ cytokine}{DMSO\ infected\ cytokine}\right) \times 100$.

Data are represented as mean ± SEM.

(D) *C. trachomatis*-induced IL-6 production is partially dependent on MyD88 in 10 LCLs (from IBS: HG01505, HG01519, HG01522, HG01526, HG01610, HG01628; from ESN: HG02971, HG03114, HG03117, HG03522). LCLs were pre-treated with 16 μM of ST2825 or DMSO 2 h before *C. trachomatis* infection. IL-6 production was measured 72 h post-infection by ELISA

in three experiments. Percent MyD88 dependence was calculated as $1 - \left(\frac{MyD88\ infected\ IL-6}{DMSO\ infected\ IL-6}\right) \times 100$. Each dot represents an individual LCL. Data are represented as mean ± SEM.

(E) Blocking the function of TLR1 or TLR6 leads to a decrease in IL-6 production following *C. trachomatis* infection in 10 LCLs. 10 LCLs were pre-treated with 10 μg/mL of anti-hTLR1-IgG, anti-hTLR6-IgG, or IgG control 2 h before *C. trachomatis* infection. IL-6 production was measured in three experiments by ELISA. Infected samples were normalized to the experimental average of the no-antibody infected control by cell line. Significance was determined by ordinary one-way ANOVA with Dunnett's multiple comparisons test on log$_2$-transformed data. Each dot represents an individual LCL. Data are represented as mean ± SEM.

(F) Knockdown of *TLR1* SMARTpool and individual SMARTpool siRNA leads to a decrease in IL-6 production following *C. trachomatis* infection. HeLa cells were treated with non-targeting (NT5) siRNA, *TLR1* siRNA SMARTpool, or the individual siRNA components of TLR1 SMARTpool 48 h prior to *C. trachomatis* infection. IL-6 production was measured 72 h post-infection by ELISA. Each experimental mean was adjusted to the grand mean by multiplying by a constant. Statistical significance was determined by Brown-Forsythe and Welch ANOVA with Dunnett's multiple comparisons test on normalized data comparing all means to the mean of the NT5 group as a control. Data are represented as mean ± SEM.

(legend continued on next page)

C. trachomatis-induced cytokine production, with TLR6 playing a more prominent role than TLR1 in LCLs.

Because the efficiency of siRNA-mediated silencing of *TLR1*, *TLR6*, and *TLR10* in LCLs was poor, we tested additional cell lines that displayed a similar profile of TLR-ligand responsiveness. Comparing IL-6 production in response to *C. trachomatis* infection, FSL-1 (a TLR2-TLR6 ligand), or Pam_3CSK_4 (a TLR1-TLR2 ligand), we determined that HeLa cells exhibited a similar IL-6 response pattern to LCLs (Figure S2). In HeLa cells, transfection of *TLR1* SMARTpool siRNAs led to robust knockdown (94% ± 7%) and reduction of IL-6 following *C. trachomatis* infection (45%, p = 0.005 from Brown-Forsythe and Welch ANOVA tests with Dunnett's multiple comparisons test on normalized data). Unfortunately, treatment with the *TLR6* SMARTpool siRNA was toxic to the cells, preventing accurate assessment. To further test the specificity of *TLR1* and attempt to circumvent *TLR6* siRNA toxicity, we tested the individual siRNA components of the SMARTpool. siRNAs with high toxicity to the cells and knockdown levels below 50% were excluded from our analysis. Three of the four *TLR1* siRNA SMARTpool components exhibited robust knockdown with no cytotoxicity (*TLR1* SMARTpool: 93%; TLR1-03: 94%; TLR1-04: 97%; TLR1-17: 80%), and all three exhibited substantially reduced levels of IL-6 (Figure 3F). Unfortunately, two of the four *TLR6* siRNA SMARTpool components were toxic to the cells, while the other two (TLR6-02 and TLR6-05) exhibited similar modest 27% and 28% knockdown. Therefore, while the role of TLR6 was still equivocal by RNAi, our results with multiple siRNAs demonstrated the importance of TLR1 in cytokine production following *C. trachomatis* infection.

Both TLR1 and TLR6 signal through heterodimerization with TLR2.[82,83] To further examine the specificity of TLR activation by *C. trachomatis*, we compared the interindividual variation in cytokine levels in response to *C. trachomatis* to the variation observed with synthetic TLR ligands with well-characterized specificity. Ten LCLs were infected with *C. trachomatis* or stimulated with FSL-1 (a synthesized ligand for TLR2-TLR6) or Pam_3CSK_4 (a synthesized ligand for TLR1-TLR2) and we measured the correlation among responses. The LCLs with higher IL-6 production following infection with *C. trachomatis* also had higher IL-6 production when stimulated with FSL-1 (Spearman coefficient of 0.72, p = 0.009), while correlation with Pam_3CSK_4 was not significant (Spearman coefficient of 0.43, p = 0.2) (Figure 3G). These data suggest that TLR2-TLR6 may be the predominant heterodimer that *C. trachomatis* signals through in LCLs.

To further test the role of TLR1-TLR2 and TLR2-TLR6 heterodimers in producing cytokines in the presence of *C. trachomatis* infection, we infected HEK-Blue hTLR2 KOTLR1-TLR6, hTLR1-TLR2, hTLR2-TLR6, and hTLR1-TLR2-TLR6 cells with *C. trachomatis*. In HEK-Blue cells, NF-κB activation is easily quantified by an NF-κB/AP-1 inducible SEAP reporter gene. *C. trachomatis* infection activated NF-κB in the hTLR1-TLR2, hTLR2-TLR6, and hTLR1-TLR2-TLR6 cell lines, with a greater response with the hTLR2-TLR6 heterodimer (Figure 3H). The hTLR2 KOTLR1-TLR6 cells did not activate NF-κB in the presence of *C. trachomatis*, further supporting the importance of TLR1 and TLR6 in *C. trachomatis*-induced cytokine production. Thus, both TLR1 and TLR6 sense *C. trachomatis*, although, consistent with our genetic data, TLR6 plays a greater role in *C. trachomatis*-induced cytokine production than TLR1. While our experiments with MyD88 inhibitor, TLR blocking antibodies, and TLR knockdown and overexpression firmly support the importance of MyD88-dependent TLR signaling in cytokine production during *C. trachomatis* infection, these experiment by themselves do not formally test the causal chain from SNP to TLR expression to cytokine levels that is implied by the associations of rs1057807 with TLR expression and rs1057807 with cytokine levels following *C. trachomatis* infection. Future studies could incorporate more complex experimental designs combining experimental perturbation in a larger number of LCLs, for example, determining whether the effect of rs1057807 on cytokines becomes less apparent with MyD88 inhibitor treatment.
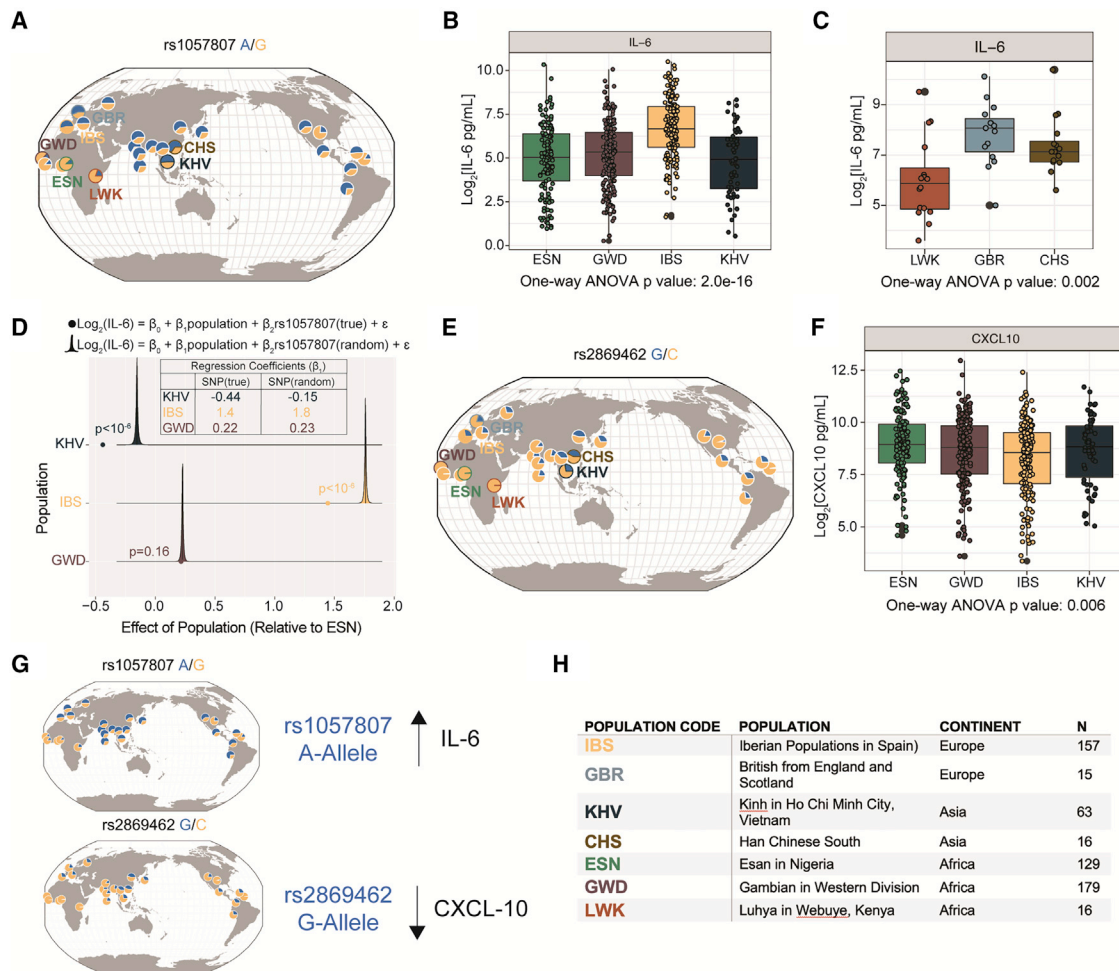
Differences across populations in cytokine responses to *C. trachomatis* infection

While there are many community-level factors that contribute to interindividual variation in *C. trachomatis* infection, such as income[84] and access to healthcare,[85] natural human genetic diversity may also contribute to these interindividual differences. In examining population frequencies across the globe for rs1057807 from the 1000 Genomes Project,[33] we found that the rs1057807 A-allele, the derived allele associated here with higher cytokine production, is less common in the GWD, MSL (Mende in Sierra Leone), YRI (Yoruba in Ibadan, Nigeria), and ESN, LWK (Luhya in Webuye, Kenya) populations across the continent of Africa and ASW (African Ancestry in Southwest US) and ACB (African Caribbean in Barbados) populations in North America compared to other regions of the world (Figure 4A) based on data from the Geography of Genetic Variants web browser.[86]

(G) LCL IL-6 production in response to *C. trachomatis* correlates with IL-6 production in response to FSL-1, the synthesized ligand for the TLR2-TLR6 heterodimer. 10 LCLs were infected with *C. trachomatis* or stimulated with 1 μg/mL of Pam_3CSK_4 (top) or 1 μg/mL FSL-1 (bottom) for 72 h. IL-6 production was measured by ELISA. Treated IL-6 levels were normalized to untreated infected IL-6 production for each cell line. Spearman correlation was performed on log_2-transformed data.

(H) *C. trachomatis* infection induces NF-κB activation in HEK-Blue TLR2-TLR6 cells and HEK-Blue TLR1-TLR2 cells. HEK-Blue TLR2-TLR6 and HEK-Blue TLR1-TLR2 cells were infected with *C. trachomatis*, and after 72 h post-infection, NF-κB activation was measured. Statistical significance was determined by ordinary one-way ANOVA with Dunnett's multiple comparisons test on raw data. Data are represented as mean ± SEM.

Figure 4. Allele frequency differences contribute to population differentiation of cytokine responses

(A) The rs1057807 A-allele is present at a lower frequency in GWD, MSL, YRI, and ESN populations from Africa, plotted using the Geography of Genetic Variants website.[86]

(B) *C. trachomatis*-induced IL-6 production is lower in ESN and GWD populations than the IBS population when measured in 528 LCLs. IL-6 production was measured at 70 h post-infection by Luminex as part of the H2P2 screen.[27] p values are from one-way ANOVA on $\log_2$-transformed IL-6 pg/mL values. Line marks the median and box indicates the first and third quartiles.

(C) *C. trachomatis*-induced IL-6 production is lower in LWK than GBR and CHS in 47 LCLs as a single batch. IL-6 production was measured 72 h post-infection by Luminex. p values are from one-way ANOVA on $\log_2$-transformed IL-6 pg/mL values. Line marks the median and box indicates the first and third quartiles.

(D) Linear modeling of $\log_2$[IL-6] as a function of population and genotype as covariates following *C. trachomatis* infection in 528 LCLs. The effect size of rs1057807(true) relative to ESN population (colored dot) is compared to rs1057807(random) relative to ESN population (density) as determined by 1 million SNP permutations. Empirical p values are calculated from distribution of permutations. The regression coefficient for rs1057807(true) effect (β_2) with A as the effect allele is +0.37, consistent with the A-allele being associated with higher cytokine levels. Regression coefficients for the population effect (β_1) when rs1057807(true) is used or permutated with rs1057807(random) are listed in the table.

(E) The rs2869462 G-allele is present at a lower frequency in GWD, MSL, YRI, and ESN populations derived from Africa, plotted using the Geography of Genetic Variants website.[86]

(F) *C. trachomatis*-induced CXCL-10 production is higher in ESN and GWD populations than the IBS population when measured in 528 LCLs. IL-6 production was measured at 70 h post-infection by Luminex as part of the H2P2 screen. p values are from one-way ANOVA on $\log_2$-transformed CXCL-10 pg/mL values. Line marks the median and box indicates the first and third quartiles.

(G) The rs1057807 A-allele, present at a lower frequency in GWD, MSL, YRI, ESN, ASW, and ACB, is associated with higher IL-6 production. The rs2869462 G-allele, present at a lower frequency in GWD, MSL, YRI, ESN, ASW, and ACB, is associated with lower CXCL10 production.

(H) Table of populations used in population genetics studies, continent of population, and number of samples in each analysis.

Because of the lower frequency of the A-allele in ESN and GWD populations screened in H2P2, we examined *C. trachomatis*-induced IL-6 production by population in the H2P2 LCLs. Both ESN and GWD show lower IL-6 production when compared to the IBS population (p = 2 × 10^{-16} from one-way ANOVA) (Figure 4B). As H2P2 of *C. trachomatis* was carried out in batches over a period of 6 months, we wanted to confirm that this association between population and IL-6 could be replicated when measured in a single batch of infections. We tested this

in 47 LCLs from three additional populations: CHS (Southern Han Chinese), GBR (British in England and Scotland), and LWK. LWK showed lower IL-6 production when compared to both the CHS and GBR populations (p = 0.002 from one-way ANOVA) (Figure 4C). Thus, the lower IL-6 expression seen with LCLs from two African populations (ESN and GWD) compared to IBS (a European population) was also observed with LWK compared to GBR.

To examine whether rs1057807 contributes to population differences for IL-6, we used a multiple linear modeling approach with covariates for population and rs1057807 genotype and compared a model with the true rs1057807 genotype data ("rs1057807[true]") to a model where rs1057807 genotypes were randomly permutated among LCLs ("rs1057807[random]"). This analysis allowed us to quantify the contributions of the SNP to the population effect on IL-6 concentration (Figure 4D). Relative to ESN, only IBS had significantly greater IL-6 (regression coefficient = 1.4; p = 1.9×10^{-8} by linear regression; also see Figure 4B). Following 1 million permutations of the rs1057807 genotype, the distribution of regression coefficients for each population was calculated and plotted (Figure 4D; results of rs1057807[random] displayed as density distributions) relative to ESN. The effect of population for IBS (relative to ESN) was significantly reduced when modeled with true genotypes than if genotype was randomly assigned (1.4 versus 1.8, empirical p < 1 × 10^{-6}). These results are consistent with the higher frequency of the pro-inflammatory A-allele in IBS (53%) compared to ESN (12%) contributing to the higher IL-6 levels in IBS. Interestingly, the effect of population became stronger (−0.44 versus −0.15, empirical p < 1 × 10^{-6}) when KHV was modeled with true genotypes compared to random genotypes. This suggests that due to the increased frequency of rs1057807-A in the KHV population (51%) relative to ESN (12%), the phenotypic difference between KHV and ESN is relatively small and would be nearly 3× larger without the effect of rs1057807. However, we note that the lower IL-6 level in KHV compared to ESN did not reach statistical significance in the linear model, and in our single batch infection described above, the Asian population used (CHS) had greater IL-6 than the LWK population. Finally, including true SNP genotypes did not alter the small effect of population between GWD and ESN, consistent with the allele frequencies being nearly the same in these populations (13% and 12%, respectively).

IL-6 is just one cytokine secreted in response to *C. trachomatis* and does not act in isolation to promote inflammation. For instance, we have previously reported another SNP, rs2869462, associated with levels of a different cytokine (CXCL10) following *C. trachomatis* infection.[27,87] The geographic distribution of rs2869462[86] is similar to rs1057807 in that the derived allele (G) has a relatively lower frequency in GWD, MSL, YRI, ESN, LWK, ASW, and ACB populations compared to several populations in Europe and Asia (Figure 4E), but in

this case, the derived allele is associated with lower levels of CXCL10. We hypothesized that LCLs might also exhibit varying levels of CXCL10 among populations. Indeed, we observed in the H2P2 dataset that LCLs from two African populations had higher CXCL10 (Figure 4F; p = 0.006 from one-way ANOVA), but population only accounts for 1.8% of the variance (from adjusted r^2 by one-way ANOVA). The relationship between the rs2869462 G-allele (derived and more common in populations outside of Africa) and lower CXCL10 is opposite what we see with rs1057807 A-allele (derived and more common in populations outside of Africa) and higher IL-6 levels (Figures 4G and 4H). Therefore, the immune response to *C. trachomatis* is complex and the contributions of human genetic diversity to this response are not easily generalized.

Positive selection of rs1057807 in broadly East Asian populations is distinct from the introgressed *TLR10/TLR1/TLR6* gene region

Differences in allele frequencies among populations is suggestive of natural selection caused by varying environment. To test for evidence of recent allele frequency changes likely caused by natural selection (<150 generations) for rs1057807, we queried singleton density scores (SDSs)[88] from 39,649 unrelated individuals in the TOPMed whole genomes dataset restricted to broadly East Asian, broadly European, and broadly African population groupings.[89] The genotypes included in the population groupings had more than 90% inferred East Asian, European, or African ancestry to avoid potential problems with selection analysis introduced by admixture.[89] The SDS scores are normalized to have a mean of 0 and variance of 1, where an SDS score > 0 corresponds to increased frequency of the derived allele and an SDS < 0 corresponds to increased frequency of the ancestral allele over recent history.[88] Interestingly, the rs1057807-derived A-allele, associated with higher IL-6, appears to have been selected for in broadly East Asian populations (SDS score = 3.1; p = 8.5 × 10^{-4} calculated from Z-score distribution) (Figure 5A; Table S4), with substantially weaker evidence for selection of the same allele in broadly European populations (SDS score = 0.92; p = 0.36 calculated from Z-score distribution) (Figure 5A; Table S4). This is consistent with the A-allele reaching higher frequency in the JPT, CHS, and CHB populations from East Asia (68%, 65%, and 62% frequency, respectively) compared to broadly European and African populations (Figure 4A). SDS did not support evidence of selection of the rs1057807-derived A-allele in the broadly African population group (SDS score = −0.35) (Figure 5A; Table S4). Thus, there is evidence for recent positive selection in broadly East Asian populations for the rs1057807 A-allele, associated with a more pro-inflammatory immune response by innate immune mechanisms (TLR-dependent cytokines including IL-6).

Others have noted introgression of Neanderthal- and Denisovan-like haplotypes[91,92] in the *TLR10/TLR1/TLR6* gene region. Notably, the introgression region borders
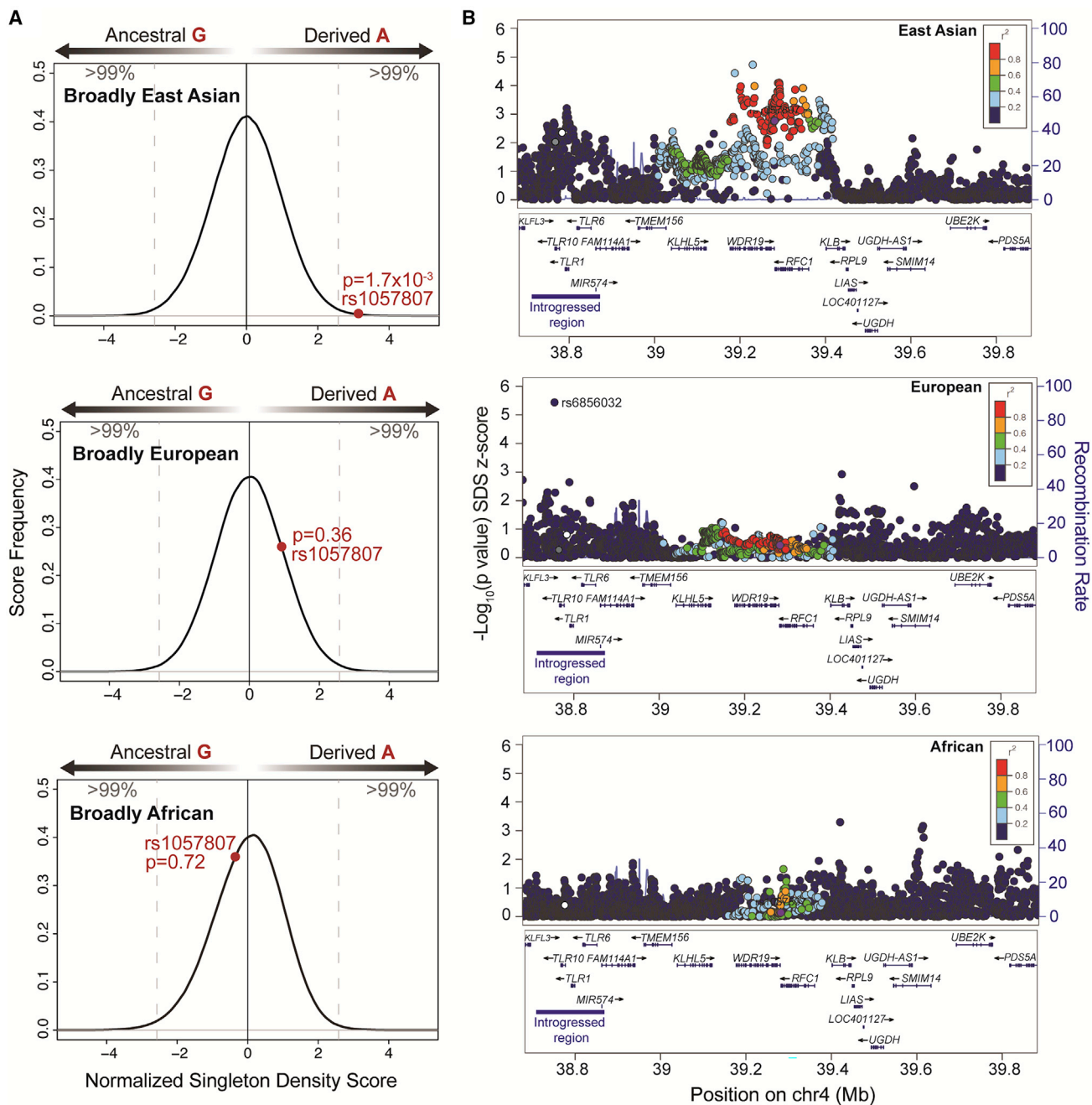
Figure 5. rs1057807 shows evidence of positive selection in broadly East Asian populations and is a distinct signal from the introgressed *TLR10/TLR1/TLR6* region

(A) The rs1057807 A-allele shows evidence of recent increase in allele frequency likely caused by natural selection in broadly East Asian populations (top). SDS values[88] from TOPMed genomes[89] are plotted on a standard normal distribution. A negative SDS value corresponds to selection of the ancestral allele, and a positive SDS value corresponds to selection of the derived allele. p values are calculated from *Z*-score distribution where normalized scores have a mean of 0 and standard deviation of 1. The rs1057807 A-allele shows weak evidence of recent increase in allele frequency likely caused by natural selection in broadly European population group (middle). SDS did not support evidence of selection of the rs1057807-derived A-allele in the broadly African population group (bottom).

(B) Evidence for positive selection of the rs1057807 A-allele in the broadly East Asian population grouping is a distinct signal from the introgressed *TLR10/TLR1/TLR6* region[90] (highlighted by blue line) (top). All genes located within 600 kb flanking region of rs1057807 are displayed. SNPs are plotted by position on chromosome 4 and by −log(p value) of the SDS p value calculated from the *Z*-score and are color-coded by linkage disequilibrium (r^2) value to rs1057807 from 1000 Genomes Asian (top), European (middle), and African (bottom) data. rs4543123[65] is shown in white. rs4274855[91] is shown in gray.

(chr4: 38760338-38846385; hg19)[90] span the *TLR10/TLR1/TLR6* genes (Figure 5B, blue line) but not the region surrounding rs1057807 (Figure 5B, purple dot). A tag SNP

of the introgressed Neanderthal haplotype as identified by Dannemann et al.,[91] rs4274855, is not in linkage disequilibrium with rs1057807 ($r^2 = 0.0$) (Figure 5B, gray
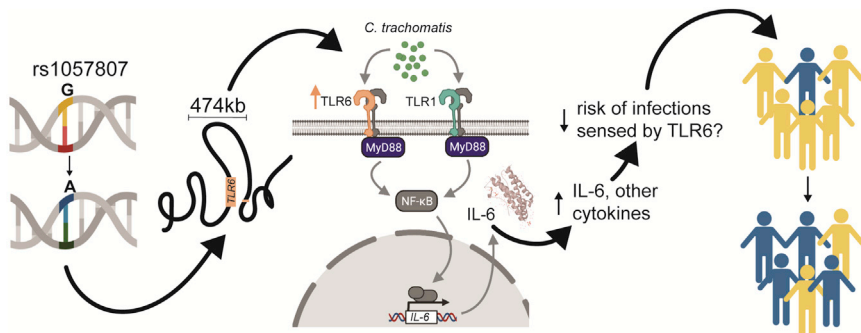
Figure 6. A model for how the rs1057807 locus impacts TLR signaling, cytokine production, and broader outcomes

dot). Additionally, the SNP with the greatest association to TLR1-TLR2 response on whole blood when stimulated with Pam_3CSK_4, rs4543123, is located in the *TLR10/TLR1/TLR6* introgressed haplotype but, as noted earlier, is also not in linkage disequilibrium with rs1057807 ($r^2 = 0.0$) (Figure 5B, white dot).[65] A regional LocusZoom plot with $-\log_{10}$(p values) based on SDS scores from TOPMed[89] suggests evidence for two distinct regions of positive selection in East Asians, one involving the *TLR10/TLR1/TLR6* introgressed region (Figure 5B, blue box) consistent with previous evidence of positive selection here[93] and the second involving the region around rs1057807 (Figure 5B, purple dot).

Discussion

We have identified that positively selected human genetic variation regulates the immune response to *C. trachomatis* (Figure 6). Specifically, a locus 500 kb away from the TLR genes forms a long-range chromatin loop to likely regulate *C. trachomatis*-induced cytokine production. Our data show that both TLR1 and TLR6 are important for *C. trachomatis*-induced cytokine production, though we show that TLR6 is more responsive to *C. trachomatis* infection than TLR1. Population studies suggest that there has been recent selection for the rs1057807 A-allele, associated with higher *C. trachomatis*-induced IL-6 production in East Asian populations. Future studies will be necessary to test whether this human genetic variation in *C. trachomatis* cytokine response may play a role in the diverse outcomes commonly seen in *C. trachomatis*-infected individuals.

Of note, the locus identified here is located in a non-coding region distant from the gene it regulates. Many genetic variants identified through GWASs are in non-coding regions of the genome, complicating functional interpretation.[94,95] Our work demonstrates the advantages of integrating eQTL datasets, linkage disequilibrium data, GWASs, and 3C to identify likely causal genes whose expression is regulated by non-coding variants. Most studies focus on the SNPs that regulate the genes they are located in or nearby.[15,19,24,87,96] However, in a large eQTL study of peripheral blood mononuclear cells of 2,112 individuals, Kirsten et al.[96] found that more than one-third of

SNPs identified were not located directly adjacent to the gene regulated. By expanding our search window to include genes located 700 kb away from our SNP of interest and by identifying a chromatin loop linking the rs1057807 region and the *TLR6* promoter, we were able to reveal the likely causal gene whose expression is regulated by genetic diversity to cause variable levels of *C. trachomatis*-induced cytokine production. We further confirmed this long-range chromatin interaction and used 3C to identify a 1,244 bp region that interacts with the *TLR6* promoter. Based on datasets that catalog caQTLs,[72] DNase hypersensitive regions,[73] and regions exhibiting enhancer activity,[74] we find the region identified by 3C is located in a likely enhancer identified by the GeneHancer dataset (identifier: GH04J039339). Work beyond the current study may further define the mechanism of this long-range interaction.

Previous work identified association signals in the *TLR10/TLR1/TLR6* locus, primarily in the *TLR1* gene, associated with cytokine levels following stimulation of whole blood with the TLR1-TLR2 agonist Pam_3CSK_4.[64] A subsequent GWAS study from the same group using Pam_3CSK_4 solidified SNPs within the *TLR10/TLR1/TLR6* locus as the main genetic determinant of TLR1-TLR2 response in whole blood.[65] In contrast, we identified a signal located 500 kb away from the *TLR10/TLR1/TLR6* region that displays no linkage disequilibrium with any SNPs located in *TLR10/TLR1/TLR6* (Figures 1C and 5B). Currently, it is not clear why these previous studies did not identify rs1057807, nor why we did not identify the previously identified SNPs in *TLR10/TLR1/TLR6*. One possibility informed by our findings is that *C. trachomatis* stimulation of IL-6 and other cytokines in cultured cells primarily involves TLR6, while the work by Wurfel et al.[65] focused on SNPs that regulate TLR1-TLR2 signaling in whole blood. Indeed, these authors noted that they did not observe any association of the SNPs they identified with cytokine levels in response to TLR2-TLR6 ligands, including FSL-1.[65] Alternatively, the difference may be attributable to the cell types used. The association described by Wurfel and colleagues[64,65] was found using stimulation of whole blood, while our association was found in an LCL model. As neither is likely to be the most relevant tissue/cell type for *C. trachomatis* infection, future work examining the association in the context of epithelial cells/tissue (for genital infection) or tears (for ocular infection) are highly worthwhile. The combination of these studies emphasizes the complexity of gene regulation and underscores the exquisite specificity of TLR signaling. Collectively, these

studies demonstrate how different genetic variants within the same vicinity can have very different effects on diversity in immune responses.

In addition to highlighting the complexity of TLR biology, our study underscores the complexity of *C. trachomatis*-induced cytokine production. Previous work has identified a role for TLR1-TLR2 and TLR2-TLR6 through studying *C. trachomatis* components that induce cytokine production. A chlamydial lipoprotein, macrophage infectivity potentiator (MIP), induces inflammatory response through TLR2/TLR1/TLR6 and CD14.[42] The main receptor involved in recombinant MIP activation was the TLR1-TLR2 heterodimer, although MIP did stimulate the TLR2-TLR6 heterodimer to a lesser extent.[42] Another study identified fourteen chlamydial lipoproteins from the *C. trachomatis* genome, and some were found to induce proinflammatory cytokine production signaled primarily through a TLR1-TLR2 heterodimer.[59] Our findings suggest a greater role for a TLR6-containing dimer being important for *C. trachomatis*-induced IL-6 production. We hypothesize that additional ligands are important for inflammatory responses during *C. trachomatis* infections.

Finally, rs1057807 has an interesting geographic distribution: multiple populations of African descent show a lower frequency of the A-allele, the derived allele associated with higher cytokine production, compared to other populations from the 1000 Genomes Project.[33] Measured IL-6 levels are also lower in LCLs derived from individuals from ESN, GWD, and LWK populations compared to IBS and GBR, and rs1057807 can account for a substantial fraction of the population variation for this trait. While the samples used in our work are not meant to be a perfect representation of all the people in a geographic location or ancestry from a particular region, our results suggest that this difference in allele frequency could play a role in phenotypic differences among human populations and between individuals. Further, while there are significant differences in cytokine levels among populations in our infection model, there is tremendous interindividual variation within populations (>85%) that is driven by other uncharacterized factors.

We speculate that positive selection near rs1057807, a distinct signal from Neanderthal and Denisovan-like introgression at the *TLR10/TLR1/TLR6* locus, has been driven by infectious diseases, in which spread was facilitated by the rise of agriculture, industrialization, and an increase in population density.[97] These infectious diseases would select for a more robust inflammatory response for pathogen clearance, perhaps by innate immune mechanisms (e.g., IL-6). Indeed, our observation that the strongest selection for the rs1057807 A-allele is in the broadly East Asian population grouping is consistent with the earlier rise of agriculture in this part of the world.[98] TLR6 is a central node in immune sensing for many different pathogens,[82,99] so variation at this locus could have been driven by numerous pathogens, including gram-positive bacteria, tuberculosis, or leprosy.[100]

Since different human populations have different evolutionary histories and have been exposed to different pathogens, our study underscores how important it is to look for genetic variants in diverse human populations and at different markers and aspects of that immune response. As human populations have undergone unique histories, including epidemic, endemic, and pandemic exposure, it is not surprising that we observe nuance in population-level responses to pathogens and that the sum of these responses does not always align with intuition. Additional caution must be taken when considering whether particular alleles or responses are "adaptive," as indeed what was adaptive for our ancestors may no longer confer the greatest fitness today.

Data and code availability

The published article includes all data and code generated or analyzed during this study. Previously published cellular GWAS data on *C. trachomatis* infection are available with the H2P2 publication[27] and on the H2P2 web page (see Web resources).

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2021.100071.

Web resources

Biorender, https://biorender.com/
H2P2, http://h2p2.oit.duke.edu
OMIM, https://www.omim.org

References

1. Hu, V.H., Holland, M.J., and Burton, M.J. (2013). Trachoma: protective and pathogenic ocular immune responses to Chlamydia trachomatis. PLoS Negl. Trop. Dis. *7*, e2020.
2. Malhotra, M., Sood, S., Mukherjee, A., Muralidhar, S., and Bala, M. (2013). Genital Chlamydia trachomatis: an update. Indian J. Med. Res. *138*, 303–316.
3. Manavi, K., McMillan, A., and Young, H. (2004). The prevalence of rectal chlamydial infection amongst men who have sex with men attending the genitourinary medicine clinic in Edinburgh. Int. J. STD AIDS *15*, 162–164.

4. Haggerty, C.L., Gottlieb, S.L., Taylor, B.D., Low, N., Xu, F., and Ness, R.B. (2010). Risk of sequelae after Chlamydia trachomatis genital infection in women. J. Infect. Dis. *201* (*Suppl 2*), S134–S155.

5. Weström, L. (1975). Effect of acute pelvic inflammatory disease on fertility. Am. J. Obstet. Gynecol. *121*, 707–713.

6. Wiesenfeld, H.C., Sweet, R.L., Ness, R.B., Krohn, M.A., Amortegui, A.J., and Hillier, S.L. (2005). Comparison of acute and subclinical pelvic inflammatory disease. Sex. Transm. Dis. *32*, 400–405.

7. Abdelsamed, H., Peters, J., and Byrne, G.I. (2013). Genetic variation in Chlamydia trachomatis and their hosts: impact on disease severity and tissue tropism. Future Microbiol. *8*, 1129–1146.

8. Conway, D.J., Holland, M.J., Bailey, R.L., Campbell, A.E., Mahdi, O.S., Jennings, R., Mbena, E., and Mabey, D.C. (1997). Scarring trachoma is associated with polymorphism in the tumor necrosis factor alpha (TNF-alpha) gene promoter and with elevated TNF-alpha levels in tear fluid. Infect. Immun. *65*, 1003–1006.

9. Conway, D.J., Holland, M.J., Campbell, A.E., Bailey, R.L., Krausa, P., Peeling, R.W., Whittle, H.C., and Mabey, D.C. (1996). HLA class I and II polymorphisms and trachomatous scarring in a Chlamydia trachomatis-endemic population. J. Infect. Dis. *174*, 643–646.

10. Mozzato-Chamay, N., Mahdi, O.S., Jallow, O., Mabey, D.C., Bailey, R.L., and Conway, D.J. (2000). Polymorphisms in candidate genes and risk of scarring trachoma in a Chlamydia trachomatis–endemic population. J. Infect. Dis. *182*, 1545–1548.

11. Natividad, A., Hanchard, N., Holland, M.J., Mahdi, O.S., Diakite, M., Rockett, K., Jallow, O., Joof, H.M., Kwiatkowski, D.P., Mabey, D.C., and Bailey, R.L. (2007). Genetic variation at the TNF locus and the risk of severe sequelae of ocular Chlamydia trachomatis infection in Gambians. Genes Immun. *8*, 288–295.

12. Natividad, A., Holland, M.J., Rockett, K.A., Forton, J., Faal, N., Joof, H.M., Mabey, D.C., Bailey, R.L., and Kwiatkowski, D.P. (2008). Susceptibility to sequelae of human ocular chlamydial infection associated with allelic variation in IL10 cis-regulation. Hum. Mol. Genet. *17*, 323–329.

13. Natividad, A., Hull, J., Luoni, G., Holland, M., Rockett, K., Joof, H., Burton, M., Mabey, D., Kwiatkowski, D., and Bailey, R. (2009). Innate immunity in ocular Chlamydia trachomatis infection: contribution of IL8 and CSF2 gene variants to risk of trachomatous scarring in Gambians. BMC Med. Genet. *10*, 138.

14. Natividad, A., Wilson, J., Koch, O., Holland, M.J., Rockett, K., Faal, N., Jallow, O., Joof, H.M., Burton, M.J., Alexander, N.D., et al. (2005). Risk of trachomatous scarring and trichiasis in Gambians varies with SNP haplotypes at the interferon-gamma and interleukin-10 loci. Genes Immun. *6*, 332–340.

15. Ohman, H., Bailey, R., Natividad, A., Ragoussis, J., Johnson, L.L., Tiitinen, A., Halttunen, M., Paavonen, J., and Surcel, H.M. (2012). Effect of IL12A and IL12B polymorphisms on the risk of Chlamydia trachomatis-induced tubal factor infertility and disease severity. Hum. Reprod. *27*, 2217–2223.

16. Roberts, Ch., Franklin, C.S., Makalo, P., Joof, H., Sarr, I., Mahdi, O.S., Sillah, A., Bah, M., Payne, F., Jeffreys, A.E., et al. (2015). Conjunctival fibrosis and the innate barriers to Chlamydia trachomatis intracellular infection: a genome wide association study. Sci. Rep. *5*, 17447.

17. Ohman, H., Tiitinen, A., Halttunen, M., Birkelund, S., Christiansen, G., Koskela, P., Lehtinen, M., Paavonen, J., and Surcel, H.M. (2006). IL-10 polymorphism and cell-mediated immune response to Chlamydia trachomatis. Genes Immun. *7*, 243–249.

18. Ohman, H., Tiitinen, A., Halttunen, M., Lehtinen, M., Paavonen, J., and Surcel, H.M. (2009). Cytokine polymorphisms and severity of tubal damage in women with Chlamydia-associated infertility. J. Infect. Dis. *199*, 1353–1359.

19. Taylor, B.D., Darville, T., Ferrell, R.E., Kammerer, C.M., Ness, R.B., and Haggerty, C.L. (2012). Variants in toll-like receptor 1 and 4 genes are associated with Chlamydia trachomatis among women with pelvic inflammatory disease. J. Infect. Dis. *205*, 603–609.

20. Taylor, B.D., Darville, T., Ferrell, R.E., Ness, R.B., and Haggerty, C.L. (2013). Racial variation in toll-like receptor variants among women with pelvic inflammatory disease. J. Infect. Dis. *207*, 940–946.

21. Darville, T., Andrews, C.W., Jr., Laffoon, K.K., Shymasani, W., Kishen, L.R., and Rank, R.G. (1997). Mouse strain-dependent variation in the course and outcome of chlamydial genital tract infection is associated with differences in host response. Infect. Immun. *65*, 3065–3073.

22. Bernstein-Hanley, I., Balsara, Z.R., Ulmer, W., Coers, J., Starnbach, M.N., and Dietrich, W.F. (2006). Genetic analysis of susceptibility to Chlamydia trachomatis in mouse. Genes Immun. *7*, 122–129.

23. Cotter, T.W., Miranpuri, G.S., Ramsey, K.H., Poulsen, C.E., and Byrne, G.I. (1997). Reactivation of chlamydial genital tract infection in mice. Infect. Immun. *65*, 2067–2073.

24. Alvarez, M.I., Glover, L.C., Luo, P., Wang, L., Theusch, E., Oehlers, S.H., Walton, E.M., Tram, T.T.B., Kuang, Y.L., Rotter, J.I., et al. (2017). Human genetic variation in *VAC14* regulates *Salmonella* invasion and typhoid fever through modulation of cholesterol. Proc. Natl. Acad. Sci. USA *114*, E7746–E7755.

25. Ko, D.C., Gamazon, E.R., Shukla, K.P., Pfuetzner, R.A., Whittington, D., Holden, T.D., Brittnacher, M.J., Fong, C., Radey, M., Ogohara, C., et al. (2012). Functional genetic screen of human diversity reveals that a methionine salvage enzyme regulates inflammatory cell death. Proc. Natl. Acad. Sci. USA *109*, E2343–E2352.

26. Ko, D.C., Shukla, K.P., Fong, C., Wasnick, M., Brittnacher, M.J., Wurfel, M.M., Holden, T.D., O'Keefe, G.E., Van Yserloo, B., Akey, J.M., and Miller, S.I. (2009). A genome-wide in vitro bacterial-infection screen reveals human variation in the host response associated with inflammatory disease. Am. J. Hum. Genet. *85*, 214–227.

27. Wang, L., Pittman, K.J., Barker, J.R., Salinas, R.E., Stanaway, I.B., Williams, G.D., Carroll, R.J., Balmat, T., Ingham, A., Gopalakrishnan, A.M., et al.; eMERGE Network (2018). An Atlas of Genetic Variation Linking Pathogen-Induced Cellular Traits to Human Disease. Cell Host Microbe *24*, 308–323.e6.

28. Cahir-McFarland, E.D., Carter, K., Rosenwald, A., Giltnane, J.M., Henrickson, S.E., Staudt, L.M., and Kieff, E. (2004). Role of NF-kappa B in cell survival and transcription of latent membrane protein 1-expressing or Epstein-Barr virus latency III-infected cells. J. Virol. *78*, 4108–4119.

29. Civelek, M., and Lusis, A.J. (2014). Systems genetics approaches to understand complex traits. Nat. Rev. Genet. *15*, 34–48.

30. Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., McVean, G.A.; and 1000

Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.

31. Gaffney, D.J., Veyrieras, J.B., Degner, J.F., Pique-Regi, R., Pai, A.A., Crawford, G.E., Stephens, M., Gilad, Y., and Pritchard, J.K. (2012). Dissecting the regulatory architecture of gene expression QTLs. Genome Biol. *13*, R7.

32. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., McVean, G.A.; and 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.

33. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R.; and 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.

34. Grubert, F., Zaugg, J.B., Kasowski, M., Ursu, O., Spacek, D.V., Martin, A.R., Greenside, P., Srivas, R., Phanstiel, D.H., Pekowska, A., et al. (2015). Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions. Cell *162*, 1051–1065.

35. Jadhav, B., Monajemi, R., Gagalova, K.K., Ho, D., Draisma, H.H.M., van de Wiel, M.A., Franke, L., Heijmans, B.T., van Meurs, J., Jansen, R., et al.; GoNL Consortium; and BIOS Consortium (2019). RNA-Seq in 296 phased trios provides a high-resolution map of genomic imprinting. BMC Biol. *17*, 50.

36. Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A., Monlong, J., Rivas, M.A., Gonzàlez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al.; Geuvadis Consortium (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature *501*, 506–511.

37. Bard, J., and Levitt, D. (1986). Chlamydia trachomatis (L2 serovar) binds to distinct subpopulations of human peripheral blood leukocytes. Clin. Immunol. Immunopathol. *38*, 150–160.

38. Levitt, D., Danen, R., and Bard, J. (1986). Both species of chlamydia and two biovars of Chlamydia trachomatis stimulate mouse B lymphocytes. J. Immunol. *136*, 4249–4254.

39. Fairfax, B.P., and Knight, J.C. (2014). Genetics of gene expression in immunity to infection. Curr. Opin. Immunol. *30*, 63–71.

40. Flutre, T., Wen, X., Pritchard, J., and Stephens, M. (2013). A statistical framework for joint eQTL analysis in multiple tissues. PLoS Genet. *9*, e1003486.

41. Nica, A.C., and Dermitzakis, E.T. (2013). Expression quantitative trait loci: present and future. Philos. Trans. R. Soc. Lond. B Biol. Sci. *368*, 20120362.

42. Bas, S., Neff, L., Vuillet, M., Spenato, U., Seya, T., Matsumoto, M., and Gabay, C. (2008). The proinflammatory cytokine response to Chlamydia trachomatis elementary bodies in human macrophages is partly mediated by a lipoprotein, the macrophage infectivity potentiator, through TLR2/TLR1/TLR6 and CD14. J. Immunol. *180*, 1158–1168.

43. Buckner, L.R., Lewis, M.E., Greene, S.J., Foster, T.P., and Quayle, A.J. (2013). Chlamydia trachomatis infection results in a modest pro-inflammatory cytokine response and a decrease in T cell chemokine secretion in human polarized endocervical epithelial cells. Cytokine *63*, 151–165.

44. Gervassi, A., Alderson, M.R., Suchland, R., Maisonneuve, J.F., Grabstein, K.H., and Probst, P. (2004). Differential regulation of inflammatory cytokine secretion by human dendritic cells upon Chlamydia trachomatis infection. Infect. Immun. *72*, 7231–7239.

45. Poston, T.B., Lee, D.E., Darville, T., Zhong, W., Dong, L., O'Connell, C.M., Wiesenfeld, H.C., Hillier, S.L., Sempowski, G.D., and Zheng, X. (2019). Cervical Cytokines Associated With Chlamydia trachomatis Susceptibility and Protection. J. Infect. Dis. *220*, 330–339.

46. Darville, T., Andrews, C.W., Jr., Sikes, J.D., Fraley, P.L., and Rank, R.G. (2001). Early local cytokine profiles in strains of mice with different outcomes from chlamydial genital tract infection. Infect. Immun. *69*, 3556–3561.

47. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics *26*, 2336–2337.

48. Turner, S.D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. J. Open Source Softw. *3*, 731.

49. Agaisse, H., and Derré, I. (2013). A C. trachomatis cloning vector and the generation of C. trachomatis strains expressing fluorescent proteins under the control of a C. trachomatis promoter. PLoS ONE *8*, e57090.

50. Saka, H.A., Thompson, J.W., Chen, Y.S., Kumar, Y., Dubois, L.G., Moseley, M.A., and Valdivia, R.H. (2011). Quantitative proteomics reveals metabolic and pathogenic properties of Chlamydia trachomatis developmental forms. Mol. Microbiol. *82*, 1185–1203.

51. Hagège, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., and Forné, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat. Protoc. *2*, 1722–1733.

52. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

53. Purcell, S., Sham, P., and Daly, M.J. (2005). Parental phenotypes in family-based association analysis. Am. J. Hum. Genet. *76*, 249–259.

54. Chen, Z., Boehnke, M., Wen, X., and Mukherjee, B. (2021). Revisiting the genome-wide significance threshold for common variant GWAS. G3 (Bethesda) *11*, jkaa056.

55. Panagiotou, O.A., Ioannidis, J.P.; and Genome-Wide Significance Project (2012). What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. Int. J. Epidemiol. *41*, 273–286.

56. Joyee, A.G., and Yang, X. (2008). Role of toll-like receptors in immune responses to chlamydial infections. Curr. Pharm. Des. *14*, 593–600.

57. Nagarajan, U.M., Ojcius, D.M., Stahl, L., Rank, R.G., and Darville, T. (2005). Chlamydia trachomatis induces expression of IFN-gamma-inducible protein 10 and IFN-beta independent of TLR2 and TLR4, but largely dependent on MyD88. J. Immunol. *175*, 450–460.

58. O'Connell, C.M., Ionova, I.A., Quayle, A.J., Visintin, A., and Ingalls, R.R. (2006). Localization of TLR2 and MyD88 to Chlamydia trachomatis inclusions. Evidence for signaling by intracellular TLR2 during infection with an obligate intracellular pathogen. J. Biol. Chem. *281*, 1652–1659.

59. Wang, Y., Liu, Q., Chen, D., Guan, J., Ma, L., Zhong, G., Shu, H., and Wu, X. (2017). Chlamydial Lipoproteins Stimulate Toll-Like Receptors 1/2 Mediated Inflammatory Responses through MyD88-Dependent Pathway. Front. Microbiol. *8*, 78.

60. Barker, J.R., Koestler, B.J., Carpenter, V.K., Burdette, D.L., Waters, C.M., Vance, R.E., and Valdivia, R.H. (2013). STING-dependent recognition of cyclic di-AMP mediates type I interferon responses during Chlamydia trachomatis infection. MBio *4*, e00018–13.

61. Prantner, D., Darville, T., and Nagarajan, U.M. (2010). Stimulator of IFN gene is critical for induction of IFN-beta during Chlamydia muridarum infection. J. Immunol. *184*, 2551–2560.

62. Cunningham, K., Stansfield, S.H., Patel, P., Menon, S., Kienzle, V., Allan, J.A., and Huston, W.M. (2013). The IL-6 response to Chlamydia from primary reproductive epithelial cells is highly variable and may be involved in differential susceptibility to the immunopathological consequences of chlamydial infection. BMC Immunol. *14*, 50.

63. Buchholz, K.R., and Stephens, R.S. (2006). Activation of the host cell proinflammatory interleukin-8 response by Chlamydia trachomatis. Cell. Microbiol. *8*, 1768–1779.

64. Wurfel, M.M., Gordon, A.C., Holden, T.D., Radella, F., Strout, J., Kajikawa, O., Ruzinski, J.T., Rona, G., Black, R.A., Stratton, S., et al. (2008). Toll-like receptor 1 polymorphisms affect innate immune responses and outcomes in sepsis. Am. J. Respir. Crit. Care Med. *178*, 710–720.

65. Mikacenic, C., Reiner, A.P., Holden, T.D., Nickerson, D.A., and Wurfel, M.M. (2013). Variation in the TLR10/TLR1/TLR6 locus is the major genetic determinant of interindividual difference in TLR1/2-mediated responses. Genes Immun. *14*, 52–57.

66. Hawn, T.R., Misch, E.A., Dunstan, S.J., Thwaites, G.E., Lan, N.T., Quy, H.T., Chau, T.T., Rodrigues, S., Nachman, A., Janer, M., et al. (2007). A common human TLR1 polymorphism regulates the innate immune response to lipopeptides. Eur. J. Immunol. *37*, 2280–2289.

67. Ward, L.D., and Kellis, M. (2016). HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res. *44* (D1), D877–D881.

68. Stranger, B.E., Nica, A.C., Forrest, M.S., Dimas, A., Bird, C.P., Beazley, C., Ingle, C.E., Dunning, M., Flicek, P., Koller, D., et al. (2007). Population genomics of human gene expression. Nat. Genet. *39*, 1217–1224.

69. Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., et al. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat. Genet. *47*, 598–606.

70. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

71. Wang, Y., Song, F., Zhang, B., Zhang, L., Xu, J., Kuang, D., Li, D., Choudhary, M.N.K., Li, Y., Hu, M., et al. (2018). The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol. *19*, 151.

72. Kumasaka, N., Knights, A.J., and Gaffney, D.J. (2016). Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. Nat. Genet. *48*, 206–213.

73. Sheffield, N.C., Thurman, R.E., Song, L., Safi, A., Stamatoyannopoulos, J.A., Lenhard, B., Crawford, G.E., and Furey, T.S. (2013). Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. Genome Res. *23*, 777–788.

74. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., et al. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford) *2017*, bax028.

75. Barton, G.M., and Medzhitov, R. (2003). Toll-like receptor signaling pathways. Science *300*, 1524–1525.

76. Beutler, B.A. (2009). TLRs and innate immunity. Blood *113*, 1399–1407.

77. Fitzgerald, K.A., and Kagan, J.C. (2020). Toll-like Receptors and the Control of Immunity. Cell *180*, 1044–1066.

78. Kawasaki, T., and Kawai, T. (2014). Toll-like receptor signaling pathways. Front. Immunol. *5*, 461.

79. Libermann, T.A., and Baltimore, D. (1990). Activation of interleukin-6 gene expression through the NF-kappa B transcription factor. Mol. Cell. Biol. *10*, 2327–2334.

80. Kawai, T., and Akira, S. (2007). Signaling to NF-kappaB by Toll-like receptors. Trends Mol. Med. *13*, 460–469.

81. Ishikawa, H., and Barber, G.N. (2008). STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. Nature *455*, 674–678.

82. Takeuchi, O., Kawai, T., Mühlradt, P.F., Morr, M., Radolf, J.D., Zychlinsky, A., Takeda, K., and Akira, S. (2001). Discrimination of bacterial lipoproteins by Toll-like receptor 6. Int. Immunol. *13*, 933–940.

83. Takeuchi, O., Sato, S., Horiuchi, T., Hoshino, K., Takeda, K., Dong, Z., Modlin, R.L., and Akira, S. (2002). Cutting edge: role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins. J. Immunol. *169*, 10–14.

84. Owusu-Edusei, K., Jr., Chesson, H.W., Leichliter, J.S., Kent, C.K., and Aral, S.O. (2013). The association between racial disparity in income and reported sexually transmitted infections. Am. J. Public Health *103*, 910–916.

85. Gaskin, D.J., Dinwiddie, G.Y., Chan, K.S., and McCleary, R. (2012). Residential segregation and disparities in health care services utilization. Med. Care Res. Rev. *69*, 158–175.

86. Marcus, J.H., and Novembre, J. (2017). Visualizing the geography of genetic variants. Bioinformatics *33*, 594–595.

87. Schott, B.H., Antonia, A.L., Wang, L., Pittman, K.J., Sixt, B.S., Barnes, A.B., Valdivia, R.H., and Ko, D.C. (2020). Modeling of variables in cellular infection reveals CXCL10 levels are regulated by human genetic variation and the Chlamydia-encoded CPAF protease. Sci. Rep. *10*, 18269.

88. Field, Y., Boyle, E.A., Telis, N., Gao, Z., Gaulton, K.J., Golan, D., Yengo, L., Rocheleau, G., Froguel, P., McCarthy, M.I., and Pritchard, J.K. (2016). Detection of human adaptation during the past 2000 years. Science *354*, 760–764.

89. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al.; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature *590*, 290–299.

90. Vernot, B., and Akey, J.M. (2014). Resurrecting surviving Neandertal lineages from modern human genomes. Science *343*, 1017–1021.

91. Dannemann, M., Andrés, A.M., and Kelso, J. (2016). Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. Am. J. Hum. Genet. *98*, 22–33.

92. Gittelman, R.M., Schraiber, J.G., Vernot, B., Mikacenic, C., Wurfel, M.M., and Akey, J.M. (2016). Archaic Hominin Admixture Facilitated Adaptation to Out-of-Africa Environments. Curr. Biol. *26*, 3375–3382.

93. Barreiro, L.B., Ben-Ali, M., Quach, H., Laval, G., Patin, E., Pickrell, J.K., Bouchier, C., Tichit, M., Neyrolles, O., Gicquel, B., et al. (2009). Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. PLoS Genet. *5*, e1000562.

94. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

95. Zhang, F., and Lupski, J.R. (2015). Non-coding genetic variants in human disease. Hum. Mol. Genet. *24* (R1), R102–R110.

96. Kirsten, H., Al-Hasani, H., Holdt, L., Gross, A., Beutner, F., Krohn, K., Horn, K., Ahnert, P., Burkhardt, R., Reiche, K., et al. (2015). Dissecting the genetics of the human transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding loci. Hum. Mol. Genet. *24*, 4746–4763.

97. Wolfe, N.D., Dunavan, C.P., and Diamond, J. (2007). Origins of major human infectious diseases. Nature *447*, 279–283.

98. Jobling, M.A., Hurles, M.E., and Tyler-Smith, C. (2004). Human Evolutionary Genetics: Origins, Peoples & Disease (New York: Garland Science).

99. Ozinsky, A., Underhill, D.M., Fontenot, J.D., Hajjar, A.M., Smith, K.D., Wilson, C.B., Schroeder, L., and Aderem, A. (2000). The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors. Proc. Natl. Acad. Sci. USA *97*, 13766–13771.

100. Kawai, T., and Akira, S. (2010). The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nat. Immunol. *11*, 373–384.