HHS Public Access

Author manuscript

Nat Med. Author manuscript; available in PMC 2018 August 01.

Published in final edited form as:

Nat Med. 2017 August ; 23(8): 975–983. doi:10.1038/nm.4370.

SNP-mediated disruption of CTCF binding at the IFITM3 promoter is associated with severe influenza risk in humans

E. Kaitlynn Allen[1], **Adrienne G. Randolph**[2], **Tushar Bhangale**[3], **Pranay Dogra**[1], **Maikke Ohlson**[4], **Christine M. Oshansky**[1,9], **Anthony E. Zamora**[1], **John P. Shannon**[1], **David Finkelstein**[5], **Amy Dressen**[3], **John DeVincenzo**[6,7], **Miguela Caniza**[8], **Ben Youngblood**[1], **Carrie M. Rosenberger**[4], and **Paul G. Thomas**[1,*]

[1]Department of Immunology, St. Jude Children's Research Hospital, Memphis, TN

[2]Department of Anesthesiology, Perioperative and Pain Medicine, Boston Children's Hospital, and Departments of Anesthesia and Pediatrics, Harvard Medical School, Boston, MA

[3]Human Genetics, Genentech, Inc., South San Francisco, CA

[4]Biomarker Discovery, Genentech, Inc., South San Francisco, CA

[5]Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN

[6]Departments of Pediatrics; Microbiology, Immunology and Molecular Biology; University of Tennessee School of Medicine; Memphis, TN USA

[7]Children's Foundation Research Institute; Le Bonheur Children's Hospital; Memphis, TN USA

[8]Department of Infectious Diseases, St. Jude Children's Research Hospital

Abstract

Previous studies reported associations of *IFITM3* SNP rs12252 with severe influenza, but evidence of association and the mechanism of risk remains controversial. We prioritized SNPs in *IFITM3* based on putative biological function and identified rs34481144 in the 5′ UTR. We found evidence of a novel association of rs34481144 with severe influenza in three influenza-infected cohorts characterized by different levels of influenza illness severity. We determined the role of rs34481144 as an expression quantitative trait loci (eQTL) for *IFITM3*, with the risk allele associated with lower mRNA expression. The risk allele was found to have decreased IRF3 binding and increased CTCF binding in promoter-binding assays, and risk allele carriage diminished transcriptional correlations among neighboring genes, indicative of CTCF boundary

[*]Correspondence to: Dr. Paul G. Thomas, Department of Immunology, St. Jude Children's Research Hospital, 262 Danny Thomas Place/MS 351, Memphis, TN 38105-3678, (901)595-6507 (phone), paul.thomas@stjude.org.

[9]Present address: Biomedical Advanced Research and Development Authority (BARDA), Office of the Assistant Secretary for Preparedness and Response (ASPR), U. S. Department of Health and Human Services (DHHS), Washington, DC USA

Author contributions

EKA, CMR, and PGT wrote the manuscript and designed figures. EKA, AGR, TB, PD, CMO, BY, CMR, and PGT designed experiments. EKA, PD, MO, CMR conducted experiments. CMO, AEZ, and JPS acquired data. EKA, TB, MO, AEZ, DF, AD, CMR, PGT analyzed data. EKA, AGR, TB, PD, BY, CMR, and PGT interpreted data. AGR oversaw recruitment, data acquisition, and analysis of PICFlu. JD and MC oversaw recruitment of FLU09. AGR, TB, PD, MO, BY edited paper. All authors approved final manuscript.

Competing financial interests

CMR, TB, MO, and AD are employees of Genentech, Inc.

activity. Furthermore, the risk allele disrupts a CpG site that undergoes differential methylation in CD8 T-cell subsets. Carriers of the risk allele had reduced CD8 T-cells in their airways during natural influenza infection, consistent with IFITM3 promoting airway CD8 T-cell accumulation, indicating that a critical function for IFITM3 may be to promote immune cell persistence at mucosal sites. Our study identifies a new regulator of IFITM3 expression that associates with CD8 T-cell levels in the airways and a spectrum of clinical outcomes.

Influenza A virus (IAV) infections can cause major morbidity and mortality in human populations[1], especially with the emergence of novel pandemic strains. In the 2009 H1N1 pandemic, while the infection was considered to be generally mild, a number of pregnant women and otherwise fit young adults were severely afflicted. Having a screening marker that could be used to identify patients at high risk of severe illness would thus be of immense benefit for prioritizing treatment.

The antiviral protein IFITM3 inhibits viral entry into the host cell cytoplasm[2,3]. IFITM3 also promotes the survival of murine lung-resident $CD8^+$ T-cells after IAV challenge, which may help clear the infection[4]. Given the importance of IFITM3 in animal studies and *in vitro* models[5,6], several groups investigated polymorphisms in IFITM3 for associations with disease outcome. The C/C genotype of rs12252 in *IFITM3* was linked to severe illness in adults during the 2009 IAV pandemic[7–9], but the mechanism for this risk allele remains unresolved. The rs12252 C allele correlation with influenza severity has been validated in two cohorts of Han Chinese, where the majority carry at least one C allele[7,10], but not in cohorts of European ancestry[11–13], suggesting that the effects of this allele may vary across ancestral populations. While its role in human disease may be ambiguous, IFITM3 remains a strong candidate for a protein influencing influenza severity based on the increased susceptibility of IFITM3-deficient mice to influenza disease and its characterized activity as a viral restriction factor.

In this study, we sought to define a biological mechanism underlying a genetic risk factor associated with severe clinical outcomes in influenza infected patients who were previously healthy. We report a novel association of SNP rs34481144 with severe influenza illness across three populations with varying degrees of severity highlighting the broad public health potential of this SNP as a screening marker. We also define the mechanism by which rs34481144 influences the regulation of IFITM3 levels, the levels of which correlate with CD8+ T-cell numbers at the site of infection.

Results

SNP in the promoter of IFITM3 is associated with severity of influenza infection in three human cohorts

Several lines of evidence have suggested a role for *IFITM3* in restricting influenza infection, including a previously reported SNP in *IFITM3* (rs12252). However, rs12252 has disputed evidence of association across cohorts and lacks a mechanism by which it confers risk. Thus, to identify other polymorphisms in *IFITM3* that may contribute to influenza disease risk, we evaluated *IFITM3* and its core promoter (Fig. 1a). After SNP prioritization described in

detail in Supplementary Fig. 1 and Methods, we found that rs34481144 satisfied our criteria. Population allele frequencies from 1000 Genomes Project[14] for rs12252 and rs34481144 are listed in Supplementary Table 1, and linkage disequilibrium (LD) between these two SNPs in these populations is included in Supplementary Table 2. We genotyped rs34481144 in influenza positive individuals from the FLU09 cohort of naturally acquired influenza infection in Memphis, TN[12] (Table 1). The FLU09 cohort was a longitudinal study of naturally acquired influenza infection with patient severity defined using clinical and symptom data collected throughout infection (Fig. 1b). When we tested rs34481144, we found evidence of association of the A allele with the total symptom score in FLU09 patients (P=0.037; OR=4.131; Fig. 1c) and a similar directional observation with lower respiratory symptoms (P=0.082; OR=6.89). Haplotype analysis established that rs34481144 and rs12252 are not in high LD (r^2=0.031, D′=1). LD data for rs12252 and rs34481144 demonstrated that the C/C genotype of rs12252 (risk) was always inherited with the G/G genotype of rs34481144 (protective) in the FLU09 study, but the G/G genotype of rs34481144 was also co-inherited with other rs12252 genotypes (C/T and T/T) indicating that the risk alleles were inherited on opposite haplotypes. Importantly, the risk allele of rs34481144 was always inherited with the protective allele of rs12252 in the FLU09 cohort, suggesting that these SNPs provide non-overlapping indications of risk.

We assessed the association of rs34481144 with clinical metrics in individuals from two additional influenza cohorts (Table 1 and Fig. 1b) at opposite ends of the influenza severity spectrum[15]. The Genentech challenge study, an influenza challenge study of healthy adults, assessed whether the rs34481144 risk allele is associated with influenza viral load. While data from studies using Ifitm3-null mice and cells support the role of IFITM3 in restricting viral replication[8], this challenge cohort presented the unique opportunity to examine the relationship between rs34481144 and early viral replication kinetics in human infection. Rapid replication was defined as detectable virus in nasopharyngeal swabs by PCR or TCID50 assay less than 24 hours from inoculation, while slow viral replication was defined as requiring 24 hours or longer. The A risk allele of rs34481144 was significantly associated with rapid viral replication in this challenge study (P=0.01; OR=3.6; Fig. 1c), supporting the association of this allele with poor outcomes and directly relating carriage of the risk allele with IFITM3's presumed function.

We sought further validation in a pediatric influenza cohort of critically ill children in the multicenter Pediatric Intensive Care Influenza (PICFlu) cohort[16]. In PICFlu patients positive for influenza infection with no predisposing conditions, the A allele of rs34481144 was associated with mortality (P=0.047; OR=2.0; Fig. 1c). The majority of the patients with fatal infections (14/17) carried at least one A allele suggesting that the A allele of rs34481144 increases risk of severe influenza illness. Due to the known differences in SNP carriage between Asian populations and European populations, we analyzed a subset of PICFlu patients who were Caucasian. In this analysis, the association of rs34481144 with mortality was strengthened (P=0.036; OR=2.43) with all patients with fatal infections carrying at least one A allele. A meta-analysis pooling all 46 severe and 347 mild individuals from these three cohorts resulted in an odds ratio of 2.6 for severe outcomes (P=0.00024). Notably, all individuals included in the association analyses of the three cohorts were influenza-infected.

The influenza risk-associated SNP rs34481144 modulates expression of IFITM3

Given that rs34481144 resides within the promoter of *IFITM3*, we next sought to test if it modulated transcription of *IFITM3* in cohorts and cell types of interest, as it was found to be an eQTL in blood in the GTEX portal[17]. The region of the promoter surrounding rs34481144 is characterized by a DNaseI hypersensitive region, H3K4me3 and H3K27Ac peaks, and multiple transcription factor binding sites (Fig. 1a). To determine if rs34481144 is associated with expression of *IFITM3* in our FLU09 cohort, we performed an eQTL analysis using PBMC expression data from 45 influenza-positive individuals. This analysis indicated that the genotype of rs34481144 is associated with expression of *IFITM3* (Fig. 2a). Carriage of a single risk allele (A) was significantly associated with lower levels of *IFITM3* expression (P=0.008). Only one individual with RNAseq data was homozygous for the A allele and this individual fell outside the 99% confidence interval of expression for the G/G genotype.

We next examined eQTL status of rs34481144 in a population of healthy donors recruited at the Broad Institute for the Phenogenetic Project and ImmVar Consortium[18] (Fig. 2b). Monocyte-derived dendritic cells from 200 subjects were stimulated with LPS, IFNβ, or influenza for 6 hours. In all unstimulated and stimulated cells, there was significantly reduced expression of *IFITM3* in subjects carrying the A risk allele (unstimulated $P=4\times10^{-7}$ Fig. 2b; LPS $P=2\times10^{-12}$, IFNβ $P=4.5\times10^{-19}$, and Influenza $P=3.5\times10^{-16}$ Supplementary Fig. 2a). We also performed an eQTL analysis using *IFITM3* levels and the rs12252 genotype but detected no association of rs12252 genotype and *IFITM3* expression except in IFNβ-stimulated cells, which show a modest association between the risk C/C genotype of rs12252 and higher levels of *IFITM3* (Supplementary Fig. 2b). Finally, this eQTL was validated in the PICFlu study of critically ill children, and it was again shown that the risk genotype of rs34481144 was associated with lower *IFITM3* expression (P=0.0003; Fig. 2c). Collectively, data from three independent cohorts provide evidence that the risk allele of rs34481144 was associated with lower expression of *IFITM3,* providing a potential mechanism for the observed higher risk of severe influenza illness[5,19].

rs34481144 genotype determines IFITM3 promoter activity

Having demonstrated that rs34481144 resides within an *IFITM3* transcriptional regulatory region (Fig. 1a), we proceeded to determine if rs34481144 could directly alter *IFITM3* promoter activity. Two luciferase constructs were created with the *IFITM3* core promoter[20] (Fig. 1b & 2d) driving expression, one with the A allele referred to as the risk promoter and the other with the G allele referred to as the protective promoter, and then transfected into Jurkat and HEK293T cell lines. In Jurkat cells, both unstimulated (P=0.02; Fig. 2e) and ConA-stimulated cells (Supplementary Fig. 3a; P<0.0001) exhibited higher promoter activity when the G allele was present in the promoter. After poly(I:C) and IFNα stimulation or influenza infection of HEK293T cells, activity of the protective promoter was consistently higher compared to the risk promoter, at both 4 hours (P<0.0001; Supplementary Fig. 3b) and 24 hours (P<0.0001; Fig. 2f). The difference in expression between the two genotypes was of an equivalent or larger magnitude than the induction of expression by any of the observed stimuli, indicating that genotype has a greater effect on expression than activation of innate immune signaling, though this is a minimal promoter and may not represent the

complete response of the full *IFITM3* promoter. These data suggest that the expression of *IFITM3* is influenced by rs34481144 genotype, with individuals carrying the A allele having reduced promoter activity.

rs34481144 genotype governs transcription factor binding

To define the mechanism for SNP modulation of *IFITM3* promoter activity, we sought to characterize genotype-specific changes in protein binding to the promoter. Electrophoretic mobility shift assays (EMSAs) were performed by incubating nuclear extract of infected HEK293T or stimulated Jurkat cells with biotinylated, 40bp double-stranded DNA probes that included either the A or G allele at rs34481144. There was overall less protein binding to the probe containing the A allele compared to the G allele in Jurkat cells (Fig. 3a) and HEK293T cells (Fig. 3b), even at early time points of infection (Supplementary Fig. 4a). Thus, the change in promoter activity based on genotype is likely due to differential binding of proteins affecting expression levels of *IFITM3*.

Because the G allele was associated with higher *IFITM3* expression, promoter activity, and protein binding, specific transcription factors involved in the innate immune response were subsequently tested for binding to this SNP. In particular, we were interested in IRF3 due to its known role in activating innate immune responses[21,22] and controlling host responses to infection[23]. Western-blotting EMSA (WEMSA) assays were used to detect specific proteins that bound to the EMSA probes encompassing rs34481144. WEMSA data show that IRF3 bound preferentially to the G allele compared to the risk allele (Fig. 3c; Supplementary Fig. 4b), consistent with the idea that these factors promote interferon-dependent antiviral immunity. STAT1, another critical transcriptional regulator of the interferon response, was found to bind equally between the two genotypes (Supplementary Fig. 4c).

Our findings suggest fewer positive regulators were binding the risk allele A, so we investigated whether inhibitory factors may preferentially bind the risk allele. Specifically, CTCF has many roles in regulating gene expression as an insulator and in its role in chromatin architecture[24–26]. Data published in ENCODE[27] using multiple cell lines including small airway epithelial cells (SAEC) indicated that CTCF binds to this region[28], so we tested CTCF binding affinity to the IFITM3 promoter. CTCF bound more abundantly to the promoter containing the A allele compared to the G (Figs. 3d & 3e; Supplementary Fig. 4d). Further evidence of CTCF binding preference in each promoter was found using the predictive binding tool CTCFBSDB 2.0[29] indicating that the 9bp sequence including the risk allele (ACCCGACC[A/G]) was almost two times more likely to be a CTCF binding site than the sequence containing the G allele (A allele PWM score=76.1; G allele PWM score=40.1; Fig. 1b). These data illustrate a shift in the binding based on genotype, with transcription factor occupancy tipped from a transcriptionally active profile with IRF3 binding in the protective promoter to an inactive profile in the presence of CTCF binding in the risk promoter.

Methylation status of rs34481144 differs between cell types

CTCF promoter binding is regulated both by DNA sequence and by DNA methylation, which reduces CTCF binding affinity[28]. In human isocitrate dehydrogenase mutant gliomas,

hypermethylation compromises the binding site of CTCF leading to increased and aberrant gene expression[30]. The protective allele of rs34481144 (G) creates a CpG site that the risk allele (A) disrupts; therefore, we hypothesized that by abolishing a CpG site, the A allele removes the potential for DNA methylation and increases CTCF binding affinity at this promoter, as an additional mechanism that reduces *IFITM3* expression levels.

We first sought to determine if rs34481144 was methylated in cell populations of interest including epithelial cells and monocytes. We examined the methylation status of the CpG site created by the G allele of rs34481144 in differentiated normal human bronchial epithelial (NHBE) cells after 24 hours of infection with influenza from three donors with at least one G allele at rs34481144 by locus-specific bisulfite sequencing. The CpG site in this subset of cells was fully demethylated (Fig. 4a; Supplementary Fig. 5a). Monocytes were also sorted from six FLU09 patients, and we found that regardless of genotype, the region was also demethylated (Supplementary Fig. 5b–c). Therefore, methylation does not appear to be a mechanism to alter occupancy of this promoter in myeloid cells and infected lung epithelial cells. In mice, lung resident memory CD8+ T-cells have increased *Ifitm3* expression and this expression is required for survival after secondary influenza infection, demonstrating that *IFITM3* in memory CD8+ T-cells acts as an important antiviral resistance mechanism by promoting adaptive immunity[4]. Based on these findings, we hypothesized that memory cells from FLU09 patients would have increased methylation at rs34481144 to block CTCF binding, allowing for increased expression of *IFITM3*. After sorting naïve (T_N) and effector memory (T_{EM}) CD8+ T-cells from FLU09 patients with A/G or G/G genotypes (Supplementary Fig. 5b), we measured methylation at rs34481144 and found differences in the methylation status among cell subsets. T_{EM} cells were more often methylated at the CpG of interest than T_N cells in individuals with the G/G genotype (P=0.03; CpG 2 in Fig. 4a; Supplementary Fig. 5d–f; Complete methylation data for all cell types investigated found in Supplementary Fig. 5.). These data demonstrate differential methylation at the rs34481144 locus is correlated with human effector and memory CD8 T-cell differentiation states.

To investigate the effect of DNA methylation at this locus on CTCF binding, we performed a WEMSA experiment assessing the ability of purified CTCF to bind to methylated versus unmethylated probes. CTCF binding affinity was greatest with the unmethylated A probe, while methylation of the A probe (in which there are two other CpG sites within the probe sequence) reduced binding partially, followed by the unmethylated G probe, with the least amount of binding on the methylated G probe (Fig. 4b).

While we had shown that rs34481144 was an eQTL for *IFITM3* transcript in total PBMCs (Figure 2), the observation that rs34481144 was differentially methylated in CD8+ T-cells led us to measure protein levels of IFITM3 in this cell type. Consistent with the importance of rs34481144 genotype on *IFITM3* expression, we found lower levels of IFITM3 protein in CD8+ T-cells from individuals with the risk A/A genotype compared to the protective G/G genotype, with individuals with the A/G genotype having an intermediate IFITM3 protein level (Fig. 4c; Supplementary Fig. 6). Interestingly, A549 cells and 293T cells are both A/A genotype as well and had similarly low levels of endogenous IFITM3 as PBMC from individuals with the A/A genotype (Supplementary Fig. 7).

Because of our finding that CD8+ T-cells were differentially methylated at rs34481144, we next sought to determine whether or not genotype had an effect on the number or frequency of lymphocytes in nasal washes of FLU09 patients. There were significantly more CD8+ T-cells in the nasal washes of influenza infected individuals with the G/G genotype compared to the A/G genotype (P=0.012; Fig. 4d), and this correlation was also seen in the frequency of CD8+ T-cells among all viable cells (P=0.009; Fig. 4d). This seems to be a CD8+ specific effect, as the correlation did not track with CD4+ T-cell number (P=0.21) or frequency (P=0.07; Supplementary Fig. 8a–b). Notably, when we used linear regression to analyze the correlation between frequency of CD8+ T-cells and rs34481144 genotype, the correlation was significant (P=0.04) even when adjusted for viral titer at day 0 and age at enrollment (P=0.04). We have observed that after resolution of infection and in uninfected patients, CD8+ T-cells are not present in the nasal wash, but CD8+ T-cells in the nasal wash are tetramer-specific at the peak of infection (data not shown). These data provide support for a mechanism whereby methylation at rs34481144 reduces CTCF binding, leading to elevated IFITM3 levels, resulting in enhanced accumulation of effector CD8+ T-cells in the airways.

Transcriptional regulation of IFITM3 neighboring genes is altered by genotype

CTCF binding is known to alter chromatin topology, changing proximity of promoters and enhancers and leading to correlated changes in transcriptional levels of genes within a broad locus, or domains[31,32]. Domain boundaries can be deduced by gene expression correlation changes within a domain region; correlation is reduced or lost when the domain is interrupted[30,33]. To determine whether expression correlation for gene pairs changed with rs34481144 genotype, we analyzed the expression levels of genes <300kb from rs34481144 to encompass any potential domain and surrounding region using RNAseq data from the FLU09 cohort to see if expression correlations of gene pairs changed with rs34481144 genotype. There were correlated gene expression levels between individuals with the G/G genotype (Fig. 4e) that were absent in individuals with the A/G genotype (Fig. 4f), indicative of a change in a boundary likely due to the effects of differential occupancy of CTCF based on rs34481144 genotype, as has been reported for other SNPs[34]. Expression correlation between these genes and *IFITM3* was subsequently tested individually, and we found the same domain region differences (Supplementary Fig. 9, Supplementary Table 3). The expression of all IFITM family members appear strongly correlated to each other regardless of genotype. We found significant expression correlations for *IFITM1* and *IFITM2* with *IFITM3* (Supplementary Fig. 10a) as well as at protein level correlations for IFITM1 and IFITM3 (Supplementary Fig. 10b), suggesting that regulatory processes identified with a single *IFITM* gene could be compounded by influencing the expression (either up or down) of all *IFITM* genes.

Discussion

Here we report that rs34481144 is associated with *IFITM3* gene expression, the methylation status of the CpG in specific cell subsets, IFITM3 protein levels directly in CD8+ T-cells, and the transcriptional landscape of the broader region including immune genes. We also identify a novel association of rs34481144 with severe influenza illness in human cohorts. We further show an association with CD8+ T-cell levels in the airways of influenza-infected

patients. The consistency of our findings across independent cohorts of varying disease severity and the correlation with CD8+ T-cell number in the airways suggest that screening for rs34481144 could be useful for identifying individuals at particular risk for severe outcomes. These results suggest that rs34481144 may have clinical implications for severe illness outcome in individuals such as those enrolled in the PICFlu cohort, who were selected based on the absence of known risk factors. Such individuals are unlikely to be categorized as “at higher risk” in the event of a pandemic.

A major limitation of the genetic analysis in this study is the size of the cohorts used for association of rs34481144 with severe influenza illness. In order to use the most applicable population for our question, we required that all individuals included in the association analyses be infected with influenza virus with cases and controls determined by the severity of illness. We argue that this is a more informative methodology for identifying underlying genetic factors contributing to risk of severe illness than comparing influenza infected cases to healthy donor controls whose history of response to infection is unknown. Cohorts with extensive influenza viral and clinical data are costly and difficult to accrue, which limited us to small numbers. We have comprehensive phenotyping of all subjects in these three cohorts, which we suggest provides further confidence for the association. Moreover we suggest our functional assays strengthen the argument for a clinical association by detailing how this SNP directly impacts expression and promoter activity of *IFITM3*, influences binding of transcription factors partially through methylation status, and is correlated with CD8+ T-cell levels in the airways of influenza-infected patients. However, validation of this association in a large, well powered cohort of the same case/control structure is a goal for our future studies.

Use of the Genentech Challenge study population allowed us to investigate the effect of the SNP on viral kinetics, a known innate mechanism by which IFITM3 can abate influenza replication. Intriguingly, while IFITM3 is conventionally only thought to play a critical role in innate restriction of viral replication, our results suggest that this locus is specifically regulated in activated and memory CD8+ T-cell subsets. Mouse models support a role for IFITM3 in protecting CD8+ T-cells in the airways and the phenotype of increased viral titer does not appear in knockout animals until after day 5, a time point associated with adaptive immune control[4,8]. In most human cohorts, including our own, the anti-influenza immune response is a secondary recall response for T-cells as individuals usually have been infected or vaccinated with influenza by the age of two. Thus, modulators of innate immunity that can promote robust adaptive recall responses may be critically important in influencing the outcome of this infection.

The genotype of rs34481144 influences CTCF binding and was found to be correlated with expression at the broader locus surrounding *IFITM3,* including several genes known to be involved in host responses to viral infection. Due to the differences in allele frequencies of *IFITM3* SNPs among ancestral populations, further study is necessary to investigate differences in the distribution of transcription factor binding in this region that affect regulation of the broader locus. Indeed, the minor allele of rs12252 is carried at much higher frequencies in Han Chinese and other populations of Asian descent when compared to rs34481144 and vice versa in populations of European descent. It is possible that these two

SNPs correlate with differentially organized chromatin domains. Additionally, given the known diverse effects of genes in this locus on immunological outcomes, we would expect other infections to be influenced by polymorphisms in this region. The minor allele of rs12252 has been associated with HIV progression in a Chinese cohort[35]. However, the predicted truncation mutant mediated by rs12252[8], which has not been observed in published studies, was able to restrict HIV replication[36], so this association may indicate a role for other effects of this locus.

One gene in this locus that showed a strong negative correlation with *IFITM3* expression in individuals with the G/G genotype is *B4GALNT4*, which produces an acetylgalactosaminyltransferase that has a role in O-glycosylation and mucin production, an important process in the host response to viral infection[37]. Currently, there is no literature on the role that B4GALNT4 may play in host response to viral infection, but it is known that mice deficient in B4GALNT3, a paralog of B4GALNT4, display significantly reduced influenza virus replication during the early stages of infection[38]. This example illustrates that the complex regulation of this locus may affect multiple antiviral host pathways, warranting additional study in response to multiple infections. This example also illustrates that evolutionary pressure for keeping the risk allele of rs34481144 may be due to the effects of another gene or genes in the broader locus that are important in a non-influenza context.

Taken together, our results suggest that differential rs34481144 methylation in CD8+ T-cells influences IFITM3 levels and may contribute to the accumulation of this important cell type at the site of influenza infection. As IFITM3 has been reported to have roles in epithelial cells[6], endothelial cells[39], and dendritic cells[40], we will explore its regulation in these cells in future studies. The current analysis indicates that the innate factor IFITM3 may play a key role in regulating the cellular response and severity of influenza infection at least in part through its expression in CD8 T-cells.

Methods

Study participants and design

The FLU09 cohort has been previously described[41] with additional recruitment for Year 5 of the FLU09 study. Briefly, inclusion criteria required that participants meet the clinical case definition of influenza virus infection at the time of enrollment, or were asymptomatic household contacts of a participant with confirmed influenza infection. Exclusion criteria included refusal to consent to nasal lavage, history (3 months prior to enrollment) of having received immunoglobulin or other blood products, and having a condition that would predispose the participant to unacceptable injury or render the participant unable to meet requirements of the study. Inclusion and exclusion criteria were established before enrollment of the participants. This study was conducted in compliance with 45 CFR46 and the Declaration of Helsinki. Institutional Review Boards of St. Jude Children's Research Hospital and the University of Tennessee Health Science Center/Le Bonheur Children's Hospital approved the study. Written, informed consent was acquired from participants' parents/guardians, and written assent from age-appropriate subjects was acquired at the time of enrollment. Index cases were asked to provide nasal swabs, nasal lavages, and blood on the day of enrollment (Day 0) and Days 3, 7, 10, and 28, whereas household contacts were

asked to provide nasal swabs on Day 0, 3, 7, and 14 and blood and nasal lavages on Days 0 and 28. The population used for these analyses was predominantly African American (81.4%) with 18.6% Caucasian participants (N=86). Meta data collected from this study included several symptoms that were ranked daily (self-reported) according to a visual analog scale. These symptoms included systemic (fever, fatigue, headache, body ache, chills, lethargy, and irritability), upper respiratory (sore throat, stuffy nose, earache, and sinus pressure), lower respiratory (cough, shortness of breath, and wheezing), and gastrointestinal symptoms (nausea, vomiting, and diarrhea). FLU09 participants were stratified into mild and severe patient groups by the total peak symptom score during infection with severe patients having a total peak score greater than 600. Severe lower respiratory tract symptom status was defined as the presence of shortness of breath and wheezing.

In the Genentech challenge study, healthy volunteers between 18 to 45 years of age and seronegative in hemagglutination inhibition (HAI) assays towards A/Wisconsin/67/2005 (H3N2) IAV were enrolled and provided written informed consent. The London, City and East Research Ethics Committee approved the study protocol. This trial (www.clinicaltrial.gov identifier number NCT01980966) was registered and conducted in full conformance with the ICH E6 guidelines for Good Clinical Practice, the EU Good Clinical Practice Directive (2005/28/EC), the Declaration of Helsinki, and applicable laws and regulations. While the study was an interventional cohort, we were not analyzing or reporting any data from the intervention, but only from the control period prior to the application of the intervention. The patient population of this study was predominantly of European ancestry (94.9%), with a small percentage self-identified as Asian (1.7%) or African (3.4%). Enrolled volunteers received an intranasal inoculation of approximately 5.0–5.5 log10 median tissue culture infective dose ($TCID_{50}$) of influenza A/Wisconsin/67/2005 H3N2 (Baxter; Orth an der Donau, Austria). In this study, we are not reporting the primary endpoint for this trial (an interventional study testing a specific agent), but rather a secondary analysis of the control individuals in the study to test our hypothesis of genetic regulation of viral control. 47 participants became positive for viral infection, based on virus detectable in nasopharyngeal swabs by qRT-PCR and/or TCID50 in the 7 days following inoculation and of those 42 passed QC for subsequent genetics analysis. Rapid replication was defined as detectable virus <24 hours from inoculation, while slow viral replication was defined as requiring ≥24 hours for nasopharyngeal swabs to be positive by PCR or TCID50 assay.

The PICFlu Study was a multicenter study of influenza critical illness in North American children admitted across 31 pediatric intensive care units (PICUs)[16]. Site Institutional Review Boards approved the study. Informed consent was obtained from at least one parent. Exclusion criteria has been described previously in detail; briefly, children were excluded for certain pre-existing lung disorders, immunocompromised status, mitochondrial disorders that could predispose to severe infection, certain genetic or neurologic disorders, preexisting cardiac disease that could affect lung function, or influenza infection considered to be acquired nosocomially by hospital personnel. Inclusion and exclusion criteria were established before enrollment of the participants. PICFlu investigators enrolled 316 influenza positive children with confirmed influenza infection who presented to the PICU with critical illness (acute respiratory failure and/or vasopressors), 265 of which had no

predisposing conditions known to increase risk for such severe illness. The population was predominantly of European ancestry (73.2%) with 13.2% African Americans, 4.9% Asian, and 8.7% mixed or other. The most severe phenotype for influenza illness was tested for association with rs34481144 genotype in this cohort, mortality. Out of the 265 patients enrolled in the study, 17 died and 248 survived.

Prioritization and Sanger sequencing of IFITM3 SNPs

In order to prioritize SNPs for genotyping in the FLU09 cohort, we followed the scheme illustrated in Fig. 1a. In brief, common SNPs (MAF>1%) within the *IFITM3* gene and core promoter[20] were tested for their status as eQTLs using the GTEx portal[17]. Those SNPs that were eQTLs for *IFITM3* were then investigated for allele frequency differences between Asian (CHB) and European (CEU) populations using 1000 Genomes Project data[14]. This step was necessary in order to tease apart the debate for rs12252 association in populations of European descent. Following this filter, we then prioritized based on regulatory potential using ENCODE data[27] in the UCSC Genome Browser. From this scheme, rs34481144 was determined as the SNP with highest priority for genotyping in FLU09. GTEX data determined rs34481144 was an eQTL for *IFITM3* (*P*=2.6E-7) in the blood with the risk allele being associated with lower *IFITM3* expression.

All individuals from the three cohorts described above were genotyped by one of two methods. The first method, used for samples from both FLU09 and the Genentech Challenge study, amplified a fragment of the *IFITM3* gene on chromosome 11 from purified patient DNA by a touchdown PCR approach using previously published primers[7] (forward: 5′-GGAAACTGTTGAGAAACCGAA-3′ and reverse: 5′-CATACGCACCTTCACGGAGT-3′), where the annealing temperature was reduced each cycle from 70°C by 1°C for 10 cycles, and then repeated at 60°C for 25 additional cycles. The resulting 352 bp PCR fragment was purified using a QIAquick PCR Purification kit (Qiagen) and sequenced by Sanger sequencing using the following forward and reverse primers, respectively: 5′-ACAGCCACCTCGTGCTCCTC-3′ and 5′-GTTGAGAAACCGAAACTACTGGG-3′. SNP genotypes were identified by inspection of sequencing chromatogram files using Sequencher (Gene Codes Corp) and CLC Main Workbench 7.5 (CLC Bio). The second genotyping method was used for samples from the PICFlu cohort has been previously published[13]. Genotype concordance was measured at 100% between the two genotyping methods used in this study.

Genotype data QC and association analysis

Genotype data for rs34481144 was obtained using Sanger sequencing. Association analysis for the FLU09 cohort was performed using linear mixed models as implemented in the software GEMMA[42]. The relatedness matrix required by the method was estimated using genotype data for the samples from Illumina HumanOmni25M array which was only used for sample QC purposes. The SNP rs34481144 was not present on the array. Estimated genomic inflation factor (λ_{gc}) was 1.019. Haplotype analysis was conducted using PLINK to determine linkage disequilibrium between the two genotyped SNPs rs12252 and rs34481144. For the Genentech Challenge Study, only subjects with self-reported European ancestry (N=42) were selected for association analysis. Association analysis was performed

using PLINK –logistic command using rapid viral replication as a binary phenotype, which was regressed on additively coded genotypes at rs34481144. For the PICFlu cohort, PLINK was used to analyze the association of rs34481144 with the dichotomous phenotype of mortality (yes/no). Meta-analysis was performed using METAL[43]. A fixed effects model based on inverse variance weighting was used to combine summary statistics across the studies and the test of heterogeneity of effects was performed (P=0.48).

eQTL analysis

Expression levels of *IFITM3* from RNAseq data was compared based on rs34481144 genotype. RNAseq data was available for 45 influenza infected individuals with genotype data. For RNAseq data, total RNA was extracted from peripheral blood mononuclear cells (PBMCs) collected on the day of enrollment using the Qiagen RNeasy kit (Qiagen, Valencia, CA). Total RNA was extracted from PBMCs from influenza-infected patients that had mild to severe clinical outcomes. Quality and quantity of RNA samples was determined prior to their processing by RNA-seq. The concentration of RNA samples was determined using NanoDrop 8000 (Thermo Scientific) and the integrity of RNA was determined by Fragment Analyzer (Advanced Analytical Technologies). 1 ug of total RNA was used as an input material for library preparation using TruSeq RNA Sample Preparation Kit v2 (Illumina). Size of the libraries was confirmed using Fragment Analyzer (Advanced Analytical Technologies) and their concentration was determined by qPCR based method using Library quantification kit (KAPA). The libraries were multiplexed and then sequenced on Illumina HiSeq2500 (Illumina) to generate 30M of single end 50 base pair reads. For each RNA seq sample, fastq files were mapped to the hg19 genome using both BWA[44] and STAR[45]. Reads were counted with HTSEQ[46] and log2 FPKM values stabilized by addition of 2 prior to transformation were generated using STATA/MP 12. These log2 values were imported into Partek Genomics Suite 6.6 (St. Louis, MO) associated with clinical metadata and analyzed via PCA. The log2 FPKM values for *IFITM3* were compared between the three genotype groups. Only 1 individual from the RNAseq dataset had the A/A genotype; therefore, we were unable to run ANOVA analysis to compare the three genotypes. A Welch Two Sample T-test was performed to compare *IFITM3* expression levels between individuals with the A/G and G/G genotype of rs34481144 using R in RStudio. 95% confidence interval and 99% confidence interval for the mean of *IFITM3* expression levels in individuals with the G/G genotype was calculated using GraphPad Prism to determine where the expression level for the A/A individual was in comparison to the range for individuals with the A/G and G/G genotypes. Data from the Phenogenetic Project at the Broad was used to analyze IFITM3 expression based on genotype of rs34481144 and rs12252 in monocyte derived dendritic cells as described previously[18]. Briefly, the patient monocyte derived dendritic cells were unstimulated, LPS-stimulated, IFNβ-stimulated, and influenza infected, and then transcriptome analysis was conducted using Nanostring technologies. Linear regression of *IFITM3* expression level was performed on additively coded (0, 1, and 2) genotype and significance was assessed using Wald test for genotype effect size.

Microarray data from PICFLU at Boston Children’s Hospital was used to analyze an association of IFITM3 expression with rs34481144 genotype. Blood was collected in PAXgene Tubes (PreAnalytiX, Qiagen/BD, Switzerland), and managed according to

manufacturer's specifications. RNA was extracted from whole blood with the PAXgene Blood RNA System kits (PreAnalytiX, Qiagen) using manufacturer's manual protocol or automated QIAcube protocol. RNA quality and concentration was determined using NanoDrop ND1000 (Thermo Scientific). RNA integrity was assessed by Agilent 2100 Bioanalyzer using the RNA 6000 Nano Assay kit according to manufacturer's instructions (Agilent Technologies, Santa Clara, CA). 63 patient RNA samples were assayed on the HuGene-1_0-st-v1 Affymetrix microarrays. CEL file data were quality controlled and RMA summarized[47] using Partek Genomics Suite 6.6. Statistical analysis was conducted using one-way ANOVA in GraphPad Prism to determine if expression was associated with genotype. Variance between genotype groups was found to be similar by Bartlett's test.

All eQTL boxplots were generated in R with the boxplot function using default plotting parameters. The top and bottom of the box represents the 75th and 25th percentiles (the upper and lower quartiles), and the center line represents the median. The whiskers extend to the data point which is no more than 1.5 times the interquartile range from the box.

Luciferase assay

Differences in promoter activity due to genotype of rs34481144 were measured using a luciferase assay. Luciferase assay of the promoter activity of *IFITM3* has been described previously[20] and modified for our SNP of interest rs34481144. Briefly, genomic DNA was extracted from HEK293T and Jurkat cells lines using Qiagen's DNA mini kit. IFITM3's core promoter (chr11:320813-321159) was amplified using (F) 5′-TATACTGCAGCTAGCGAGCCCTGAACCGGGACAGTG-3′; (R) 5′-TATACTGCACTCGAGTGGTGTCCAGCGAAGACCAGC3′. PCR product was validated using the primer sequences (F) 5′-GAGCCCTGAACCGGGACAGTG-3′ and (R) 5′-TGGTGTCCAGCGAAGACCAGC-3′. Site-directed mutagenesis was used to mutate the genotype of rs34481144 in the PCR product from the A allele to G allele before cloning into the plasmid with primers 5′-ACCATCCCAGTAACCCGACCGCCGCTGGTCTTCGCTGGACA-3′ and 5′-TGTCCAGCGAAGACCAGCGGCGGTCGGGTTACTGGGATGGT-3′ using QuickChange (Agilent Technologies) protocol according to manufacturer's instructions. Two plasmids were constructed by independently inserting both promoters (A and G alleles) between the NheI and XhoI sites of pGL4 basic vector (Promega) with the coding region for firefly luciferase. Colonies were selected for using LB with ampicillin, and Sanger sequencing was used to validate. HEK293T and Jurkat cell lines were transfected with each reporter plasmid DNA using TransIT Transfection Reagent (Mirius Bio LLC, Madison, WI). Briefly, cells were seeded into 96-well flat-bottom plates at 50,000 cells/well and 0.15μg pGL4 (IFITM3prom-A or IFITM3prom-G) plasmid DNA and 0.15μg pRL plasmid DNA were added for 24 hours. All transfections were carried out in triplicate. After transfection, cells were washed with PBS, and then treated. HEK293Ts were treated with INFα (1000U/mL) and influenza (1 MOIA/Brisbane/59/07) for 4 and 24 hours, and then Jurkats were stimulated with ConA (2.5μg/mL). Uninfected controls were included as well. Dual luciferase activity was measured using the Dual Luciferase Assay (Promega) on the GloMax platform. To analyze if genotype has an effect on luciferase activity across treatments, statistical analysis was performed using Two-Way ANOVA in Prism software (GraphPad).

For the analysis of unstimulated Jurkats only, a two-tailed, unpaired T-test was performed to compare the difference between genotypes (P=0.02), and equal variance was found between groups. Two-way ANOVA was used to look at differences based on genotype across unstimulated and ConA-stimulated cells (Extended Data Figure 3A; P<0.0001).

Electrophoretic mobility shift assay

Oligonucleotides were synthesized for the region surrounding the SNP rs34481144 in the promoter region of *IFITM3* (IDT) to create 40bp probes. The sequences of the probe with the A genotype are as follows: (F) 5′-AGTAACCCGACCACCGCTGGTCTTCGCTGGACACCATGAA-3′ and (R) 5′-TTCATGGTGTCCAGCGAAGACCAGCGGTGGTCGGGTTACT -3′. The sequences of the probe with the G genotype are as follows: (F) 5′-AGTAACCCGACCGCCGCTGGTCTTCGCTGGACACCATGAA-3′ and (R) 5′-TTCATGGTGTCCAGCGAAGACCAGCGGCGGTCGGGTTACT-3′. Oligonucleotides were biotinylated at the 5′-end (Integrated DNA Technologies). Unlabeled probes were also ordered using the same sequences for specific competition of labeled probes. The complementary oligonucleotides were annealed to produce double-stranded oligonucleotide probes. HEK293T and A549 cell lines were infected with 5 MOI A/Brisbane/59/07 for various time points from 30 minutes to 24 hours for this experiment. Jurkats were stimulated with PMA (50ng/mL) for 2–48 hours. Nuclear extracts were prepared using the NE-PER Nuclear and Cytoplasmic Extraction Reagents kit (Thermo Fisher Scientific). Binding reactions were performed with LightShift Chemiluminescent Kit according to manufacturer's instructions (Thermo Fisher Scientific). Briefly, the binding reaction was prepared with 1X Binding buffer (100mM Tris, 500nM KCl, 10mM DTT; pH7.5), 4% Glycerol, 1mM MgCl2, 5mM EDTA, and 1μg Poly (dI•dC), and then 3uL nuclear extract was added before incubation with the 20fmol of each of the test probes (A and G). For specific competitor binding reactions, 300 fold molar excess of unlabeled probe was added to the binding reaction mixture before the test probe was added. The reactions were loaded on Novex® 6% DNA Retardation Gels TBE ready gel (Life Technologies) and run at 250V for 20 minutes on ice, then transferred on ice to nylon membrane (380mA for 45 minutes). The DNA was then crosslinked to the membrane at 120mJ/cm^2 using Stratalinker® UV Crosslinker (Stratagene) before continuing with the LightShift Chemiluminescent kit to detect the biotinylated probes.

For the Western-Blotted EMSA (WEMSA)[48], probes were incubated with nuclear extract, CTCF purified full-length recombinant protein (100ng; Abnova), or IRF3 purified full length recombinant protein (60–100ng; Abnova) along with the same binding buffer as described for EMSAs. The samples were incubated at 4°C for 20 minutes before they were run on DNA retardation gels at 250V on ice for 20 minutes. The samples were then transferred on ice to a nitrocellulose membrane for Western blotting (30V for 1 hour). Following transfer, the membrane was blocked with 5% milk in TBST for 1 hour rocking at room temperature before primary antibody was applied (1:1000 dilution in 5% milk) overnight at 4°C, rocking. Antibodies used for WEMSAs include CTCF (Cell Signaling 2899; http://www.antibodypedia.com/gene/15816/CTCF/antibody/106154/2899), IRF3 (Cell Signaling 11904; https://www.citeab.com/antibodies/701329-11904-irf-3-d6i4c-xp-rabbit-mab), and

STAT1 (EMD Millipore; http://1degreebio.org/reagents/product/848076/?qid=1294237). Membrane was washed with TBST before secondary was added for 1 hour at room temperature and was washed before substrate was added for imaging.

For WEMSAs using methylated probes, methods are the same as described above. Probes were methylated using standard NEB protocol for CpG Methyltransferase (M.SssI) enzyme (NEB). Briefly, nuclease-free water, 10X NEBuffer 2, 1600μM S-adenosylmethionine (SAM), 1μg double-stranded probe, and 1μL 4U/μL SssI methylase were combined and mixed, then incubated for one hour at 37°C with the reaction stopped by heating at 65°C for 20 minutes. Probes were cleaned and concentrated using DNA Clean & Concentrator™ (Zymo). Probes were then diluted to 10fmol/uL for use in the WEMSA.

EMSA and WEMSA membranes were all imaged with the Amersham Imager 600 (GE Healthcare). WEMSA bands were quantified using ImageJ software[49] to compare the difference between binding to the two genotypes on each membrane. Each experiment was then normalized to the lower intensity and plotted according to protein to display the difference in binding based on genotype.

Cell sorting from PBMCs

Classical monocytes (CD3-CD14+CD16-) and CD8+ T-cell subsets (Naïve-CD3+CD8+CCR7+CD45RO-and TEM- CD3+CD8+CCR7-CD45RO+) were sorted from thawed PBMCs. For sorting experiments, cells were initially stained with Live/Dead Aqua (Life Technologies) per manufacturer's instructions. Briefly, cells were centrifuged, washed, and resuspended in 1 mL of PBS to which 1 μL of reconstituted dye was added. Cells were then incubated for 30 minutes at room temperature, protected from light. Following incubation, the cells were washed with 1 mL of staining buffer (2% FBS, 1mM EDTA, 1X Penicillin Streptomycin (Pen-Strep)). Cells were then stained for surface markers, as described above, in sorting buffer for 20 minutes at room temperature using the following human mAbs: APC/Cy7- CD3 (BioLegend 344818, Clone SK7; https://www.biolegend.com/en-us/products/apc-cy7-anti-human-cd3-antibody-6940), PB-CD8a (Biolegend 301033, Clone RPA-T8; https://www.biolegend.com/en-us/products/pacific-blue-anti-human-cd8a-antibody-2851), FITC-CCR7 (BioLegend 353216, Clone G043H7; https://www.citeab.com/antibodies/524602-353216-fitc-anti-human-cd197-ccr7-antibody-g043h7), APC-CD45RO (BioLegend 304210, Clone UCHL1; https://www.biolegend.com/en-us/products/apc-anti-human-cd45ro-antibody-856), BV785-CD14 (BioLegend 301840, Clone M5E2; http://www.biolegend.com/brilliant-violet-785-anti-human-cd14-antibody-7965.html), and PE/Cy7-CD16 (BioLegend 302016, Clone 3G8; https://www.biolegend.com/en-us/products/pe-cy7-anti-human-cd16-antibody-1905). Following surface staining, cells were washed, diluted to 10×10^6 cells per mL and filtered through sterile 50μm filters (Sysmex) to ensure that cells were in single cell suspensions. After staining, cell subsets were sorted into tubes using a Sony sorter (Model sy3200, Sony Biotech Synergy sorter, Sony Biotech, San Jose, CA) equipped with a 70 μm nozzle. Representative sort plot with gating schema shown in Supplemental Fig. 5b.

Methylation

DNA was extracted from monocytes and CD8+ T-cell subsets sorted as described above from six individuals from the FLU09 cohort, three with the A/G genotype and 3 with the G/G genotype, as well as five healthy donors for CD8+ T-cell analysis only using Qiagen's All Prep DNA/RNA Mini or Micro kit depending on the number of cells sorted. Individuals with the A/A genotype would not have the CpG site available for methylation and so were excluded from this analysis.

We also extracted DNA from undifferentiated and differentiated normal human bronchial epithelial (NHBE) cells from two donors with A/G and G/G genotypes. Expanded vials of NHBE donors 17623 (Donor 1) and 78338 (Donor 2) were plated in transwell plates and grown in growth media until 100% confluent. Growth media made up from Bronchial Epithelial Basal Medium (BEBM) media (Lonza), BEGM BulletKit (Lonza), 5×10^{-8}M Retinoic acid (Sigma), and 25ng/mL Epidermal growth factor (Millipore) with the RA and EGF replacing the aliquots in the BulletKit. When the cells were 100% confluent, an air-liquid interface was established by removing media on cells and adding ALI media to the basolateral layer of the transwell plate. ALI media was made with similarly but with ALI media uses the BulletKit EGF aliquot. Cells differentiated into cells representative of the human airway. Differentiated cells were washed with warm saline to remove any mucus produced at this state. One EpiAirway culture of differentiated NHBEs (MatTek) was also used for this experiment. Differentiated cells were received and equilibrated overnight in MatTek medium. Cells were washed with warm saline to remove mucus produced. For the differentiated NHBE cells, we also infected the cells with 3MOI H1N1 A/Brisbane/59/07 for 24 hours before extracting DNA with Qiagen's All Prep DNA/RNA Mini kit. Genotyping of NHBE and EpiAirway samples used same methods as described above.

Genomic DNA (100ng) was subjected to bisulfite (BS) treatment using the EZ DNA methylation Gold kit (Zymo Research). Following BS treatment, the DNA was used to set up loci specific PCR using forward primer 5′-GTAATTTGATTTTTAGGAATTTG -3′ and reverse primer 5′-CTAAACATTCCCTAAAACC -3′. The PCR product was gel extracted using the Zymoclean Gel DNA Recovery Kit (Zymo Research) and ligated overnight at 4°C to the pGEM vector using T4 DNA ligase following the manufacturers protocol (pGEM T vector system-I, Promega). Highly competent *E. coli* (Agilent Technologies) were transformed with the ligated DNA and spread on LB agar plates containing Ampicillin, X-Gal and IPTG and incubated for 16hrs at 37°C. White colonies were picked and inoculated into 1ml of 2x LB broth and incubated overnight at 37°C in a shaking incubator at 220 rpm. Plasmid DNA was isolated from the cultured bacteria using the Directpret 96 Miniprep Kit (Qiagen). DNA was quantified using a Nanodrop and sent for sequencing using M13 forward primer 5′-TCACACAGGAAACAGCTATGAC -3′. Sequencing results were analyzed using QUMA online server (http://quma.cdb.riken.jp/) to assess the extent of methylation at the selected loci. To determine if there was a difference in methylation in T_N and T_{EM} cells in the G/G individuals, a two-tailed unpaired T-test was used to analyze the difference in GraphPad Prism. Variances between groups were not statistically different.

Western blotting

CD8+ T-cells were enriched from thawed PBMCs of healthy donors by negative selection using the EasySep™ Human CD8+ T Cell Isolation Kit (Stem Cell) per manufacturer's instructions. An aliquot of the enriched CD8+ T-cells was evaluated by flow cytometry and the purity was determined to be on average 90.4% for all healthy donors. Representative flow cytometry gating schema to determine purity of CD8+ T-cells shown in Supplemental Fig. 6. The remaining cells were resuspended in RIPA buffer (Cell Signaling) with cOmplete protease inhibitors (Roche) and stored at −80°C.

Cell lines were infected with 3 MOI A/Brisbane/59/07 for 24 hours to compare IFITM3 levels between resting state and infection. Cells were washed with sterile DPBS twice before lysing with RIPA buffer (Cell Signaling) with cOmplete protease inhibitors (PI; Roche) then stored at −80°C. PBMCs from patients with different rs34481144 genotypes were also lysed using RIPA buffer with PI and stored at −80°C to determine protein expression levels of IFITM1 and IFITM3.

Protein concentration was measured in triplicate for each sample. Samples were loaded onto NuPAGE™ 4–12% Bis-tris protein gel (Thermo) in groups by genotype or infection status under reducing conditions, separated by SDS-PAGE, and immunoblotted using antibodies to IFITM3 (59212S, Cell Signaling, 1:500), IFITM1 (13126, Cell Signaling, 1:500), and β-Actin (4970S, Cell Signaling, 1:6000). Imaged on ChemiDoc Touch (BioRad), and band intensities were quantified with the Image Lab 5.2.1 software (BioRad). Data are represented as mean ± SEM for each genotype group.

CD8+ and CD4+ T-cell phenotyping from human nasal wash samples

Nasal lavage cells from FLU09 patients on day 0 (the day patients presented for medical attention) were thawed and resuspended in RPMI medium containing 10% fetal bovine serum (FBS), 1% Pen-Strep, 1% sodium pyruvate, and benzonase endonuclease (EMD Millipore, 1:10000). Approximately 1×10^5 cells were plated and blocked with human FcR blocking reagent (Mitenyi Biotec, 1:100) in SORT buffer (PBS, 1% bovine serum albumin) for 15 minutes at 4°C. Cells were stained with a panel of surface human mAbs in SORT buffer for 30 minutes on ice, including PerCP-Cy5.5 - CD3 (Biolegend 344808, 1:100, clone SK7; https://www.biolegend.com/en-us/products/percp-cy5-5-anti-human-cd3-antibody-6932), Pacific Blue - CD4 (Biolegend 344620, 1:200, clone SK3; http://www.biocompare.com/9776-Antibodies/1077251-Pacific-Bluetrade-antihuman-CD4/), and APC-eFlour780 - CD8 (eBioscience 47-0087-42, 1:200, clone SK1; http://www.ebioscience.com/human-cd8a-antibody-apc-efluor-780-sk1.htm). Subsequently, cells were fixed with a 1% formaldehyde-PBS solution on ice for 10 min. After two washes, cells were resuspended in 50μl of SORT buffer for flow cytometric analysis (BD LSR2). Representative flow cytometry gating schema for NWC samples shown in Supplemental Fig. 7a. Additional analysis was performed using FlowJo (TreeStar, Ashland, OR). Single stain and fluorescence minus-one (FMO) controls were performed. Total cell number and frequency were then calculated using the cell count performed directly after nasal lavage was sampled from the participant and before freezing samples down. Data was analyzed and plotted using R in RStudio. The boxplot was generated using default plotting parameters.

The top and bottom of the box represents the 75th and 25th percentiles (the upper and lower quartiles), and the center line represents the median. The whiskers extend to the data point which is no more than 1.5 times the interquartile range from the box. Welch's two-sample T-test was used to analyze differences in number and frequency for each subset based on genotype. For statistical testing, CD8+ and CD4+ frequencies were log-transformed to ensure equal variances between groups. Linear regression modeling with correction for age at enrollment and viral titer at day 0 was also used for this analysis in R.

Neighboring gene analysis using RNAseq data

To assess whether CTCF binding at our SNP changed domain structure, we examined gene expression correlations between individuals with A/G genotypes (N=11) and G/G genotypes (N=33). It has been reported that the boundaries or topological domains caused by looping are on average about 180kb[30,33]. For our analysis, we included all genes which passed RNAseq QC whose gene bodies fall within 260kb of our SNP (chr11:193,000-612,222) were assessed for correlation of gene expression. Corrplot in R was used to generate plots using Spearman correlation analysis. All results shown in the plots are significant after FDR adjustment with insignificant results (P>0.05) left blank. Results are shown based on location of each gene on chromosome 11.

Genes in this region were further analyzed individually for correlation with *IFITM3* using GraphPad Prism to confirm correlation results. *IFITM3* and each gene were subjected to correlation test (Spearman) in patients with A/G genotype and then G/G genotype.

Statistical analysis

Statistical methods are found in the Figure legends of each section and in the relevant methods descriptions. In all cases, we considered group numbers, sample size, variance, and number of comparisons in selecting an appropriate test. For the human genetic association analyses, statistical methods were not used to determine sample sizes for cohorts used. No samples were excluded from analysis. Sample sizes were adequate for the association analysis of one SNP. Experiments were not performed in a blinded fashion. Variance between groups was determined for each statistical test (by F test for t-tests and by Bartlett's test for one-way ANOVAs) and was corrected for in cases on unequal variance by log-transforming the data.

Code availability

R code used for the eQTL plotting and analysis, cellular analysis and plotting, and correlation plot is available by request to emma.allen@stjude.org. R code was conducted using the Very Secure Dishes version of R in RStudio.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the participants who contributed in each of the studies described in this manuscript. For FLU09, they acknowledge the clinical work of Lisa Harrison and the clinical team at the UTHSC, Le Bonheur Children's Hospital. They thank Dr. John J. Treanor (University of Rochester Medical Center, Rochester, NY) for providing counsel regarding study design; Resha Bajracharya, Thomas H. Oguin III, and Lana McClaren for technical assistance; Ashley Webb, Jeri Carol Crumpton, Kimberly Friedman, Trushar Jeevan, and Sook-San Wong for their help with virologic assays; and Dr. Pamela McKenzie and Dr. Kyle Johnson for assisting in the study management. They thank Mark Weilnau, Jerry Parker, Richard Eila, and Ramin Tabrizi for data management. We thank the Genentech FluA team for access to clinical samples. For PICFlu, they acknowledge the hard work and collaboration of the PALISI PICFlu Study Site Investigators who critically reviewed the initial study proposal and all modifications, and enrolled and collected data on the patients in this study:

Arkansas Children's Hospital, Little Rock, AR (Ronald C. Sanders, MD, Glenda Hefley, RN, MNsc); Phoenix Children's Hospital, Phoenix, AZ (David Tellez, MD, Courtney Bliss, MS, Aimee Labell, MS, RN, Danielle Liss, BA, Ashely L. Ortiz, BA); Children's Hospital of Los Angeles, Los Angeles, CA (Barry Markovitz, MD, Jeff Terry, MBA, Rica Sharon P. Morzov, RN, BSN, CPN); Children's Hospital Central California, Madera, CA (Ana Lia Graciano, MD, Melita Baldwin, BS); Children's Hospital and Research Center, Oakland, CA (Heidi Flori, MD, Becky Brumfield, RCP, Julie Simon, RN); Children's Hospital of Orange County, Orange, CA (Nick Anas, MD, Adam Schwarz, MD, Chisom Onwunyi, RN, BSN, MPH, CCRP, Stephanie Osborne, RN, CCRC, Tiffany Patterson, BS, CRC, Ofelia Vargas-Shiraishi, BS, CCRC); UCSF Benioff Children's Hospital, University of California San Francisco, San Francisco, CA (Anil Sapru, MD, Maureen Convery, BS, Victoria Lo, BA); Children's Hospital Colorado, Aurora, CO (Angela Czaja, MD, Peter Mourani, MD, Valeri Batara Aymami, RN, MSN-CNS, Susanna Burr, CRC, Megan Brocato, CCRC, Stephanie Huston, BS, RPSGT, CRC, Emily Jewett, Sandra B. Lindahl, RN, Danielle Loyola, RN, BSN, MBA, Yamila Sierra, MPH); Connecticut Children's Medical Center, Hartford, CT (Christopher Carroll, MD, MS, Kathleen A. Sala, MPH, Sherell Thornton-Thompson, BA); Yale-New Haven Children's Hospital, New Haven, CT (John S. Giuliano Jr., MD); Holtz Children's Hospital, Miami, FL (Gwenn McLaughlin, MD); Children's Healthcare of Atlanta at Egleston, Atlanta, GA (Matthew Paden, MD, Chee-Chee Manghram, Stephanie Meisner, RN, BSN, CCRP, Cheryl L. Stone, RN, Rich Toney, RN); The University of Chicago Medicine Comer Children's Hospital, Chicago, IL (Juliane Bubeck Wardenburg, MD, PhD, Andrea DeDent, PhD); Kosair Children's Hospital, Louisville, KY (Vicki Montgomery, MD, FCCM, Tracy Evans, RN, Kara Richardson, RN); Boston Children's Hospital, Boston, MA (Adrienne Randolph, MD, MSc, Anna A. Agan, BA, Ellen M. Smith, BS, Ryan M. Sullivan, RN, BSN, CCRN, Grace Yoon, RN, NNP, Michael Kiers, RN, Shannon M. Keisling, BA); Johns Hopkins Children's Center, Baltimore, MD (Melania Bembea, MD, MPH, Elizabeth D. White, RN, CCRP); Children's Hospital and Clinics of Minnesota, Minneapolis, MN (Stephen C. Kurachek, MD, Angela A. Doucette, CCRP, Erin Zielinski, MS); St. Louis Children's Hospital, St. Louis, MO (Allan Doctor, MD, Mary Hartman, MD, Rachel Jacobs, BA); Children's Hospital of Nebraska, Omaha, NE (Edward Truemper, MD, Machelle Dawson, RN, BSN, MEd, CCRC); Children's Hospital at Dartmouth-Hitchcock, Lebanon, NH (Daniel L. Levin, MD, J. Dean Jarvis, MBA, BSN); The Children's Hospital at Montefiore, Bronx, NY (Chhavi Katyal, MD); Golisano Children's Hospital, Rochester, NY (Kate Ackerman, MD, L. Eugene Daugherty, MD, Laurel Baglia, PhD); Nationwide Children's Hospital, Columbus, OH (Mark W. Hall, MD, Kristin Greathouse, BSN, MS, Lisa Steele, RN, BSN, CCRN); Penn State Children's Hospital, Hershey, PA (Neal Thomas, MD, Jill Raymond, RN, MSN, Debra Spear, RN); Children's Hospital of Philadelphia, Philadelphia, PA (Julie Fitzgerald, MD, Mark Helfaer, MD, Scott Weiss, MD, Jenny L. Bush, RNC, BSN, Mary Ann Diliberto, RN, Jillian Egan, RN, Brooke B. Park, RN, BSN, Martha Sisko, RN, BSN, CCRC); Monroe Carell Jr. Children's Hospital at Vanderbilt, Nashville, TN (Frederick E. Barr, MD, Judi Arnold, RN); Dell Children's Medical Center of Central Texas, Austin, TX (Renee Higgerson, MD, LeeAnn Christie, RN); Children's Medical Center, Dallas, TX (Marita Thompson, MD); Texas Children's Hospital, Houston, TX (Laura L. Loftis, MD, Nancy Jaimon, RN, MSN-Ed, Ursula Kyle, MS); University of Virginia Children's Hospital (Douglas F. Willson, MD, Christine Traul, MD, Robin L. Kelly, RN); Children's Hospital of Wisconsin, Milwaukee, WI (Rainer Gedeit, MD, Briana E. Horn, Kate Luther, MPH, Kathy Murkowski, RRT, CCRC); Centre Hospitalier Universitaire Sainte-Justine, Montreal, Quebec, Canada (Philippe A. Jouvet, MD, Anne-Marie Fontaine, BSc); Centre Hospitalier de l'Université Laval, Quebec, Quebec, Canada (Marc-André Dugas, MD). This work was funded by the National Institute of Allergy, and Infectious Diseases, National Institutes of Health under HHS Contracts HHSN266200700005C and HHSN272201400006C for the St. Jude Center of Excellence for Influenza Research and Surveillance (PGT), Genentech, Inc. (PGT, CMR), National Institutes of Health (NIH AI084011 AGR), the Centers for Disease Control and Prevention (CDC; AGR), and ALSAC. CMR, TB, AD, and MO are employees of Genentech, Inc.

References

1. Centers for Disease Control and Prevention (CDC). Estimates of deaths associated with seasonal influenza --- United States, 1976–2007. MMWR Morb Mortal Wkly Rep. 2010; 59:1057–1062. [PubMed: 20798667]

2. Amini-Bavil-Olyaee S, et al. The antiviral effector IFITM3 disrupts intracellular cholesterol homeostasis to block viral entry. Cell Host Microbe. 2013; 13:452–464. [PubMed: 23601107]
3. Desai TM, et al. IFITM3 Restricts Influenza A Virus Entry by Blocking the Formation of Fusion Pores following Virus-Endosome Hemifusion. PLoS Pathog. 2014; 10
4. Wakim LM, Gupta N, Mintern JD, Villadangos JA. Enhanced survival of lung tissue-resident memory $CD8^+$ T cells during infection with influenza virus due to selective expression of IFITM3. Nat Immunol. 2013; 14:238–245. [PubMed: 23354485]
5. Bailey CC, Huang IC, Kam C, Farzan M. Ifitm3 limits the severity of acute influenza in mice. PLoS Pathog. 2012; 8:e1002909. [PubMed: 22969429]
6. Feeley EM, et al. IFITM3 inhibits influenza A virus infection by preventing cytosolic entry. PLoS Pathog. 2011; 7:e1002337. [PubMed: 22046135]
7. Zhang YH, et al. Interferon-induced transmembrane protein-3 genetic variant rs12252-C is associated with severe influenza in Chinese individuals. Nat Commun. 2013; 4:1418. [PubMed: 23361009]
8. Everitt AR, et al. IFITM3 restricts the morbidity and mortality associated with influenza. Nature. 2012; 484:519–523. [PubMed: 22446628]
9. Yang X, et al. Interferon-Inducible Transmembrane Protein 3 Genetic Variant rs12252 and Influenza Susceptibility and Severity: A Meta-Analysis. PloS One. 2015; 10:e0124985. [PubMed: 25942469]
10. Wang Z, et al. Early hypercytokinemia is associated with interferon-induced transmembrane protein-3 dysfunction and predictive of fatal H7N9 infection. Proc Natl Acad Sci. 2014; 111:769–774. [PubMed: 24367104]
11. Mills TC, et al. IFITM3 and susceptibility to respiratory viral infections in the community. J Infect Dis. 2014; 209:1028–1031. [PubMed: 23997235]
12. López-Rodríguez M, et al. IFITM3 and severe influenza virus infection. No evidence of genetic association. Eur J Clin Microbiol Infect Dis. 2016; :1–7. DOI: 10.1007/s10096-016-2732-7
13. Randolph AG, et al. Evaluation of IFITM3 rs12252 Association with Severe Pediatric Influenza Infection. J Infect Dis. 2017; doi: 10.1093/infdis/jix242
14. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. Nature. 2015; 526:68–74. [PubMed: 26432245]
15. Allen EK, Thomas PG. Immunity to influenza. Preventing infection and regulating disease. Am J Respir Crit Care Med. 2015; 191:248–251. [PubMed: 25635487]
16. Hall MW, et al. Innate immune function and mortality in critically ill children with influenza: a multicenter study. Crit Care Med. 2013; 41:224–236. [PubMed: 23222256]
17. Consortium, T. Gte. et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science. 2015; 348:648–660. [PubMed: 25954001]
18. Lee MN, et al. Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells. Science. 2014; 343:1246980. [PubMed: 24604203]
19. Huang IC, et al. Distinct patterns of IFITM-mediated restriction of filoviruses, SARS coronavirus, and influenza A virus. PLoS Pathog. 2011; 7:e1001258. [PubMed: 21253575]
20. Shen C, et al. A Functional Promoter Polymorphism of IFITM3 Is Associated with Susceptibility to Pediatric Tuberculosis in Han Chinese Population. PLoS ONE. 2013; 8:e67816. [PubMed: 23874452]
21. Hiscott J, et al. Triggering the Interferon Response: The Role of IRF-3 Transcription Factor. J Interferon Cytokine Res. 1999; 19:1–13. [PubMed: 10048763]
22. Barnes B, Lubyova B, Pitha PM. Review: On the Role of IRF in Host Defense. J Interferon Cytokine Res. 2002; 22:59–71. [PubMed: 11846976]
23. Ciancanelli MJ, et al. Life-threatening influenza and impaired interferon amplification in human IRF7 deficiency. Science. 2015; 348:448–453. [PubMed: 25814066]
24. Kim S, Yu NK, Kaang BK. CTCF as a multifunctional protein in genome regulation and gene expression. Exp Mol Med. 2015; 47:e166. [PubMed: 26045254]
25. Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet. 2014; 15:234–246. [PubMed: 24614316]

26. Ea V, Baudement MO, Lesne A, Forné T. Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization. Genes. 2015; 6:734–750. [PubMed: 26226004]
27. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science. 2004; 306:636–640. [PubMed: 15499007]
28. Wang H, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res. 2012; 22:1680–1688. [PubMed: 22955980]
29. Ziebarth JD, Bhattacharya A, Cui Y. CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. Nucleic Acids Res. 2013; 41:D188–D194. [PubMed: 23193294]
30. Flavahan WA, et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature. 2016; 529:110–114. [PubMed: 26700815]
31. Lupiáñez DG, et al. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. Cell. 2015; 161:1012–1025. [PubMed: 25959774]
32. Ji X, et al. 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell Stem Cell. 2016; 18:262–275. [PubMed: 26686465]
33. Rao SSP, et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell. 2014; 159:1665–1680. [PubMed: 25497547]
34. Tang Z, et al. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell. 2015; 163:1611–1627. [PubMed: 26686651]
35. Zhang Y, et al. Interferon-induced transmembrane protein-3 rs12252-C is associated with rapid progression of acute HIV-1 infection in Chinese MSM cohort. AIDS Lond Engl. 2015; 29:889–894.
36. Compton AA, et al. Natural mutations in IFITM3 modulate post-translational regulation and toggle antiviral specificity. EMBO Rep. 2016; 17:1657–1671. [PubMed: 27601221]
37. Tsuboi S, Fukuda M. Roles of O-linked oligosaccharides in immune responses. BioEssays. 2001; 23:46–53. [PubMed: 11135308]
38. Nakamura S, et al. Influenza A Virus-Induced Expression of a GalNAc Transferase, GALNT3, via MicroRNAs Is Required for Enhanced Viral Replication. J Virol. 2016; 90:1788–1801.
39. Sun X, et al. Constitutively expressed IFITM3 protein in human pulmonary endothelial cells poses an early infection block to human influenza viruses. J Virol. 2016; JVI.01254-16. doi: 10.1128/JVI.01254-16
40. Infusini G, et al. Respiratory DC Use IFITM3 to Avoid Direct Viral Infection and Safeguard Virus-Specific CD8+ T Cell Priming. PloS One. 2015; 10:e0143539. [PubMed: 26600246]
41. Oshansky CM, et al. Mucosal immune responses predict clinical outcomes during influenza infection independently of age and viral load. Am J Respir Crit Care Med. 2014; 189:449–462. [PubMed: 24308446]
42. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012; 44:821–824. [PubMed: 22706312]
43. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–2191. [PubMed: 20616382]
44. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl. 2009; 25:1754–1760.
45. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. Bioinforma Oxf Engl. 2013; 29:15–21.
46. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinforma Oxf Engl. 2015; 31:166–169.
47. Irizarry RA, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003; 4:249–264. [PubMed: 12925520]
48. Deckmann K, Rörsch F, Geisslinger G, Grösch S. Identification of DNA–protein complexes using an improved, combined western blotting-electrophoretic mobility shift assay (WEMSA) with a fluorescence imaging system. Mol Biosyst. 2012; 8:1389–1395. [PubMed: 22358373]
49. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012; 9:671–675. [PubMed: 22930834]

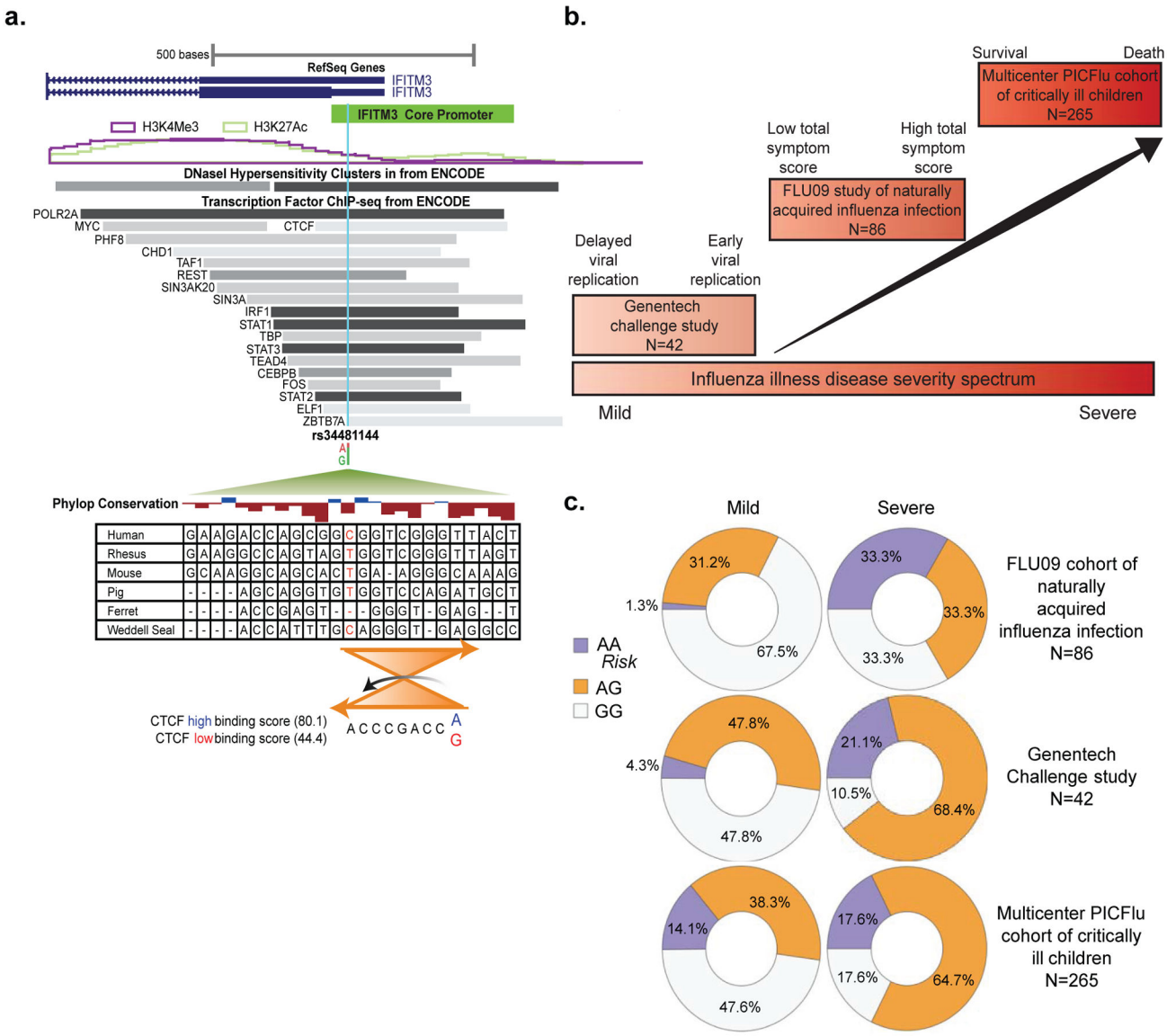

Fig. 1. IFITM3 regulatory SNP rs34481144 lies within the promoter region and is associated with clinical illness in three human cohorts

a. *IFITM3* graphic of UCSC Genome browser. Custom track "rs34481144" was added to indicate our associated SNP (genotype from reverse strand) with the allele frequency from our FLU09 cohort (A in red and G in green). The light blue vertical line across transcription factor binding sites (from the 161 proteins tested and available through ENCODE data; GRCH37/hg19) indicates the location of the SNP. ENCODE H3K27Ac and H3K4Me3 tracks (from 7 cell lines) indicate known promoter and enhancer regions. Green triangle indicates sequence region directly surrounding rs34481144, and PhyloP and sequence data show basewise conservation across the region on the forward strand orientation. CTCF binding sequence (CTCFBSDB 2.0) at the bottom with genotype (reverse strand orientation with the *IFITM3* promoter) changed in the sequence tested for strength of putative binding. Red allele indicating lower predicted binding affinity score. b. The three cohorts used for association analysis positioned along the influenza illness severity spectrum with severity

designation and cohort size described. c. Genotype frequency distribution of rs34481144 in three independent influenza cohorts, FLU09 cohort of naturally acquired influenza infection (N=86), the Genentech challenge study (N=42), and the multicenter PICFlu cohort of critically ill children (N=265), in mild influenza illness phenotypes (left side) compared to severe influenza illness phenotypes (right side). FLU09 P=0.037; Genentech Challenge Study P=0.01; PICFlu P=0.047.

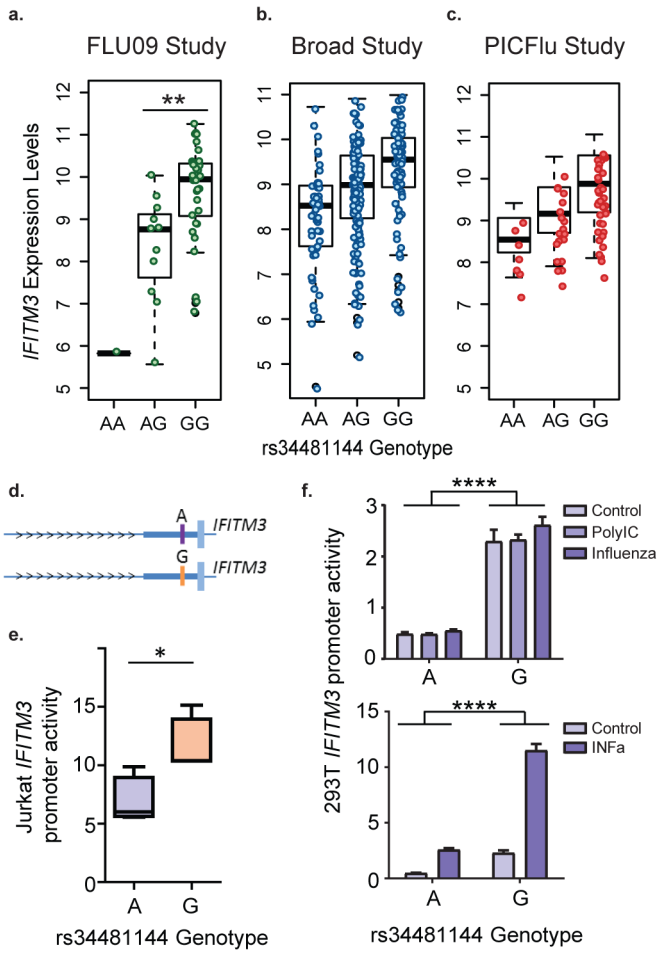

Fig. 2. Regulatory effect of rs34481144 genotype on IFITM3

a. *IFITM3* expression levels measured by RNAseq (log2(FPKM)) in PBMCs of 45 influenza-positive individuals from the FLU09 cohort plotted based on genotype of rs34481144. Asterisk indicating the two groups compared using Welch Two Sample T-test. ** *P*-value=0.008. b. *IFITM3* expression levels (log2(*IFITM3* counts) using Nanostring technology) in monocyte-derived dendritic cells from 534 healthy donors in the Phenogenetic Project and ImmVar consortium plotted by genotype of rs34481144 and analyzed using linear regression modeling ($P=4\times10^{-7}$). c. *IFITM3* expression levels (log2signal from Affymetrix microarray) in whole blood from 62 critically ill children in the PICFlu cohort plotted by rs34481144 genotype and analyzed with one-way ANOVA (*P*=0.0003). Luciferase expression from two reporter constructs with the rs34481144 alleles (d) included sequences from the core promoter of *IFITM3* (Fig. 1b) and were subcloned 5′ of a minimal promoter and the firefly luciferase gene in the pGL4 plasmid. Luciferase plasmids were transfected into (e) unstimulated Jurkat cells and (f) HEK293T cells that were left unstimulated or were stimulated with poly-IC (1000U/mL), live influenza virus (1 MOI), or IFN-α (1000 U/ml) for 24 hours. Firefly luciferase expression was normalized to Renilla luciferase expression expressed from co-transfected plasmid. All luciferase experiments included at least three technical replicates and were repeated in three independent experiments. Two-way ANOVA was used to test if genotype affected promoter activity

across all treatments. Error bars representing SD. ****P<0.0001. For the analysis of unstimulated Jurkats only, a two-tailed, unpaired T-test was performed to compare the difference between genotypes (P=0.02).

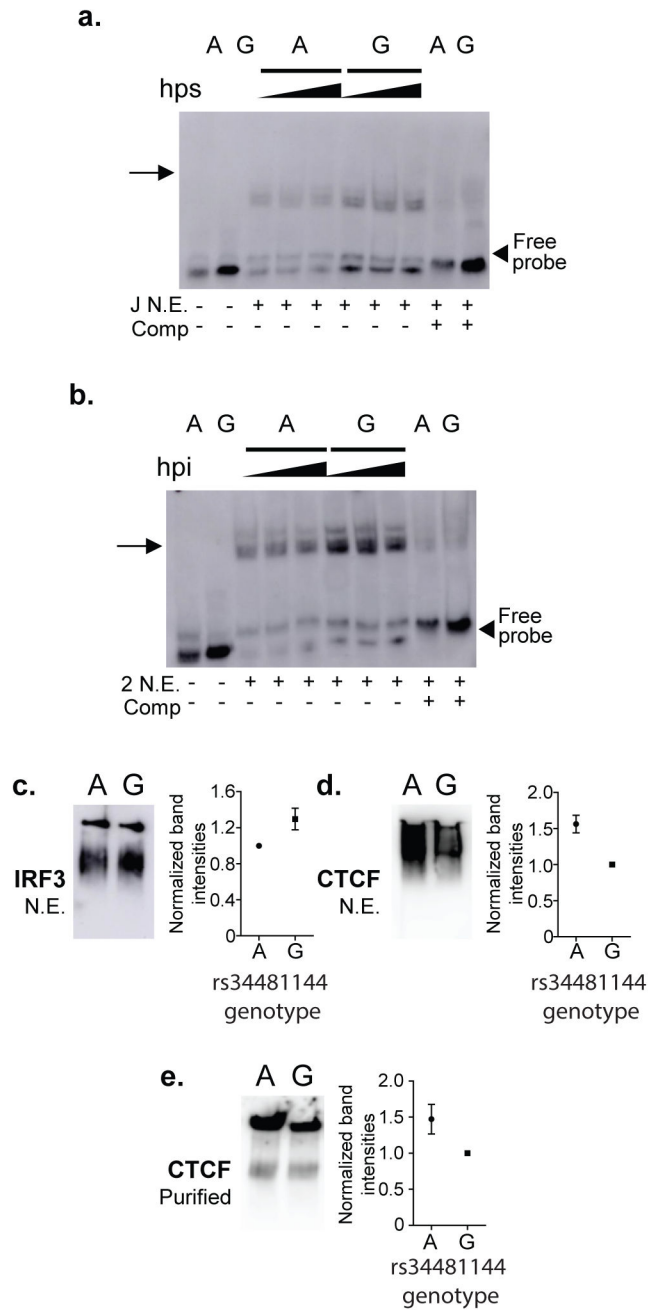

Fig. 3. Identification of changes in protein binding to the IFITM3 promoter region due to rs34481144 genotype

40-bp Biotin-labeled double-stranded DNA probes which differ only by rs34481144 genotype (A or G indicated) were used as probes in EMSA and WEMSA assays. a. EMSA using nuclear extract (N.E.) from Jurkat cells stimulated with PMA for 18, 24, and 48 hours indicated by hps (hours post-stimulation) with or without specific competitor (Comp). b. EMSA using N.E. from HEK293T cells infected with influenza for 8, 10, and 12 hours indicated by hpi (hours post-infection) with or without specific competitor. c. WEMSA using N.E. from 293T cells infected with influenza for 4 hours, and blotted with an IRF3

antibody. d. WEMSA using N.E. from HEK293T cells infected with influenza for 6 hours, and blotted with a CTCF antibody. e. WEMSA using purified recombinant CTCF and blotted for CTCF. (c–e) Data were log transformed and were then represented as mean ± SEM. Binding in WEMSAs are quantified representing at least three replicate experiments, demonstrating that they consistently show the same binding properties between rs34481144 alleles in each case.

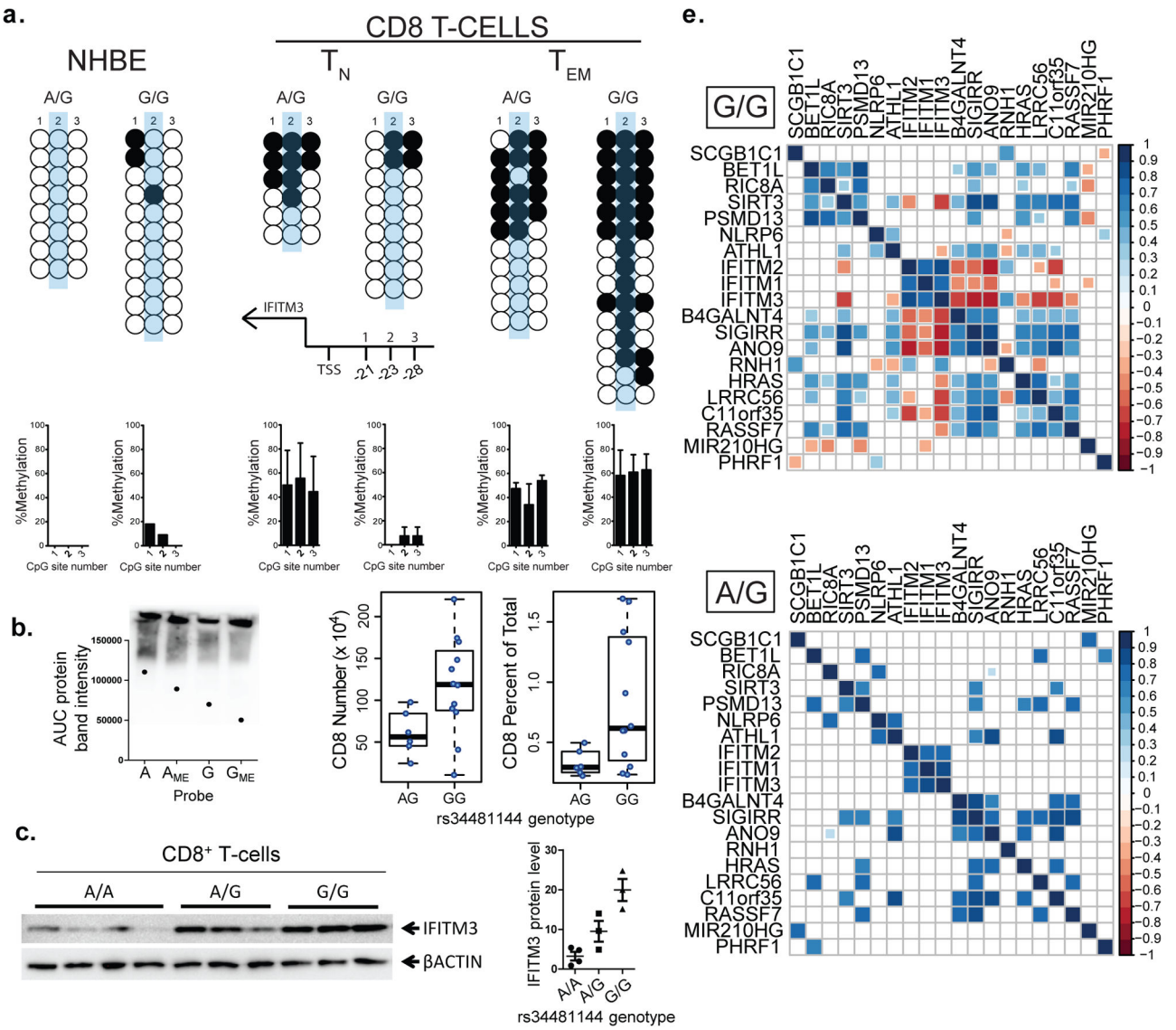

Fig. 4. Genotype at rs34481144 alters methylation of the *IFITM3* promoter, correlates with CD8+ T-cell abundance in nasal washes, and results in broader effects on the locus

a. Representative bisulfite sequencing DNA methylation analysis of CpGs upstream of IFITM3 transcriptional start site. Each row represents a clone that was picked for sequencing and each column represents CpG dinucleotide in the selected region. NHBE cells (two Lonza and one MatTek donors) and CD8+ T-cell subsets (Naïve (TN) and effector memory T-cells (TEM)) sorted from six FLU09 patients PBMC samples (three A/G and three G/G) were bisulfite converted and sequenced for DNA methylation. Top left panel illustrates rs34481144 (CpG 2) with two flanking CpG sites in relation to IFITM3. NHBE and CD8+ T-cell subsets methylation data separated by rs34481144 genotype (A/G and G/G). Methylated CpGs are shown as a black circle, and unmethylated CpGs are shown as a white circle. Accompanying bar graphs show percent methylation at each of the CpGs across each cell subset with error bars representing SEM. Two-tailed unpaired T-test was used to analyze the difference between TN and TEM in individuals with G/G genotype (P=0.03). b.

Identification of CTCF binding to 40-bp, double stranded probes with differing genotypes of rs34481144 (A or G) and differing methylation status. Band intensity for each reaction is plotted and overlaid on the WEMSA. c. Anti-IFITM3 western blot using lysates from CD8+ T-cells enriched from PBMCs of healthy donors and grouped by rs34481144 genotype. To the right, quantification of IFITM3 protein band intensities, relative to those of β-Actin. Gels are cropped to include only relevant samples. Data represent mean ± SEM of each genotype group. d. CD8+ T-cell number and frequency from FLU09 patient Day 0 NWC samples based on rs34481144 genotype. Welch Two-sample T-test was used to analyze the differences between genotype (frequency P=0.0096; number P=0.012). Spearman correlation analysis using Corrplot for gene expression of neighboring genes within 180kb of rs34481144 (chr11:237,460-405,459) in patients with G/G genotype (e) and A/G genotype (f), with only significant correlations (after FDR adjustment) shown and ordered by genomic position. Positive correlation represented by blue; Negative correlation represented by red; see Supplemental Table 3 for individual correlation analysis (R and P-value) with *IFITM3*.

Table 1

Demographics table for the three cohorts used for association analyses

Study	N	Severe phenotype description	% Severe	% Female	Age Range (years)	% Caucasian	% African American	% Asian
Flu09	86	High total symptom score	10.5	60.5	<1–70	18.6	81.4	0
Genentech Challenge Study	42	Rapid viral replication	45.2	35.7	18–45	100	0	0
PICFlu	265	Mortality	6.9	41.1	<1–18.9	73.2	16.7	4.2