Genetic analysis of isoform usage in the human anti-viral response reveals influenza-specific regulation of *ERAP2* transcripts under balancing selection

Chun Jimmie Ye[*,§,1], Jenny Chen[*,2,3], Alexandra-Chloé Villani[2,4], Meena Subramaniam[1,5], Rachel E. Gate[1,5], Tushar Bhangale[6], Mark N. Lee[2,4,7], Towfique Raj[2,7,8,13], Raktima Raychowdhury[2], Weibo Li[2], Noga Rogel[2], Selina H. Imboywa[7], Portia I. Chipendo[7], Cristin McCabe[2,8], Michelle H. Lee[7], Irene Y. Frohlich[7], Barbara E. Stranger[9], Philip L. De Jager[2,7,8,14], Aviv Regev[2,10,11], Tim Behrens[6], Nir Hacohen[§,2,4]

[1]Institute for Human Genetics, Institute for Health and Computational Sciences, Department of Biostatistics and Epidemiology, Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA

[2]Broad Institute of MIT and Harvard, 415 Main Street, Cambridge MA 02142, USA

[3]Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139

[4]Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, MA 02139

[5]Biomedical Informatics Program, University of California, San Francisco

[6]Genentech Inc., South San Francisco, CA

[7]Harvard Medical School, Boston, MA 02116

[8]Departments of Neurology and Psychiatry, Brigham and Women's Hospital, Boston, MA 02115, USA.

[9]Section of Genetic Medicine, Department of Medicine, Institute for Genomics and Systems Biology, Center for Data Intensive Science, The University of Chicago, Chicago, IL 60637

[10]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

[11]Howard Hughes Medical Institute, Chevy Chase, MD 20815

[13]Present address: Ronald M. Loeb Center for Alzheimer's Disease, Departments of Genetics and Genomic Sciences, and Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[14]Present address: Center for Translational & Computational Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York, NY 10032.

[*]These authors contributed equally to this work.

[§]Corresponding authors. Email: nhacohen@mgh.harvard.edu (N.H); jimmie.ye@ucsf.edu (C.J.Y.)

Abstract

While the impact of common genetic variants on transcript abundance in response to cellular stimuli has been analyzed in depth, less is known about how stimulation modulates the genetic control of isoform usage. Using RNA-seq profiles of monocyte-derived dendritic cells from 243 individuals, we uncovered thousands of unannotated isoforms synthesized in response to viral infection or stimulation with Type 1 interferon. We identified more than a thousand single nucleotide polymorphisms associated with isoform usage (isoQTLs), many of which are independent of expression QTLs for the same gene. Compared to eQTLs, isoQTLs are enriched for splice sites and untranslated regions, and depleted of upstream sequences. In five loci, they provide a possible mechanism of action underlying DNA variants associated with immune-related disorders. Among these five is the *ERAP2* locus, where the major haplotype is under balancing selection and associated with Crohn's disease risk. At baseline and following Type 1 interferon stimulation, the major haplotype is associated with absence of *ERAP2* expression; but in response to influenza infection, the major haplotype results in the expression of two previously uncharacterized, alternatively transcribed, spliced and translated short isoforms. Thus, genetic variants at a single locus could modulate independent gene regulatory processes in the innate immune response, and in the case of *ERAP2*, may confer a historical fitness advantage in response to virus, but increase risk for autoimmunity in the modern environment.

Introduction

An important aspect of eukaryotic gene regulation is the control of alternative gene isoforms. This is achieved through several mechanisms at the transcript level, including: alternative promoters for transcription initiation, alternative splicing of pre-messenger RNA, alternative polyadenylation, and selective degradation of isoforms. These processes regulate the relative abundances of multiple coding and non-coding RNAs from the same underlying DNA sequence, often resulting in altered function of the protein products in response to developmental or environmental changes[1-4].

A case in point is the role of alternative isoform usage in the human immune response. For example, studies have shown that alternative splicing is critical across many immune processes, such as B cell functions reflected in the balance between IgM and IgD immunoglobulin isoforms[5], naïve and memory T cell functions controlled by *CD45* isoforms[6], and innate immune responses to pathogens regulated by different isoforms of *MYD88*[7]. Genetic variants that affect isoform usage have been associated with immune disorders[8] including the association of systemic lupus erythematosus (SLE) with common variants in a splice site at the *IRF5* locus[9].

Previous studies have identified shared and divergent transcriptional programs in the antibacterial and antiviral response of innate immune cells[10,11], with genetic variation imparting stimulation specific effects on gene expression[11-14]. While genetic maps of alternative splicing are beginning to emerge, most notably in lymphoblastoid cell lines[15,16], across healthy human tissues[17,18], and in macrophages stimulated with bacteria[19], variability in isoform usage across individuals and its genetic basis in the human antiviral response have not been studied.

Here, we integrate RNA-sequencing with dense genotyping to systematically investigate the genetic control of isoform usage in monocyte derived dendritic cells (MoDCs) at rest, and in response to influenza-infection or Type 1 interferon. Because the Type 1 interferon pathway is known to be engaged by a broad array of microbial products, our study design is unique in allowing the separation of the universal and influenza-specific effects on the interferon-induced response. Since the human transcriptome has never been annotated under these conditions, we first used *de novo* assembly to catalog and quantify all synthesized isoforms in resting and stimulated cells. Then, by harnessing the natural transcriptomic and genetic variation in the ImmVar cohort[11,20,21], we mapped genetic variants (isoQTLs) associated with isoform usage. Systematic characterization of isoQTLs, especially in comparison to eQTLs, provides mechanistic insights into the genetic control of different aspects of gene regulation and enables the functional interpretation of loci associated with immune disease and under natural selection.

Results

Influenza and Type I interferon stimulate widespread alternate isoform usage

We used paired-end RNA-seq to profile the transcriptomes of primary monocyte-derived dendritic cells (MoDCs) from healthy donors at rest (N = 99), and following stimulation with either influenza ΔNS1 (a strain engineered to maximize the IFNβ-induced host response to infection by the deletion of a key virulence factor[22]) (N = 250) or interferon beta (IFNβ), a cytokine that stimulates anti-viral effectors (N = 227). A total of 552 pass-filter samples (out of 576), 84 from all three conditions, 127 from both stimulation conditions, and 46 from only one condition, were analyzed (**Table S1**). To define the corpus of transcripts in human dendritic cells at rest and in response to stimulation, including previously unannotated transcripts, we assembled the transcriptome *de novo* in each sample (individual-condition pair), retained only expressed isoforms (> 5 transcripts per million in any sample), then combined isoforms across all samples to enable direct comparisons between conditions. Overall, we identified 35,411 transcripts: 18,644 present in resting cells, and 29,841 (flu) and 25,127 (IFNβ) present in

stimulated cells. The 35,411 transcripts correspond to 15,123 genes, including 8,338 previously unannotated transcription start sites (TSSs) (corresponding to 5,414 genes), 16,062 previously unannotated splice sites (corresponding to 11,704 isoforms and 6,703 genes), and 1,653 previously unannotated transcripts (corresponding to 1,281 genes) (**Table S2**).

Compared to IFNβ induction, flu infection elicited a prominent change in isoform usage independent of gene expression, estimated as the ratio of isoform abundance over total gene abundance. Relative to baseline, the usage of 3x as many isoforms (4,937 vs 1,651) were altered in flu-infected compared to IFNβ-stimulated cells (beta regression, FDR < 0.01, isoform abs($\log_2$(fold change)) > 1, **Table S3**, **Fig. S1**). In response to both conditions, more than 50% of isoforms with differential usage were previously unannotated, highlighting the inadequacy of current annotations in describing the full diversity of gene isoforms in the human antiviral response. Of the differentially expressed genes with more than one isoform, 36% (flu, 1326/3680) and 22% (IFNβ, 433/1995) had at least one isoform that differed in usage (**Fig. 1A**), suggesting independent regulatory mechanisms that control overall gene abundance and specific isoform usage in response to stimuli.

Isoforms that differed in usage partitioned into five prominent clusters (**Fig. 1B**, k-means clustering). Isoforms with decreased usage in response to both stimuli (cluster IV) were enriched for mitochondria function (GO:0044429, $P < 2.1\times10^{-7}$, **Table S4**) while those in response to one stimulus were either enriched for broad biological function (flu, cluster V, **Table S4**) or not enriched for known gene ontology entries (IFNβ, cluster III). Isoforms with increased usage in response to influenza were either shared with (cluster II) or independent of IFNβ stimulation (cluster I). Isoforms upregulated in response to flu (cluster I and II) were enriched for known influenza responsive genes (KEGG: Influenza A, $P < 4.2\times10^{-4}$ and $P < 7.6\times10^{-2}$, **Table S5**) while isoforms specific to flu-infection were also enriched for viral sensing genes including the Toll-like receptor (TLR) pathway (KEGG: Toll Like Receptor Signaling Pathway, P <

3.9×10^{-4}, **Table S5**). Among the genes within the flu-specific cluster was *TLR4* (**Fig. 1C**), the toll-like receptor classically associated with sensing bacterial ligands but have been shown to also sense viral products[23]. In flu-infected cells, the usage of longer isoforms with an upstream alternative start site (*TLR4*/Iso1 and *TLR4*/Iso2) was decreased, while the usage of *TLR4*/Iso3, *TLR4*/Iso4, and *TLR4*/Iso7 was increased. Although *TLR4*/Iso4 encodes the annotated 839 amino acid (aa) product, isoforms *TLR4*/Iso3 and *TLR4*/Iso7 encode shorter, 799 aa products each with a truncated extracellular domain missing a predicted signal peptide. Additionally, we also found flu-specific isoform control of *CASP8* (**Fig. 1D**), where usage of short isoforms (*CASP8*/Iso4, *CASP8*/Iso5, *CASP8*/Iso6) was decreased only in flu-infected cells. Although *CASP8* is known to induce the apoptotic program via Fas-associated death domain (FADD) protein in response to extrinsic cytokine signals, *CASP8*/Iso4 has a unique N-terminal extension of 59 amino acids, which has been reported to allow for selective recruitment to the endoplasmic reticulum[24]. These results demonstrate that changes in isoform usage independent of overall gene abundance are pervasive and affect prominent innate immune sensors and regulators in viral versus interferon response.

Genetic determinants of isoform usage are associated with distinct regulatory features compared to those of gene expression

We assessed whether common genetic variants, known to affect gene expression[11], could affect isoform usage in both resting and stimulated MoDCs. We associated over 10M imputed variants with two transcriptional traits, isoform percentage and total gene abundance, to identify isoform usage quantitative trait loci (isoQTLs) and expression quantitative trait loci (eQTLs), respectively. After adjusting for unwanted variation from latent effects (**Fig. S2** and **S3**), we identified 2,393 isoforms corresponding to 1,345 genes (linear regression, permutation FDR < 0.05, **Table S6**) with local isoQTLs (+/- 500kb of TSS) and 8,350 genes (linear regression, permutation FDR < 0.05, **Table S7**) with local eQTLs in at least one condition. A substantial proportion of leading isoQTL SNPs (58% baseline, 42% flu, 39% IFNβ)

were not significant eQTLs, suggesting that the genetic control of isoform usage and overall gene abundance are largely independent.

Genetic variants could modulate isoform usage through several mechanisms including perturbing the usage of alternate promoters, splice sites, or regulatory elements in the untranslated regions (UTRs). We compared the *cis* properties of isoQTLs and eQTLs to identify the mechanisms by which each class of variants acts. When normalized by exon and intron lengths, leading SNPs for local isoQTLs were enriched across the entire gene body (**Fig. 2A)**, in distinct contrast with leading SNPs for local eQTLs, which were enriched near TSS and transcription end site (TES). Further, when compared to a set of SNPs matched for allele frequency and distance to TSS, leading SNPs for local isoQTLs were most enriched for splice sites (4.8x baseline, 2.8x flu, 2.7x IFNβ), synonymous (1.6x baseline, 1.6x flu, 2.1x IFNβ) and missense variants (2.0x baseline, 1.4x flu, 1.9x IFNβ), and 5' (1.7x baseline, 1.1x flu, 1.4x IFNβ) and 3' (1.5x flu, 1.4x IFNβ) UTRs (**Fig. 2B**). Compared to eQTLs, isoQTLs are not enriched for binding sites of transcription factors known to play a role in myeloid cell response (**Fig. S4**). These results suggest that genetic variants associated with isoform usage likely do so via *cis* regulatory sequences that modulate alternative splicing and transcript stability.

Because alternate isoforms could result in protein products of drastically different function that may affect the expression of other genes, we next assessed the *trans* regulatory properties of isoQTLs. Testing only significant local eQTLs and isoQTLs for possible *trans* associations resulted in higher multiple testing power to detect distal QTLs, with the strongest signal observed in flu-infected cells (**Fig. 2C**, **Table S8**). In stimulated cells, the expressions of 95 (flu) and 55 (IFNβ) genes were distally associated (> 1M base pairs) with isoQTLs (linear regression, permutation FDR < 0.1) (**Table S8**), in contrast to 492 and 49 genes distally associated with eQTLs (linear regression, permutation FDR < 0.1). Leveraging the distribution of permuted P-values, we estimated that 20% (flu) and 16% (IFNβ) of genes are associated in

trans with isoQTLs and 41% (flu) and 39% (IFNβ) of genes are associated in *trans* with eQTLs. These results suggest that local eQTLs are more likely to affect *trans* gene expression than local isoQTLs and that both local eQTLs and isoQTLs are more likely to affect *trans* gene expression in influenza-infected than IFNβ-stimulated cells.

Genetic control of alternative isoform usage in responses to virus and interferon

To assess how the genetic control of isoform usage differs in response to stimuli, we analyzed 84 donors whose cells were assayed in all three conditions to enable equally powered comparisons across conditions. Genetic variants imparted stronger effects on isoform usage in resting and IFNβ-infected cells than in flu-infected cells as indicated by more isoQTLs detected (815 in resting, 784 in IFNβ and 427 in flu, permutation FDR < 0.05) and an increase in the proportion of variance of isoform usage explained (R_{iso}^2) by the associated variants (**Fig. 3A**). The correlation of R_{iso}^2 was lowest between flu-infected and resting (baseline) cells (Pearson $\rho_{flu.baseline}$=0.46 compared to $\rho_{IFN.baseline}$ =0.69 and $\rho_{IFN.flu}$ =0.66) suggesting flu-specific genetic control of isoform usage independent of Type 1 interferon signaling. Isoforms with higher R_{iso}^2 in stimulated cells are upregulated in response to stimuli, suggesting that the activation of specific gene regulatory programs that control isoform usage are sensitive to genetic effects unobserved in inactive states (**Fig. 3B**).

To directly assess how stimulation modifies the effects of individual genetic variants on isoform usage, we mapped SNPs associated with the difference in isoform usage between conditions, herein referred to as local response-isoQTLs (risoQTLs). Compared to resting cells, we identified 74 (flu) and 22 (IFNβ) significant local risoQTLs corresponding to 50 and 14 genes (permutation FDR < 0.1, **Table S9)**. Amongst the 13 genes that share local risoQTLs in both stimulated conditions were *IFI44L* and *WARS* (**Fig. 3C**), two genes whose splicing has previously been studied. *IFI44L* is a type 1 interferon-stimulated

gene that has been shown to have moderate effects in inhibiting human hepatitis virus replication *in vitro*[25] and whose splicing has been shown to be influenced by the most significant risoQTL (rs1333973)[26]. Although *WARS*, a tryptophanyl-tRNA synthesase, is primary involved in protein synthesis with other family members known to encode for catalytic null enzymes[27], it also has a known spliced form induced by IFNγ that has anti-angiogenic activity[28]. Among the 37 genes that have risoQTLs in flu-infected but not interferon-stimulated cells was *ZBP1*, a sensor of influenza infection that triggers cell death and inflammation and contributes to virus-induced lethality[29]. While influenza-infected and interferon-stimulated cells are expected to share some risoQTLs reflecting a common gene regulatory program (as interferons are induced by viral infection), influenza-specific risoQTLs confer genetic control of previously unknown viral sensing pathways independent of downstream effector (type-1 interferon) signaling.

Association of eQTLs and isoQTLs with immune-related diseases

Previous analyses of the overlap between expression QTLs and genome-wide association studies (GWAS) have aided the localization and functional interpretation of causal variants in GWAS loci. Because disease-causing variants that modulate isoform usage could have more profound effects on gene regulatory networks by altering protein structure, we compared leading local isoQTL and local eQTL SNPs with the latest GWAS catalog[30]. While local eQTLs in stimulated cells were enriched in multiple diseases including inflammatory bowel disease (flu $P < 5.46\times10^{-7}$, IFNβ $P < 5.21\times10^{-5}$), rheumatoid arthritis (flu $P < 2.49\times10^{-5}$, IFNβ $P < 0.03$), and Parkinson's disease (flu $P < 4.06\times10^{-7}$, IFNβ $P < 7.44\times10^{-5}$) (**Fig. 4A**), local isoQTLs were enriched in late onset Alzheimer's disease (flu $P < 1.15\times10^{-6}$, IFNβ $P < 1.75\times10^{-4}$), vitiligo (flu $P < 4.94\times10^{-5}$, IFNβ $P < 0.42$), and systemic lupus erythmatosus (SLE) (flu $P < 0.1$, IFNβ $P < 3.62\times10^{-3}$) (**Fig. 4B**). Notably, the significant overlap with Alzheimer's loci is only found for isoQTLs and not for eQTLs (flu $P < 0.9$, IFNβ $P < 1$) due to the overlap of genes with known splicing variants (*CD33*[31,32], *CD46*[33]) and genes with unvalidated isoQTLs controlling transcript usage

(*CR1L*). These results suggest a role for both variants that affect isoform usage and gene expression in mediating autoimmune and neurodegenerative disease risk.

One of the disease-associated isoQTLs lies in the *IRF7* locus within an extended haplotype known to be associated with SLE (rs58688157 lead SNP, $P < 2.97\times10^{-11}$)[34] (**Fig. 4B**). A linked SNP (rs1061502; LD $R^2 = 0.93$, D' = 0.97 to rs58688157) is the most significant association to overall *IRF7* expression in IFNβ-stimulated ($P < 2.87\times10^{-49}$) and flu-infected cells ($P < 2.21\times10^{-25}$) (**Fig. 4C, Fig. S5**). IsoQTL analysis further revealed that rs1061502^T also increased the usage of *IRF7*/Iso4 (flu beta = 5.3%, $P < 9.42\times10^{-52}$, IFNβ beta = 5.8%, $P < 4.81\times10^{-46}$, **Fig. 4C**, panel 3, purple) while decreasing the usage of *IRF7*/Iso3 (flu beta = -3.4%, 1.15×10^{-15}, IFNβ beta = -7.0%, 2.04×10^{-29}, **Fig. 4C**, panel 3, green). Further, although the overall *IRF7* abundance was similar between the two stimulated conditions, *IRF7*/Iso4 (purple) was the dominant isoform in flu-infected cells, but not IFNβ-stimulated cells (10.7x fold, $P < 10^{-306}$) (**Fig. 4C**, bottom panel). rs1061502 was also a distal eQTL (permutation FDR < 0.2) for a cluster of genes including *NMI*, type-1 interferon *IFNA2*, *IFIT5*, and *C5* only in flu-infected, but not in IFNβ-stimulated cells (**Fig. 4D**). These results replicate and expand our previous findings that rs12805435 (LD $R^2 = 0.95$, D' = 0.98 to rs1061502) is associated in *cis* with *IRF7* expression in these two conditions, and in *trans* with a cluster of *IRF7*-regulated genes only in flu-infected cells[11]. The flu-specific *trans* associations could be due to the additive effects of flu-specific induction of *IRF7*/Iso4 independent of IFNβ signaling and induction of overall *IRF7* expression by rs1061502^T. *IRF7*/Iso4 encodes a 516-amino acid protein product and differs from other dominant isoforms in IFNβ-stimulated cells (*IRF7*/Iso1 and *IRF7*/Iso3) in the 5'UTR and the coding sequence in the DNA-binding domain. Given the link between a type-1 interferon signature and SLE, this raises an intriguing notion that SNPs affecting a specific *IRF7* isoform could impact viral responses and autoimmune inflammation through similar mechanisms.

An *ERAP2* risoQTL controls differential transcript usage during influenza infection

The *ERAP2* locus is characterized by two frequent and highly differentiated (40 SNPs in perfect LD) haplotypes observed in every major human population (B: 53% and A: 47%) (**Fig. S6**). The major allele (G) of rs2248374, a splice-site variant tagging Haplotype B, creates an alternate 3' donor splice site inducing the splicing of an extended exon 10 with two premature termination codons[35]. As a result, transcripts from Haplotype B are degraded by nonsense-mediated decay resulting in one of the most significant eQTLs and isoQTLs in most tissues and cell types[11,15,17,21]. Intriguingly, while Haplotype B is associated with increased risk for Crohn's disease[36] (**Fig. 5A**), it is also maintained by long term balancing selection (between 1.4M[35] and 5.1M years[37]). This raises the important question: in what environmental condition does balancing selection act to maintain the seemingly loss-of-function (LOF), disease-causing haplotype in humans?

Given the known role of *ERAP2* in antigen presentation[38], we examined the genetic control of *ERAP2* transcripts in the human antiviral response. In resting and IFNβ-stimulated cells, we confirmed the known genetic association of rs2248374^{G} allele with lower *ERAP2* expression (**Fig. 5B**). Remarkably, while the overall abundance of *ERAP2* was elevated in stimulated conditions, two previously uncharacterized short isoforms (*ERAP2*/Iso2, *ERAP2*/Iso3, **Fig. 5B, Fig. S7**) were transcribed from Haplotype B only in flu-infected and not IFNβ-stimulated cells, resulting in the partial rescue of *ERAP2* expression. The short isoforms differed from the constitutive full-length isoform (*ERAP2*/Iso1 transcribed from Haplotype A) by the initiation of transcription at exon 9 and the alternate splicing of an extended exon 10, and differed from each other by alternative splicing at a secondary splice site at exon 15. The initiation of transcription at exon 9 results in an alternate in-frame translation start site at exon 11 thus rendering the premature termination codon in exon 10 inactive. The influenza-dependent genetic control of *ERAP2* isoform usage is further supported by (*i*) correlation between overall flu transcript abundance, a proxy for degree of infection, and *ERAP2*/Iso2 (R^2 = 0.57) and *ERAP2*/Iso3 (R^2 = 0.7) (**Fig. 5C** triangles), (*ii*) evidence of alternative translation starting at exon 11 as shown by the detection of flu-specific protein isoforms (50

kDa) in flu-infected cells from Haplotype B homozygotes and heterozygotes (**Fig. 5D**), and (*iii*) evidence of transcription of *ERAP2*/Iso2 or *ERAP2*/Iso3 (marked by an extended exon 10) in monocyte derived macrophages infected by H3N2 over a time course (fluomics, GEO GSE97672) (**Fig. 5E**).

The complex genetic signals at the *ERAP2* locus is consistent with three perfectly linked variants on Haplotype B affecting *ERAP2* transcription and splicing in response to viral stimulation independent of Type 1 interferon signaling (**Fig. 5F**). Rs2548538, an intronic variant that overlaps chromatin marks from LCLs[39], likely causes alternate transcript initiation at exon 9. Rs2248374^{G}, the known splice site mutation, creates an alternate preferred splice site resulting in alternative splicing of an extended exon 10. Rs2549797^{G}, a splice-site mutation that creates a competing alternate splice site, results in ~40% of the transcripts with an extended exon 15. The signature of natural selection, the previous disease associations, and the viral specific transcription suggest a critical antiviral role for the short *ERAP2* isoforms that could also result in an overactive autoinflammatory response in Crohn's disease.

Discussion

Although maps of genetic variants associated with overall transcript abundance have been generated in many tissue types, the genetic control of alternate isoform usage has not been extensively studied. Using *de novo* transcript reconstruction, we found a large number of previously uncharacterized transcripts in human dendritic cells, especially in response to influenza and interferon stimulation, indicating that the current reference human transcriptome is far from complete. We further found genetic variants (isoQTLs) associated with alternate isoform usage are widespread, approximately half of which are not associated with the overall abundance of the corresponding gene, indicative of independent genetic control of gene regulation at most loci of the genome. The enrichment of isoQTLs for known splice sites and disease loci suggest a highly clinically relevant set of candidate loci that would induce targetable changes in protein sequence.

IsoQTLs, like eQTLs, can affect gene expression at other loci in the genome suggesting important downstream effects on gene regulation. The most striking example is at the *IRF7* locus where a splice-site SNP affects *IRF7* splicing in response to influenza and interferon, but only affects the expression of downstream genes in response to flu. This suggests that both genetic effects on isoform usage and stimulation dependent regulation of *IRF7* expression are necessary for the observed *trans* effects. Although C-terminal splice forms of *IRF7* have been shown to differentially transactivate type-1 interferons and chemokines[40], *IRF7*/Iso4 is not known to have specific antiviral properties *in vivo* even though its ectopic expression is known to activate IFNAs in fibroblasts[41]. The association of the variant with SLE indicates a possible role for viral exposure to prime the immune system of individuals carrying the risk allele toward autoimmunity.

Different genetic variants in a locus could also affect multiple facets of gene regulation in response to stimulation in establishing transcriptome diversity and susceptibility to disease. This was clearly demonstrated at the *ERAP2* locus where multiple variants on the Crohn's disease-associated haplotype lead to differential expression and splicing of the transcript in response to influenza. Previous experimental evidence has shown that full length *ERAP2* is a prototypical aminopeptidase that heterodimerizes with *ERAP1*[38] to perform peptide trimming during MHC class I presentation. The lack of the aminopeptidase domain in the flu-specific *ERAP2* isoforms suggests that it could interact with *ERAP1* to negatively influence antigen presentation or adopt previously unknown immunological function. Altogether, this dataset can help elucidate the mechanisms underlying disease alleles by providing deeper molecular data for each gene in baseline and inflammation.

Figure 1

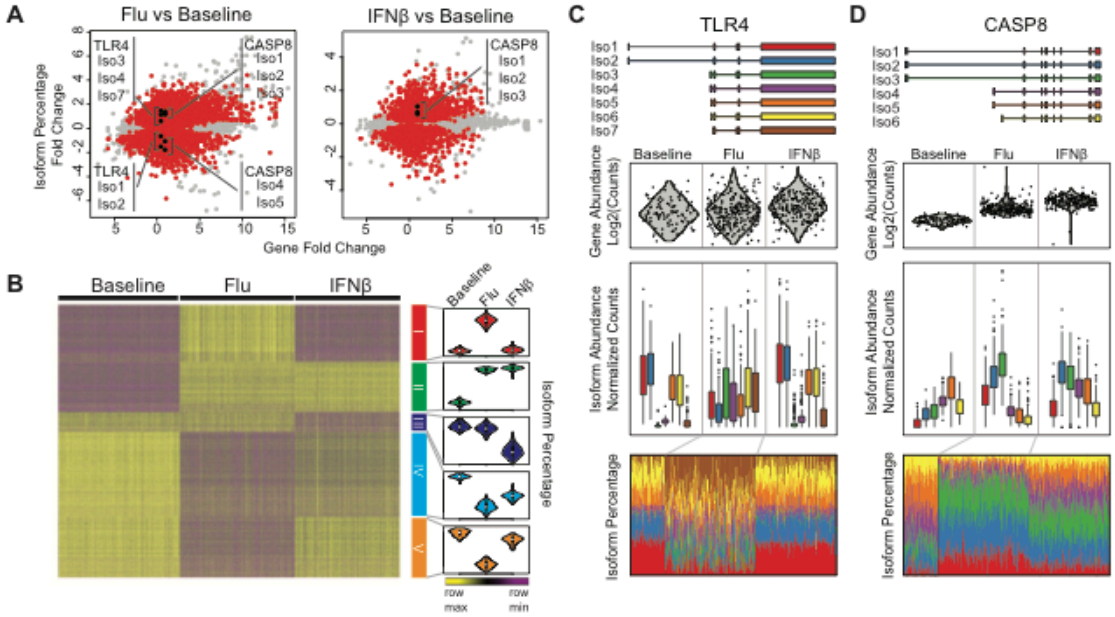


Figure 1 - Transcriptome changes in response to stimulation. (A) Scatter plot of fold change of overall gene abundance (ln(fold change), x-axis) versus fold change in isoform percentage (y-axis) in flu-infected (left) and IFNβ-simulated (right) cells compared to baseline. Each dot represents one isoform. Isoforms that significantly differed in their usage percentages (beta regression, FDR < 0.01) are highlighted in red. (B) Clustering of isoform percentages in baseline, flu-infected, and IFNβ-stimulated cells. Heatmap colors are row-scaled (yellow: row max, purple: row min) (left). Violin plots (right) summarize isoform percentages of all isoforms within a cluster, within each condition. Only isoforms that significantly changed (beta regression, FDR < 0.01, abs($\log_2$(fold change)) > 1) in usage are shown. *De novo* constructed isoforms (top panel), overall gene abundance (middle panel), and isoform abundance and percentage (bottom panels) for (C) *TLR4* and (D) *CASP8*.

Figure 2

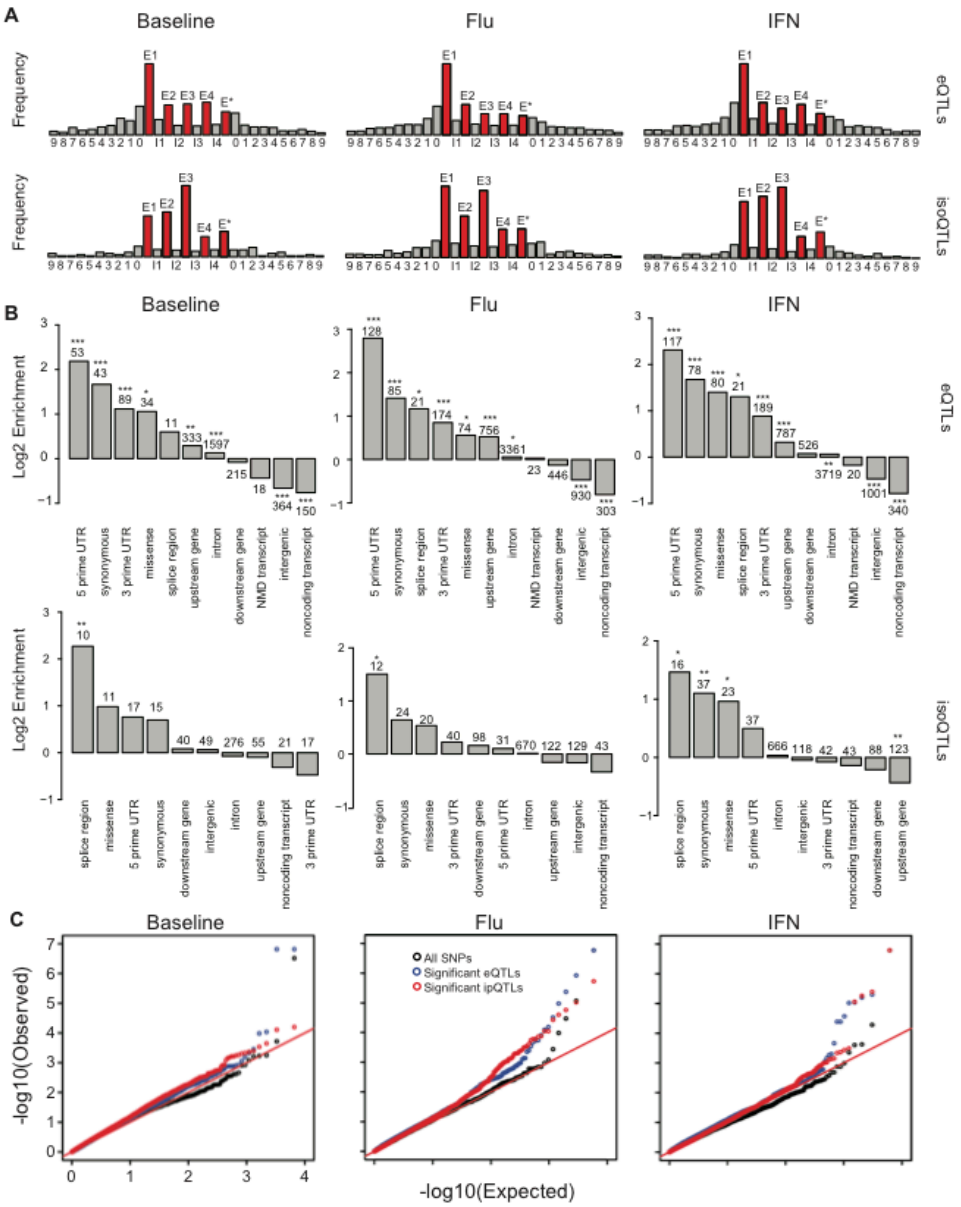


Figure 2 - Properties of local eQTLs and isoQTLs. (A) Frequency (y-axis) of the location of significant local eQTLs and isoQTLs (permutation FDR < 0.05) with respect to meta gene structure (x-axis). Genes are normalized to five exonic regions (E1 - E4: exons 1 - 4, E*: exon 5 - last exon) and four intronic regions (introns 1, 2, 3, and 4 – last intron). Upstream and downstream sequences are divided into 100kb windows. (B) Log_2 fold enrichment (y-axis) of significant local eQTLs and isoQTLs (permutation FDR < 0.05) for genomic annotations. eQTL enrichments are calculated using a background set of SNPs matched for distance to TSS and allele frequency. isoQTL enrichments are calculated with respect to a background set of eQTLs matched for distance to TSS and allele frequency. Multiple testing significance is indicated (*: P < 0.05, **: P < 0.01, ***: P < 0.001). (C) Observed (y-axis) versus expected (x-axis) quantile-quantile plots of distal associations (> 1from TSS) for all SNPs (black), eQTLs (blue) and isoQTLs (red).

Figure 3

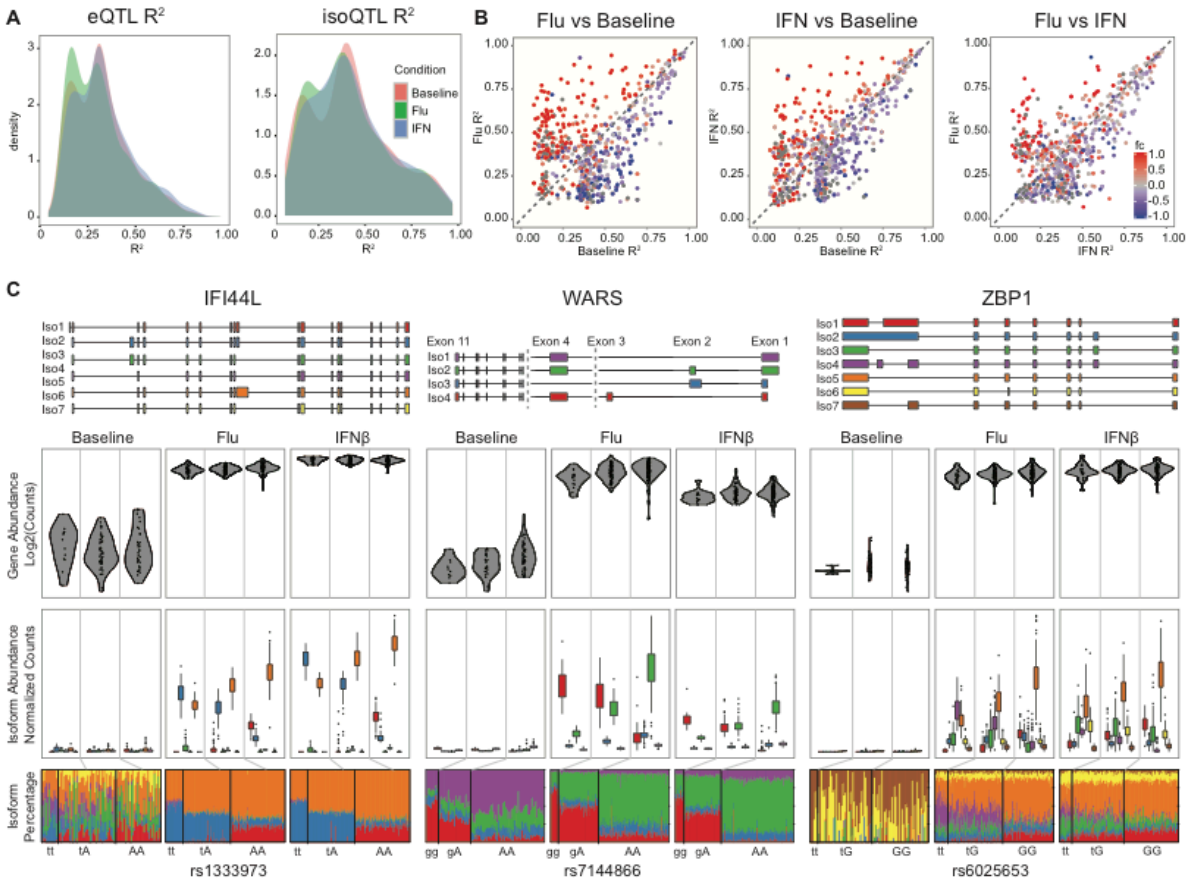


Figure 3 – Comparison of local isoQTLs between conditions. (A) Effect size (R^2) distribution of significant *cis* eQTLs and isoQTLs (permutation FDR < 0.05) at baseline (red), post flu-infection (green) or IFNβ-stimulation (blue). (B) Correlation of effect sizes (R^2) for significant local isoQTLs (permutation FDR < 0.05) between pairs of conditions. Transcripts are colored by differential expression (red: up-regulated, blue: down-regulated). (C) *De novo* constructed transcript structure (top panel) and box-whisker plots (bottom 3 panels) between transcript quantitative traits (y-axis: $\log_2$(normalized gene abundance), normalized isoform abundance, or isoform percentage) and genotype (x-axis) for 3 genes (*IFI44L*, *WARS* and *ZBP1*) with risoQTLs.

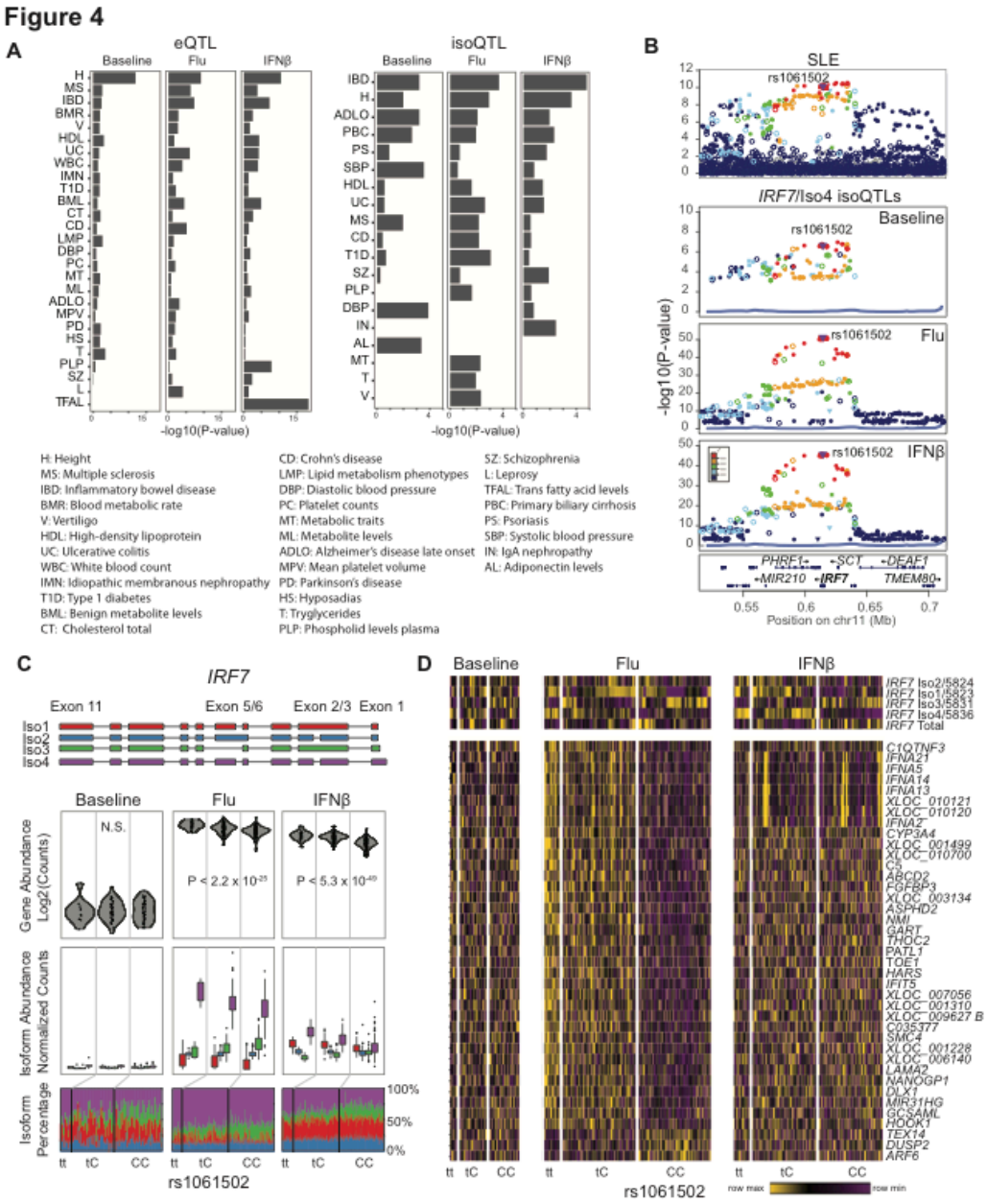


Figure 4 - GWAS enrichment of local eQTLs and isoQTLs. (A) Enrichment of leading eQTL and isoQTL SNPs for GWAS hits. Diseases enriched (FDR < 0.1) for significant local e/isoQTLs (permutation FDR < 0.05) detected at baseline, post flu-infection or IFNβ-stimulation are plotted. X-axis: $-\log_{10}$(hypergeometric enrichment P-value). (B) LocusZoom plots of the *IRF7* region for SLE GWAS associations (top panel) and *IRF7*/Iso4 isoQTLs for baseline, flu-infected or IFNβ-stimulated cells (bottom 3 panels). Y-axis: $-\log_{10}$(P-value) of association. X-axis: genomic location. Points are colored based on their LD to rs1061502. (C) Transcript structure and box-whisker plots (bottom 3 panels) between *IRF7* transcript quantitative traits (y-axis: $\log_2$(normalized gene abundance), normalized isoform abundance, or isoform percentage) and rs1061502 genotype (x-axis). (D) Heatmap of genes distally associated (permutation FDR < 0.05) with risoQTL rs1061502. Heatmap colors are row-scaled TPM values (yellow: row max, purple: row min).

Figure 5

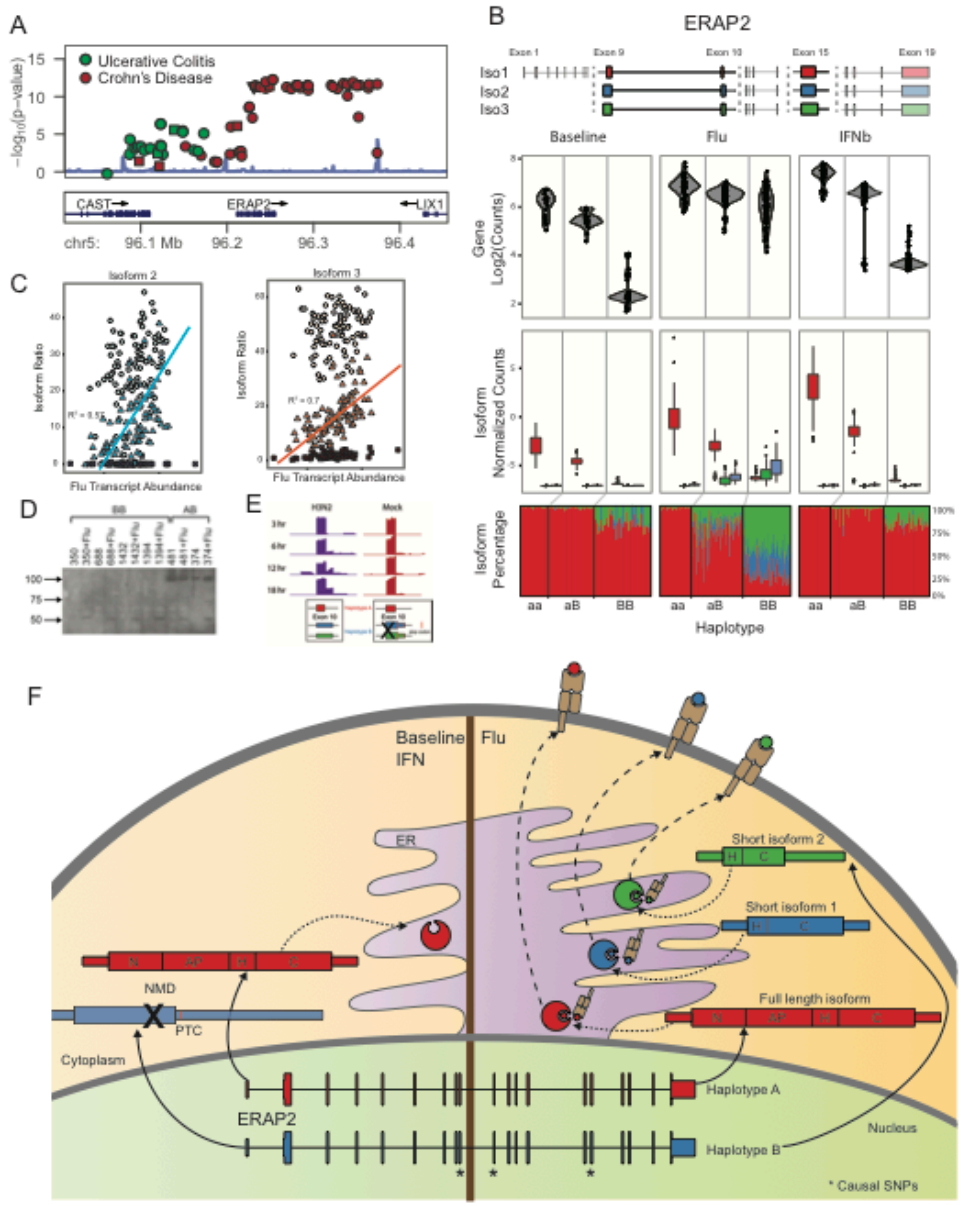


Figure 5 - Genetics of *ERAP2* regulation. (A) LocusZoom plot of Crohn's disease and ulcerative colitis associations at the *ERAP2* locus. Y-axis: -$\log_{10}$(P-value) of association. X-axis: genomic location. (B) Transcript structure (top panel) and box-whisker plots (3 bottom panels) between *ERAP2* transcript quantitative traits (y-axis: $\log_2$(normalized gene abundance), normalized isoform abundance, or isoform percentage) and genotype (x-axis). (C) Correlation between *ERAP2*/Iso2 (blue) and *ERAP2*/Iso3 (orange) usage percentage (y-axis) and overall flu transcript abundance (x-axis) segregated by genotype (aa: squares, aB: triangles, BB: circles). (D) Western blot of MoDCs before and after flu-infection from 5 Haplotype B homozygotes and 2 heterozygotes. Full length *ERAP2* protein isoform is expected at 120 kDas. Two flu-specific *ERAP2* protein isoforms are expected at 49 and 29 kDas. (E) A schematic of the hypothesized regulation and function of two *ERAP2* haplotypes. N - N terminal domain. AP - amino peptidase domain. H - hinge domain. C - C terminal domain. PTC - premature termination codon.

References

1. Black, D.L. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72**, 291-336 (2003).
2. Maquat, L.E. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat Rev Mol Cell Biol* **5**, 89-99 (2004).
3. Matlin, A.J., Clark, F. & Smith, C.W. Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* **6**, 386-98 (2005).
4. Wang, E.T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-6 (2008).
5. Enders, A. *et al.* Zinc-finger protein ZFP318 is essential for expression of IgD, the alternatively spliced Igh product made by mature B lymphocytes. *Proc Natl Acad Sci U S A* **111**, 4513-8 (2014).
6. Berard, M. & Tough, D.F. Qualitative differences between naive and memory T cells. *Immunology* **106**, 127-138 (2002).
7. Martinez, N.M. & Lynch, K.W. Control of alternative splicing in immune responses: many regulators, many predictions, much still to learn. *Immunol Rev* **253**, 216-36 (2013).
8. Xiong, H.Y. *et al.* RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806 (2015).
9. Graham, R.R. *et al.* Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. *Proc Natl Acad Sci U S A* **104**, 6758-63 (2007).
10. Amit, I. *et al.* Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326**, 257-63 (2009).
11. Lee, M.N. *et al.* Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science* **343**, 1246980 (2014).
12. Fairfax, B.P. & Knight, J.C. Genetics of gene expression in immunity to infection. *Curr Opin Immunol* **30**, 63-71 (2014).
13. Barreiro, L.B. *et al.* Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 1204-1209 (2012).
14. Quach, H. *et al.* Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* **167**, 643-656 e17 (2016).
15. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506-11 (2013).
16. Li, Y.I. *et al.* RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600-4 (2016).
17. Consortium, G.T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
18. Rivas, M.A. *et al.* Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* **348**, 666-9 (2015).
19. Nedelec, Y. *et al.* Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell* **167**, 657-669 e21 (2016).
20. Raj, T. *et al.* Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**, 519-23 (2014).
21. Ye, C.J. *et al.* Intersection of population variation and autoimmunity genetics in human T cell activation. *Science* **345**, 1254665 (2014).
22. Shapira, S.D. *et al.* A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell* **139**, 1255-67 (2009).
23. Doyle, S. *et al.* IRF3 mediates a TLR3/TLR4-specific antiviral gene program. *Immunity* **17**, 251-63 (2002).

24. Breckenridge, D.G., Nguyen, M., Kuppig, S., Reth, M. & Shore, G.C. The procaspase-8 isoform, procaspase-8L, recruited to the BAP31 complex at the endoplasmic reticulum. *Proc Natl Acad Sci U S A* **99**, 4331-6 (2002).
25. Schoggins, J.W. *et al.* A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472**, 481-5 (2011).
26. Lalonde, E. *et al.* RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. *Genome Res* **21**, 545-54 (2011).
27. Lo, W.S. *et al.* Human tRNA synthetase catalytic nulls with diverse functions. *Science* **345**, 328-32 (2014).
28. Wakasugi, K. *et al.* A human aminoacyl-tRNA synthetase as a regulator of angiogenesis. *Proc Natl Acad Sci U S A* **99**, 173-7 (2002).
29. Kuriakose, T. *et al.* ZBP1/DAI is an innate sensor of influenza virus triggering the NLRP3 inflammasome and programmed cell death pathways. *Sci Immunol* **1**(2016).
30. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* **45**, D896-D901 (2017).
31. Hernandez-Caselles, T. *et al.* A study of CD33 (SIGLEC-3) antigen expression and function on activated human T and NK cells: two isoforms of CD33 are generated by alternative splicing. *J Leukoc Biol* **79**, 46-58 (2006).
32. Raj, T. *et al.* CD33: increased inclusion of exon 2 implicates the Ig V-set domain in Alzheimer's disease susceptibility. *Hum Mol Genet* **23**, 2729-36 (2014).
33. Russell, S.M., Sparrow, R.L., McKenzie, I.F. & Purcell, D.F. Tissue-specific and allelic expression of the complement regulator CD46 is controlled by alternative splicing. *Eur J Immunol* **22**, 1513-8 (1992).
34. Morris, D.L. *et al.* Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. *Nat Genet* **48**, 940-6 (2016).
35. Andres, A.M. *et al.* Balancing selection maintains a form of ERAP2 that undergoes nonsense-mediated decay and affects antigen presentation. *PLoS Genet* **6**, e1001157 (2010).
36. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119-24 (2012).
37. Cagliani, R. *et al.* Genetic diversity at endoplasmic reticulum aminopeptidases is maintained by balancing selection and is associated with natural resistance to HIV-1 infection. *Hum Mol Genet* **19**, 4705-14 (2010).
38. Saveanu, L. *et al.* Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat Immunol* **6**, 689-97 (2005).
39. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
40. Lin, R., Mamane, Y. & Hiscott, J. Multiple regulatory domains control IRF-7 activity in response to virus infection. *J Biol Chem* **275**, 34320-7 (2000).
41. Au, W.C., Moore, P.A., LaFleur, D.W., Tombal, B. & Pitha, P.M. Characterization of the interferon regulatory factor-7 and its potential role in the transcription activation of interferon A genes. *J Biol Chem* **273**, 29210-7 (1998).

Supplementary Materials

Study subjects
Donors were recruited from the Boston community and gave written informed consent for the studies. Individuals were excluded if they had a history of inflammatory disease, autoimmune disease, chronic metabolic disorders or chronic infectious disorders. Donors were between 18 and 56 years of age (avg. 29.9).

Reagents
Influenza A (PR8 ΔNS1) was prepared as described[1]. Recombinant human IFNβ was obtained from PBL Assay Science (Piscataway, NJ). Antibodies used were anti-IRF1 (sc-497x; Santa Cruz Biotechnology; Dallas, TX), anti-STAT2 (sc-476x; Santa Cruz Biotechnology) and anti-IRF9 (sc-10793x; Santa Cruz Biotechnology).

Preparation and stimulation of primary human monocyte-derived dendritic cells
As previously described[2], 35-50 mL of peripheral blood from fasting subjects was collected between 7:30-8:30 am. The blood was drawn into sodium heparin tubes and peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll-Paque (GE Healthcare Life Sciences; Uppsala, Sweden) centrifugation. PBMCs were frozen in liquid N_2 in 90% FBS (Sigma-Aldrich; St. Louis, MO) and 10% DMSO (Sigma-Aldrich). Monocytes were isolated from PBMCs by negative selection using the Dynabeads Untouched Human Monocytes kit (Life Technologies; Carlsbad, CA) modified to increase throughput and optimize recovery and purity of $CD14^{+}CD16^{lo}$ monocytes: the FcR Blocking Reagent was replaced with Miltenyi FcR Blocking Reagent (Miltenyi; Bergisch Gladbach, Germany); per mL of Antibody Mix, an additional 333 ug biotinylated anti-CD16 (3G8), 167 ug biotinylated anti-CD3 (SK7) and 167 ug biotinylated anti-CD19 (HIB19) antibodies (Biolegend; San Diego, CA) were added; the antibody labeling was modified to be performed in 96-well plates; and Miltenyi MS Columns or Multi-96 Columns (Miltenyi) were used to separate magnetically-labeled cells from unlabeled cells in an OctoMACS Separator or MultiMACS M96 Separator (Miltenyi) respectively. The number of PBMCs and monocytes was estimated using CellTiter-Glo Luminescent Cell Viability Assay (Promega; Madison, WI). A subset of the isolated monocytes was stained with PE-labeled anti-CD14 (M5E2; BD Biosciences; Franklin Lakes, NJ) and FITC-labeled anti-CD16 (3G8; Biolegend), and subjected to flow cytometry analysis using an Accuri C6 Flow Cytometer (BD Biosciences). A median of 94% $CD14^{+}$ cells and 99% $CD16^{lo}$ cells was obtained. The remaining monocytes were cultured for seven days in RPMI (Life Technologies) supplemented with 10% FBS, 100 ng/mL GM-CSF (R&D Systems; Minneapolis, MN) and 40 ng/mL IL-4 (R&D Systems) to differentiate the monocytes into monocyte-derived dendritic cells (MoDCs). $4x10^4$ MoDCs were seeded in each well of a 96-well plate, and stimulated with influenza virus for 10 h, 100 U/mL IFNβ for 6.5 h or left unstimulated. Cells were then lysed in RLT buffer (Qiagen; Hilden, Germany) supplemented with 1% β-mercaptoethanol (Sigma-Aldrich).

RNA isolation and Sequencing
RNA from all samples was extracted using the RNeasy 96 kit (Qiagen, cat. # 74182), according to the manufacturer's protocols. 576 total samples were sequenced (99 baseline, 250 influenza infected and 227 interferon stimulated). 552 pass filter samples (94 baseline, 243 influenza and 215 interferon) were sequenced to an average depth of 38M 76bp paired end reads using the Illumina TruSeq kit with 86% mapping to transcriptome and 97% mapping to the genome (**Table S1**).

Adjusting for expression heterogeneity
We empirically determined the number of principal components to adjust for each stimulation condition and either overall gene abundance or isoform percentage. Because of the smaller number of individuals in the baseline study, the number of principal components adjusted is fewer (**Fig. S2**). Because the isoform

percentage implicitly adjusts for confounders that affect overall gene abundance and isoform abundance levels (i.e. other eQTLs), the number of adjusted PCs is also fewer (**Fig. S3**).

Transcriptome reconstruction

After aligning reads to the genome, we reconstructed transcriptomes for each sample individually using StringTie[3] using default parameters and quantified the abundances of annotated transcripts using kallisto[4]. For genes expressed at > 5 transcripts per million (TPM) in any sample, we removed isoforms expressed at < 5 TPM across all samples. In order to preserve isoforms that may be uniquely expressed in a single condition (e.g. baseline, flu, IFN), transcriptomes within the same conditions were first merged before transcriptomes across all three conditions were merged, using cuffcompare[5]. As a final step, cuffmerge[5] (--overhang-tolerance 0) was used to remove redundant isoforms.

Transcriptome quantification

Differential expression testing was carried out with sleuth[6] using 100 bootstraps per sample. Gene-level quantification was estimates by summing isoform counts and differential expression testing was carried out with DESeq2[7].

Beta regression

Differential isoform ratio testing was carried out in R using beta regression package *betareg*[8] and P-values were calculated using likelihood ratio test and adjusted with a false discovery rate adjustment.

DNA extraction and genotyping

As previously described[2], genomic DNA was extracted from 5 mL whole blood (DNeasy Blood & Tissue Kit; Qiagen), and quantified by Nanodrop. Each subject was genotyped using the Illumina Infinium Human OmniExpress Exome BeadChips, which includes genome-wide genotype data as well as genotypes for rare variants from 12,000 exomes as well as common coding variants from the whole genome. In total, 951,117 SNPs were genotyped, of which 704,808 SNPs are common variants (Minor Allele Frequency [MAF] > 0.01) and 246,229 are part of the exomes. The genotype success rate was greater than or equal to 97%. We applied rigorous subject and SNP quality control (QC) that includes (1) gender misidentification, (2) subject relatedness, (3) Hardy-Weinberg Equilibrium testing, (4) use concordance to infer SNP quality, (5) genotype call rate, (6) heterozygosity outlier, (7) subject mismatches. In the European population, we excluded 1,987 SNPs with a call rate < 95%, 459 SNPs with Hardy-Weinberg equilibrium P-value < 10^{-6}, 234 SNPs with a MisHap P-value < 10^{-9}, and 63,781 SNPs with MAF < 1% from (a total of 66,461 SNPs excluded). In the African-American population, we excluded 2,161 SNPs with a call rate < 95%, 298 SNPs with Hardy-Weinberg equilibrium P-value < 10^{-6}, 50 SNPs with a MisHap P-value < 10^{-9}, and 17,927 SNPs with MAF < 1% from (a total of 20,436 SNPs excluded). In the East Asian population, we excluded 1,831 SNPs with a call rate < 95%, 213 SNPs with Hardy-Weinberg equilibrium P-value < 10^{-6}, 47 SNPs with a MisHap P-value < 10^{-9}, and 84,973 SNPs with MAF < 1% from (a total of 87,064 SNPs excluded). After QC, 52 subjects across all three populations and approximately 18,000 – 88,000 SNPs in each population were filtered out from our analysis.

Population stratification: Underlying genetic stratification in the population was assessed by multi-dimensional scaling using data from the International HapMap Project[9] (CEU, YRI and CHB samples) combined with IBS cluster analysis using the Eigenstrat 3.0 software[10].

The quality control of the genotyping data were performed using PLINK[11].

Genotype imputation

To accurately evaluate the evidence of association signal at variants that are not directly genotyped, we used the BEAGLE[12] software (version: 3.3.2) to imputed the post-QC genotyped markers using reference

Haplotype panels from the 1000 Genomes Project[13] (The 1000 Genomes Project Consortium Phase I Integrated Release Version 3), which contain a total of 37.9 Million SNPs in 1,092 individuals with ancestry from West Africa, East Asia, and Europe. For subjects of European and East Asian ancestry, we used haplotypes from Utah residents (CEPH) with Northern and Western European ancestry (CEU), and combined panels from Han Chinese in Beijing (CHB) and Japanese in Tokyo (JPT), respectively. For imputing African American subjects, we used a combined haplotype reference panel consisting of CEU and Yoruba in Ibadan, Nigeria (YRI). For the admixed African American population, using two reference panels substantially improves imputation performance [ref]. After genotype imputation, we filtered out poorly imputed (BEAGLE $r^2 < 0.1$) and low MAF SNPs (MAF < 0.01), which resulted in 7.7M, 6.6M, 12.7M common variants in European, East Asian and African American, respectively. This set of genotyped and imputed markers was used for all the subsequent association analysis.

QTL mapping

QTL mapping was performed using the Matrix eQTL[14] package using empirically determined number of PCs as covariates for each analysis (**Fig. S2** and **S3**). For *cis* QTLs, isoform usage or overall gene abundance were regressed against all genetics variants with a MAF > 5% in a 1MB (+/- 500kb) window. For *trans* QTLs, overall gene abundance was regressed against genetic variants (all, only *cis* isoQTLs or only *cis* eQTLs) with a MAF > 5% outside of a 1MB window. Empirical P-values were calculated by comparing the nominal P-values with null P-values determined by permuting each isoform/gene 1000 times[15]. False discovery rates were calculated using the *qvalue*[16] package as previously described[17].

QTL annotation

QTLs were annotated using Variant Effect Predictor and Ensembl release 79[18]. Exonic and intronic locations of QTLs were determined using UCSC's canonical transcripts (table *knownCanonical*) as a reference[19]. Enrichments were calculated against background set of SNPs that were matched in allele frequency (binned by 4%) and distance to nearest transcription start site (binned by 10kb).

GWAS associations

The GREGOR suite[20] was used for calculating the enrichment of eQTLs and isoQTLs containing a GWAS loci across baseline, flu, and IFNβ stimulations. GWAS associations for disease with FDR < 0.1 are reported.

Estimating flu transcript abundance

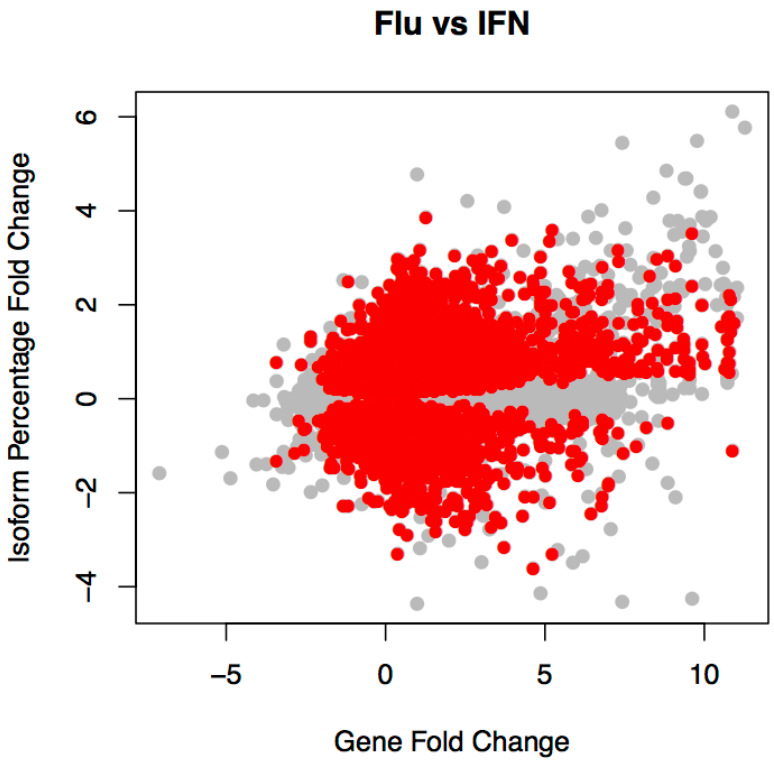
Flu transcript abundance was estimated by running RSEM[21] on a custom reference of the influenza PR8 genome.

ERAP2 Western Blot

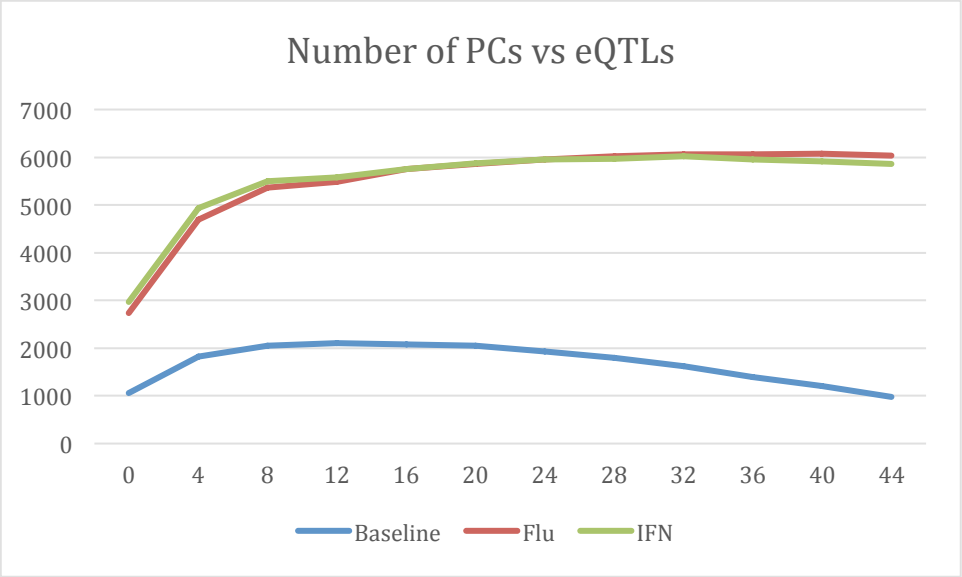
Protein extracts were fractionated by SDS-PAGE (4-12% Bis-Tris gel, Thermo scientific, NP0335BOX) and transferred to PVDF membrane (BioRad, cat. #162-0177). After blocking with 2% BSA in TBST (Tris buffered saline containing 0.1% tween-20) for 1 hour, membranes were incubated with primary antibody (either ERAP2, R&D Systems, cat# AF3830, 1:3000) or b-actin (Abcam, cat. #ab6276, 1:15,000) overnight at 4C. Membranes were then washed and incubated with a 1:5000 dilution of HRP conjugated secondary antibody (either donkey anti-goat from Santa Cruz Biotech cat. #sc2020, or with goat anti-mouse from Jackson immune Research cat. #115-035-146) for 1 hour. Membranes were washed and developed with ECL system (VWR, cat. #89168-782) according to the manufacturer's protocol.

Data Availability

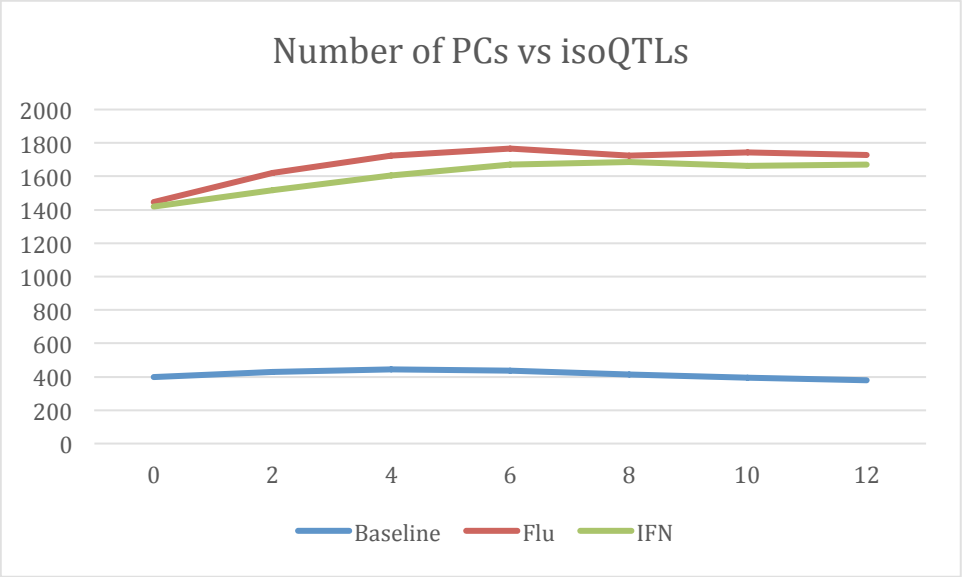
Processed RNA-Sequencing data is available from GEO under accession GSE92904. Raw fastq data is available from dbGAP under accession phs000815.v1.p1.



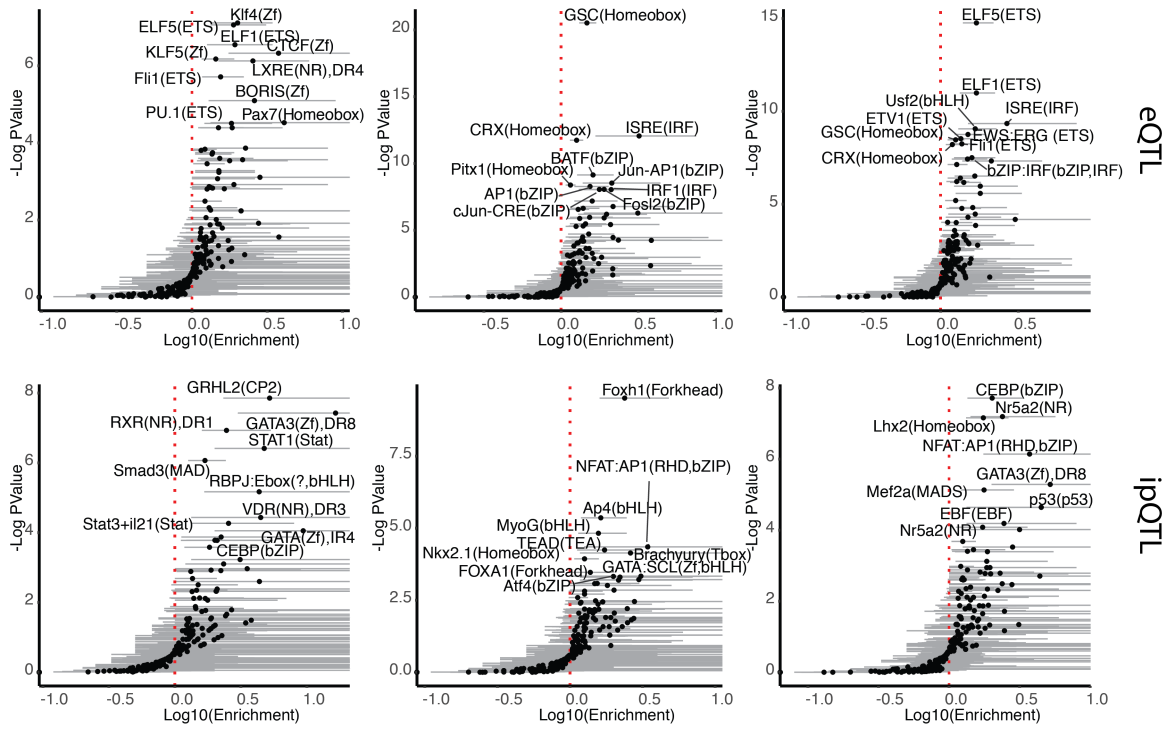
Supp. Fig. 1 – Scatter plot of fold change of overall gene abundance (ln(fold change), x-axis) versus fold change in isoform percentage (y-axis) in flu-infected cells compared to IFNβ-simulated cells. Each dot represents one isoform. Isoforms that significantly differed in their usage percentages (beta regression, FDR < 0.01) are highlighted in red.



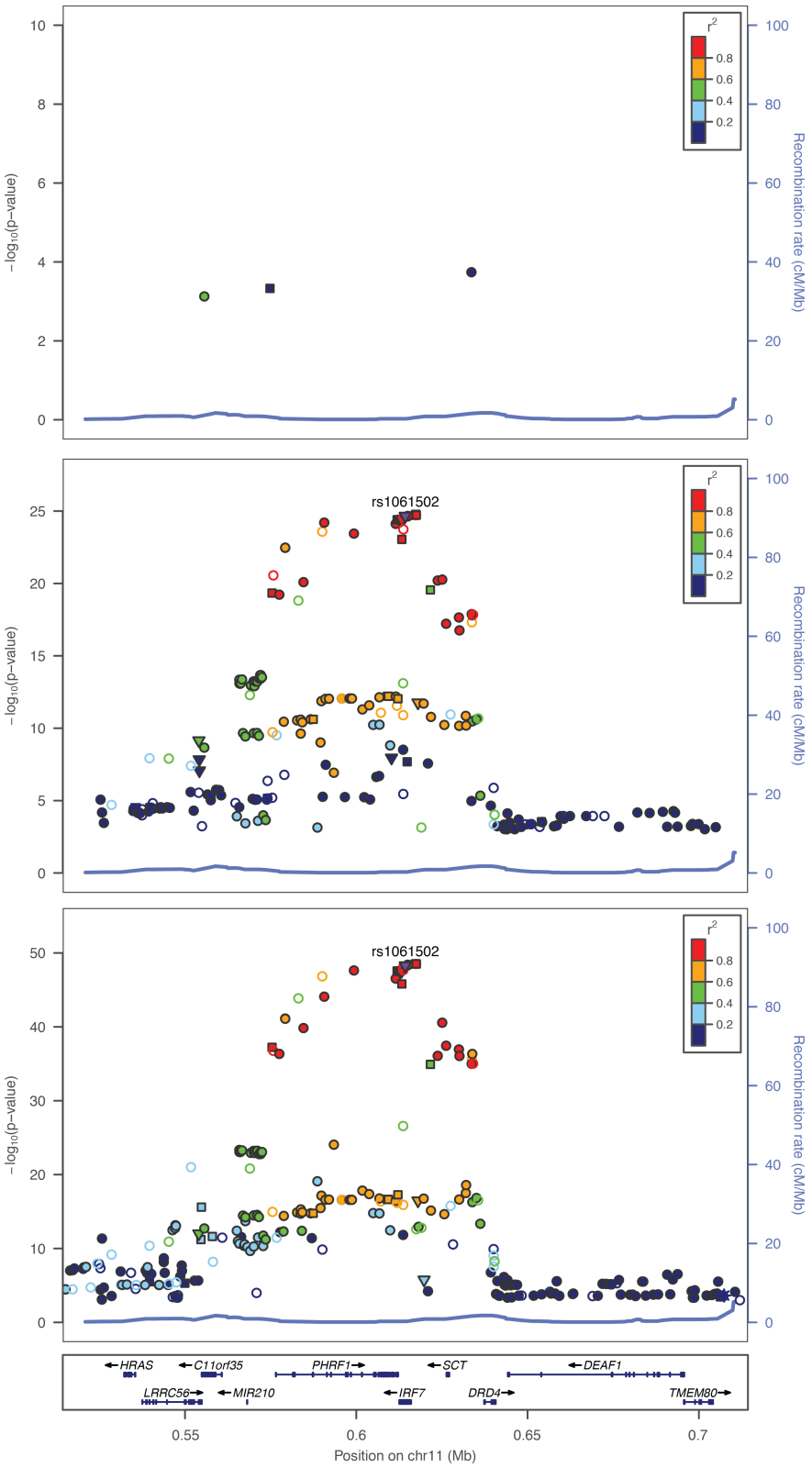
Supp. Fig. 2 – Empirically determined number of PCs to adjust for eQTLs.



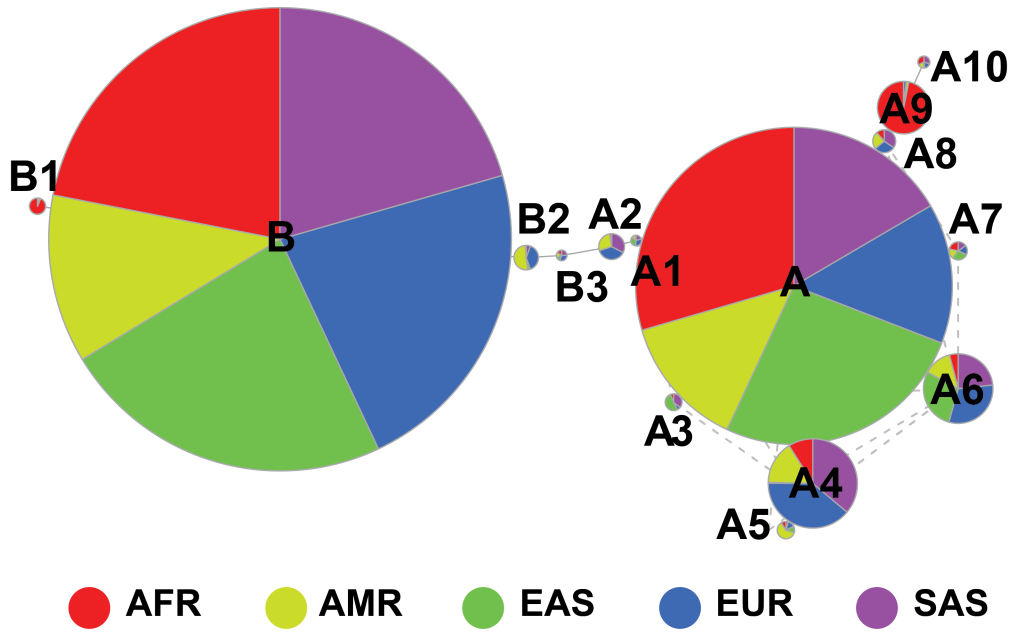
Supp. Fig. 3 – Empirically determined number of PCs to adjust for isoQTLs.



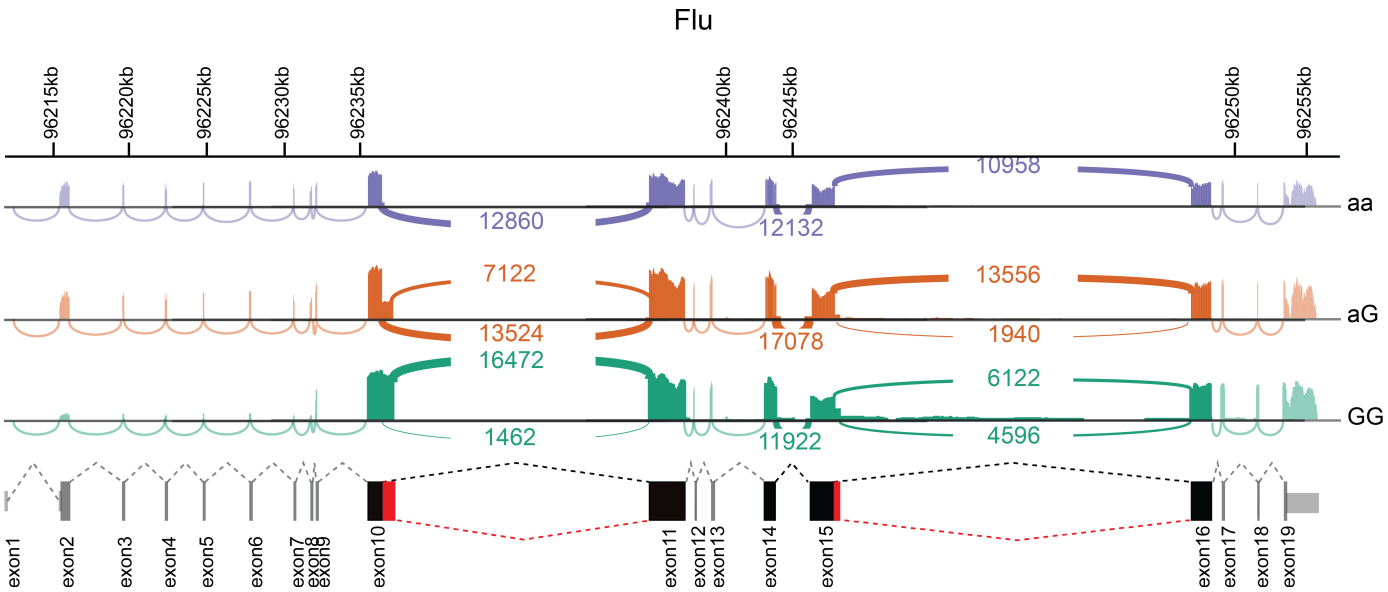
Supp. Fig. 4 – Enrichment of eQTLs and isoQTLs for known transcription factor binding sites.



Supp. Fig. 5 – LocusZoom plot for *IRF7* eQTLs.



Supp. Fig. 6 - Distribution of *ERAP2* haplotypes based on 1000 Genomes.



Supp. Fig. S7 – Sashimi diagram of *ERAP2* expression in response to influenza as a function of patient genotype.

References

1. Shapira, S.D. *et al.* A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell* **139**, 1255-67 (2009).
2. Lee, M.N. *et al.* Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science* **343**, 1246980 (2014).
3. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290-5 (2015).
4. Bray, N.L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-7 (2016).
5. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-78 (2012).
6. Pimentel, H., Bray, N.L., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat Methods* **14**, 687-690 (2017).
7. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).
8. Cribari-Neto, F. & Zeileis, A. Beta Regression in R. *Journal of Statistical Software* **34**, 1-24 (2010).
9. International HapMap, C. The International HapMap Project. *Nature* **426**, 789-96 (2003).
10. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
11. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
12. Browning, B.L. & Browning, S.R. Genotype Imputation with Millions of Reference Samples. *Am J Hum Genet* **98**, 116-26 (2016).
13. Siva, N. 1000 Genomes project. *Nat Biotechnol* **26**, 256 (2008).
14. Shabalin, A.A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-8 (2012).
15. Churchill, G.A. & Doerge, R.W. Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963-71 (1994).
16. JDSwcfAJ, B., A, D. & D, R. qvalue: Q-value estimation for false discovery rate control. (R package version 2.9.0, 2015).
17. Storey, J.D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**, 9440-5 (2003).
18. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
19. Karolchik, D. *et al.* The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32**, D493-6 (2004).
20. Schmidt, E.M. *et al.* GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics* **31**, 2601-6 (2015).
21. Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).