Supporting Information

Katibah et al. 10.1073/pnas.1412842111

SI Methods

Human Cells and Constructs. Cells were grown under standard conditions (37 °C, 5% CO_2) in DMEM supplemented with 10% (vol/vol) tetracycline-free FBS and primocin (Invivogen). The HEK293 Flp-In T-Rex stable cell line expressing 3xF-IFIT5 (1) was treated with 1 μg/mL doxycycline (Sigma) for 48 h, and if indicated, 1,000 U/mL human IFN-β (PBL IFN Source) overnight before extract preparation. Transient transfections used calcium phosphate and pcDNA3.1 3xF-IFIT5 expression vectors (1). The specificity of site-directed mutagenesis was confirmed by sequencing.

Purifications from Cell Extracts. Cells were lysed by freeze-thaw cycles in hypotonic buffer (HLB) containing 20 mM Hepes, pH 7.3, 2 mM $MgCl_2$, 10% (vol/vol) glycerol, 0.2 mM ethylene glycol tetraacetic acid, 1 mM DTT, 0.2% Triton X-100, and protease inhibitors. Extracts were supplemented to 0.2 M NaCl on ice and incubated for 5 min before centrifugation. Soluble cell extracts were normalized with HLB containing 0.2 M NaCl, 0.1% IGEPAL CA 630, and 0.1% Triton X-100 (IP buffer) and then incubated with anti-FLAG M2 resin (Sigma) at 4 °C for 1.5 h. Resin was washed three times for 5 min at room temperature with IP buffer containing 0.2% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate and eluted with FLAG peptide (Sigma) for 45 min at 4 °C in wash buffer. RNA was recovered by extraction with TRIzol (Invitrogen), visualized by denaturing gel electrophoresis followed by staining with SYBR Gold (Invitrogen), and imaged using a Typhoon Trio phosphorimager. For immunoblots, secondary antibodies (IR700 or IR800; Rockland Immunochemicals) were imaged on a LI-COR Odyssey.

Cross-Linking and Cross-Linked RNA Purification. Cells were cross-linked with 0.25% formaldehyde for 10 min in PBS, and then unreacted formaldehyde was neutralized with 0.3 M glycine, pH 7, for 5 min. Cells were resuspended in RIPA buffer containing 50 mM Tris, pH 8, 1% IGEPAL CA 630, 0.5% sodium deoxycholate, 0.05% SDS, 1 mM EDTA, 150 mM NaCl, 1 mM DTT, and protease inhibitors on ice for 10 min and lysed by passage four times through a narrow gauge syringe. Soluble cell extracts were incubated with anti-FLAG M2 magnetic resin (Sigma) at 4 °C for 2 h. Resin was washed three times for 10 min and then two times for 5 min at room temperature in RIPA buffer containing 0.1% SDS, 1 M NaCl, and 2 M urea. Resin was eluted with 3xFLAG peptide in RIPA buffer for 45 min at 4 °C. Protein–RNA cross-links were reversed by adding 100 mM Tris, pH 8, 10 mM EDTA, 1% SDS, and 2 mM DTT to eluted samples and heating to 70 °C for 45 min. RNA was recovered by extraction with TRIzol and then again with phenol:chloroform:isoamyl alcohol [25:24:1 (vol/vol)] followed by ethanol precipitation.

Bacterial Protein Expression and Purification. *Escherichia coli* expression of IFIT5 with a composite N-terminal six-histidine and triple-FLAG tag was done using pET28a(+) (Novagen). Proteins were expressed in BL21(DE3)-RP (Agilent) in 2× yeast extract tryptone grown for 18–20 h at 18 °C with 0.5 mM isopropyl β-D-1-thiogalactopyranoside and purified by binding to Ni-nitrilotriacetic acid agarose (Qiagen) in 20 mM Hepes, pH 7.6, 10% (vol/vol) glycerol, 1 mM tris(2-carboxyethyl)phosphine, 1.5 M NaCl, 20 mM imidazole, and EDTA-free complete protease inhibitor mixture (Roche). After extensive washing with binding buffer, IFIT5 was eluted in binding buffer with 300 mM imidazole and 150 mM NaCl and then buffer-exchanged with an Amicon 30K centrifuge filter to remove imidazole. Protein concentrations were normalized by Bradford assay. Aliquots of purified proteins were snap-frozen in liquid nitrogen and stored at −80 °C.

Recombinant RNAs. RNA was gel-purified after transcription using T7 RNA polymerase. A triphosphate 5′ end was created by T7 RNA polymerase initiation with a Φ2.5 promoter capable of initiation with adenosine (2). Radioactively labeled 5′-ppp West Nile Virus (WNV)30 RNA was produced by T7 RNA polymerase transcription in reactions with ^{32}P-GTP and ^{32}P-UTP. A monophosphate 5′ end was added to tRNA iMet 64-nt RNA generated from a self-cleaving 5′ hammerhead ribozyme precursor (1) using [^{32}P]ATP labeling with T4 polynucleotide kinase (New England Biolabs). A monophosphate 5′ end was added to WNV30 RNA by treatment of the T7 RNA polymerase transcript with calf intestinal phosphatase (New England Biolabs) before phenol-chloroform extraction, ethanol precipitation, and labeling with [^{32}P]ATP and T4 polynucleotide kinase. Capped RNAs were prepared using the Vaccinia RNA capping system (New England Biolabs), with α-phosphate–labeled ^{32}P-GTP. The 2′-*O*-methylated cap1 was prepared by adding Vaccinia 2′-*O* methyltransferase VP39 (New England Biolabs) to the capping reaction. 5′-OH WNV30 RNA was generated from a self-cleaving 5′ hammerhead ribozyme precursor transcribed by T7 RNA polymerase with ^{32}P-GTP and ^{32}P-UTP. RNA concentration was determined by absorbance, with purified RNA integrity and concentration verified by denaturing gel electrophoresis and SYBR Gold staining or autoradiography. RNA 5′-end structures were tested by using terminator (XRN-1), 5′ polyphosphatase in buffer B, and/or tobacco acid pyrophosphatase (TAP) all from Epicenter and resolving the digestion products by denaturing gel electrophoresis. WNV 30 is AGUAGUUCGCCUGUGUGAGCUGACAAACUU.

EMSA. Before use, RNA transcripts were folded by heating to 70 °C for 5 min, freezing on ice, adding 5 mM $MgCl_2$ and 200 mM NaCl, and slow equilibration to room temperature. Binding was performed in 20 mM Tris·HCl, pH 8.0, 200 mM NaCl, 5 mM $MgCl_2$, 10% (vol/vol) glycerol, 2 mM DTT, and 0.1 mg/mL BSA with a radiolabeled RNA concentration of 0.1 nM. Binding reactions were incubated at room temperature for 10 min before electrophoresis on 0.5× TBE (45 mM Tris base, 45 mM boric acid, and 1 mM EDTA), 5% (vol/vol) glycerol, and 5% (wt/vol) (75:1) acrylamide:bis-acrylamide gels at 4 °C for 2 h in 0.5× TBE running buffer with 5 mM $MgCl_2$. Gels were dried, exposed to a phosphor screen (Fuji), and imaged on a Typhoon Trio. Quantification of gels was performed with Image Quant. Binding affinity calculations used Graph Pad Prism 6. RNA binding affinity measurements were highly reproducible in independent assays (data not shown).

Thermostable Group II Intron Reverse Transcriptase Sequencing. Template-switching reactions were done essentially as described (3). The initial template primer substrate consists of a 41-nt RNA oligonucleotide (5′-AGA UCG GAA GAG CAC ACG UCU AGU UCU ACA GUC CGA CGA UC/3SpC3/-3′) with Illumina Read1 and Read2 primer binding sites and a 3′ blocking group (three carbon spacer; IDT) annealed to a complementary ^{32}P-labeled 42-nt DNA primer to leave an equimolar mixture of A, C, G, or T single-nucleotide 3′ overhangs (3). Reactions contained ∼35 ng IFIT5-bound RNA or 25 ng synthetic miRNA control (miRNAx), 100 nM of the initial template-primer substrate, and 2 μM (10 U/μL) TeI4c-MRF RT with a C-terminal truncation of the DNA endonuclease

domain (TGIRT) in 20 μL with final concentrations of 450 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris·HCl, pH 7.5, 1 mM DTT, and 1 mM dNTPs (an equimolar mix of 1 mM dATP, dCTP, dGTP, and dTTP). Components other than dNTPs were incubated at room temperature for 30 min. Reactions were initiated by adding dNTPs, incubated at 60 °C for 10 min, terminated by adding 5 M NaOH to a final concentration of 0.25 M and incubating at 95 °C for 3 min, and then neutralized with 5 M HCl. The labeled cDNAs were analyzed by electrophoresis in a denaturing 6% (wt/vol) polyacrylamide gel, which was scanned with a Typhoon FLA9500 phosphorImager. cDNAs were electroeluted from gel slices using a D-tube Dialyzer Maxi with an molecular weight cut-off of 6–8 kDa (EMD Millipore) and ethanol precipitated in the presence of 0.3 M sodium acetate and glycogen carrier (50 μg; Thermo Scientific). The purified cDNAs were then circularized with CircLigase II (Epicentre) using the manufacturer's protocol with an extended incubation time of 5 h at 60 °C, extracted with phenol-chloroform-isoamyl alcohol (25:24:1) (vol/vol), ethanol precipitated, and amplified by PCR with Phusion-HF (New England Biolabs) using Illumina multiplex (5′-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG TTC AGA GTT CTA CAG TCC GAC GAT C-3′) and barcode (5′-CAA GCA GAA GAC GGC ATA CGA GAT BARCODE GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3′) primers. PCR conditions were initial denaturation at 98 °C for 5 s, followed by 15 cycles of 98 °C for 5 s, 60 °C for 10 s, and 72 °C for 10 s.

For comparison of WT protein purifications, amplification for sequencing was done using cDNAs pooled from two gel slices of ∼55–82 and 84–200 nt, excluding only the 83-nt cDNA products from template switches to Read1,2 RNA (Fig. S5*D*). For WT vs. mutant IFIT5, amplification was performed separately for cDNA size pools a = ∼55–82, b = ∼84–150, and c = ∼150–230 nt (Fig. S5*E*). A biological replicate of the WT vs. mutant IFIT5 purifications (Fig. S5*C*) was used for analysis of ∼55–82 and 84–230 nt cDNAs pooled before amplification (Table S1). Pretreatment of IFIT5-bound RNAs with T4 polynucleotide kinase (Epicentre; manufacturer's conditions) to remove 3′ phosphates prior to template switching did not substantially alter RNA-seq profiles for wild-type or mutant IFIT5.

RNA Sequence Analysis. Samples were sequenced on an Illumina MiSeq. We collected 1,000,000 or more 250-nt paired-end reads for all samples summarized in Table 2. For the biological replicate of WT vs. mutant IFIT5 purifications, 100-nt paired-end reads were collected. For all samples, only read 1 was used for analysis. Read quality was assessed using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc). Reads were trimmed and filtered to remove contaminant primer sequences and substandard base calls from the 3′ ends using Fastx-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit) and Ea-utils (http://code.google.com/p/ea-utils), respectively. Adapter filtering parameters in Ea-utils were relaxed to preserve information related to posttranscriptionally added nucleotides at 5′ and 3′ RNA ends, resulting in 2.8–11.5% of the reads retaining adapter sequence. After trimming and filtering, reads were aligned to the human genome reference sequence (Ensembl GRCh37) with Bowtie 2 using local alignment to identify posttranscriptional additions (4, 5). The genome sequence and annotations were obtained from the University of California, Santa Cruz genome database (http://genome.ucsc.edu). The alignments were analyzed for read abundance and posttranscriptional nucleotide additions using SAMtools (6), BEDtools (7), and Picard (http://picard.sourceforge.net). Sequence reads from mitochondrial tRNA and mitochondrial rRNA were predominantly detected in samples not cross-linked before cell lysis, and therefore, read percentages in Table 2 and Table S1 were calculated after removing sequences mapped to the mitochondrial genome. Ensembl annotation categories containing less than 1.5% of total reads in any library were not listed in the tables. Coverage plots were generated using SAMtools and R (www.R-project.org), and read alignments were visualized using the Integrated Genomics Viewer (IGV) (8). tRNA read start-site plots were generated using BEDtools and Excel. Posttranscriptionally added poly-U tails at tRNA 3′ ends, consisting of two or more consecutive Us following the last genome encoded nucleotide, were identified using SAMtools and quantified using a custom computer script. The abundance and length of poly-U tails may be underestimated due to the presence of a low frequency of other nucleotides within the poly-U tracts (Fig. S7).

1. Katibah GE, et al. (2013) tRNA binding, structure, and localization of the human interferon-induced protein IFIT5. *Mol Cell* 49(4):743–750.
2. Coleman TM, Wang G, Huang F (2004) Superior 5′ homogeneity of RNA from ATP-initiated transcription under the T7 phi 2.5 promoter. *Nucleic Acids Res* 32(1):e14.
3. Mohr S, et al. (2013) Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* 19(7):958–970.
4. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.
5. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
6. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
7. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.
8. Robinson JT, et al. (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24–26.

Table S1. Biological replicate sequencing of pooled RNA

RNA class	WT	E33A	E33A/D334A
tRNA	83	63	30
snaR	2.4	2.0	1.3
protein_coding*	5.6	19	20
rRNA*	5.9	9.9	42
snRNA*	0.3	0.8	1.5
snoRNA*	0.3	0.7	1.0
misc_RNA*	0.5	1.6	1.3
tRNA % poly-U tailed	0.4	0.3	0.2

Table values are rounded percentages. RNA classes included were >1.5% of reads in at least one sample analyzed in Table S1 or Table 2.
*Transcript categories from Ensembl GRCh37.

Fig. S1. SYBR Gold staining of IFIT5-bound RNAs. SYBR Gold staining of input and bound RNAs for 3xF-IFIT5 variants expressed in HEK293T cells by transient transfection, parallel to Fig. 1*E*. FLAG antibody immunoblot of an aliquot of the bound sample is shown below the gel for bound RNA as a recovery control.

Fig. S1

Fig. S2. Bacterial expression and purification of recombinant IFIT5 proteins. (*A*) Cylinder representation of RNA-bound IFIT5 [Protein Data Bank (PDB) ID code 4HOT] with substituted side chains indicated in the expanded cut-away view of the RNA binding channel. (*B*) Purified recombinant IFIT5 proteins detected by SDS/PAGE and Coomassie R-250 staining.

Fig. S2

Fig. S3. Effect of IFIT5 mutations on RNA 5′-end discrimination. EMSA assays comparing binding of the indicated purified recombinant IFIT5 to WNV30 RNAs with 5′-p and 5′-ppp (*A*) or cap0, cap1, and 5′-p (*B*).

Fig. S3

Fig. S4. IFIT5 proteins lack substantial binding affinity for 5′ OH RNA. (*A*) Verification of 5′-end structure on the WNV30 RNAs used for EMSAs in *B*, parallel to Fig. 2*B*. The radiolabeled phosphate is indicated by an asterisk (*), except that the 5′-OH and 5′-ppp RNAs contained internal radiolabel from nucleotide incorporation during transcription. Treatments included 5′ polyphosphatase (5′ PPtase), terminator exonuclease (XRN-1), or buffer without enzyme (–). Products were resolved by denaturing PAGE. (*B*) EMSAs for binding of purified recombinant WT IFIT5 and the indicated IFIT5 variants to WNV30 RNAs bearing a 5-OH, 5′-p, or 5′-ppp group.

Fig. S4

Fig. S5. RNA purification and cDNA synthesis for TGIRT-seq. (*A* and *B*) SYBR Gold staining of RNAs in input cell extract and enriched by FLAG antibody purification from a HEK293 cell line with doxycycline-inducible 3xF-IFIT5 expression (1) or the parental cells. In *A*, cells were pretreated with IFN-β as indicated. Immunoblots of cell extract confirmed 3xF-IFIT5 expression, IFN-β treatment (Y701-phosphorylated Stat1), and equal loading (Calnexin). In *B*, cells were treated with formaldehyde for in vivo protein–RNA cross-linking followed by stringently denaturing affinity purification and reversal of the cross-link. (*C*) SYBR Gold staining of RNAs in input cell extract and enriched by FLAG antibody purification from transiently transfected HEK293 cells. Duplicate transfections and purifications were performed to obtain biological replicates for 3xF-IFIT5 WT, E33A, and E33A/D334A. A FLAG antibody immunoblot of an aliquot of the purified samples is the recovery control. RNA was transferred from the denaturing gel for the Northern blot analysis of snaR A-family transcripts shown in the bottom panel, with one major and two minor RNA forms distinguished by migration (minor forms are indicated with an asterisk). (*D* and *E*) Template-switching cDNA synthesis. In D, PCR amplification and sequencing were done using pooled cDNA excised as two gel slices of ∼55–82 and ∼84–200 nt, excluding only the 83-nt cDNA product from template switching to Read1,2 RNA. In the legend above the gel, N indicates native extract purification and XL indicates purification after in vivo cross-linking. In *E*, amplification and sequencing were performed separately for cDNA from size pools a = ∼55–82, b = ∼84–150, and c = ∼150–230 nt. TGIRT-seq from the biological replicate in which cDNAs were processed as in *D* is summarized in Table S1.

Fig. S5

1. Katibah GE, et al. (2013) tRNA binding, structure, and localization of the human interferon-induced protein IFIT5. *Mol Cell* 49(4):743–750.

Fig. S6. Broad representation of IFIT5-bound tRNAs. Profiles of tRNA read abundance were plotted using RNA libraries from in vivo cross-linking or native extract of cells without (–) or with (+) prior IFN-β treatment. (*Left*) The tRNA loci with maximal read abundance are indicated with read number as the *y* axis. (*Right*) All tRNA loci with mapped reads were rank-ordered in normalized abundance. The number of different tRNA species identified by sequence reads is indicated in parenthesis in the plots on the right and represent most of the 625 reference human genome tRNA and tRNA pseudogene loci searched.

Fig. S6

Fig. S7. Read sequence alignments. The figure shows screen shots of IGV sequence alignments for reads mapping to loci for which coverage plots are shown in Fig. 3. The blue bar at the top delineates the mature tRNA sequence encoded in the genome, with the arrow indicating 5′ to 3′ direction of the tRNA, which differs across alignments depending on the DNA strand to which the reads are mapped by the Bowtie 2 aligner. The total number of reads mapped to the locus is indicated near the top of each panel. To fit the entire alignment on one page, loci with more than 1,500 mapped reads were down-sampled to 1,500 reads in IGV. Reads were sorted by their start site on the chromosome, which can be from either the 5′ or 3′ RNA end depending on the orientation of the gene on the chromosome. In the coverage plot profiles, nucleotides matching the genome sequence are represented in gray color, and mismatches are represented in different colors (A, green; C, blue; G, brown; T, red). Soft-clipped sites, which demarcate the beginning of extra 5′ and 3′ nucleotides that do not match the genomic sequence, are indicated by a short black bar, and read continuity between a genome sequence gap, such as a spliced intron, is indicated by a black line. Pol III ter, predicted RNA polymerase III termination site. For the spliced tRNAs, reads were mapped with or without intron removal from the gene sequence to highlight inaccurate splice junctions and modified nucleotides near the junction that affect the sequence alignment. The spliced ArgTCT tRNA reads contain potential examples of missplicing with a shifted splice junction and/or one extra nucleotide inserted at the junction (highlighted in the inset sequence alignment). For the spliced LeuCAA tRNA reads, misalignments of the splice junction are caused by misincorporation relative to genomic sequence by reverse transcription of a modified nucleotide (G37/yW). Additionally, the Bowtie 2 aligner misaligns short truncated 5′ exon sequences after the gap resulting from intron excision. These misalignments are noted as "misaligned 5′ exon" in the alignment of the spliced tRNAs. Misalignment of these short 5′ exon sequences could be ameliorated by adjustment of the alignment parameters, but at the cost of introducing misalignments elsewhere in the tRNA sequence. Examples of untrimmed adapter sequence, nontemplated nucleotide addition by the TGIRT at cDNAs 3′ ends (corresponding to tRNA 5′ ends), and rare second template switches are indicated in the alignments. Mismatches at positions corresponding to modified nucleotides known to be present in the tRNA (1–3) are indicated by arrows indicating the tRNA position and modified nucleotide. The spectrum of misincorporated nucleotides at modification sites is shown in the coverage plot, with a misincorporated nucleotides threshold of 10%. In at least some cases (e.g., m^1A and $m^{2,2}G$), the spectrum of mismatches appears to be characteristic of the modified base and may be useful for identifying unknown base modifications in other coding and noncoding RNAs. A position of potential posttranscriptional modification of a conserved guanosine residue in the snaR is indicated in the alignment. Cm, 2′-*O*-methylcytidine; D, dihydrouridine; I, inosine; i^6A, N^6-isopentenyl adenosine; t^6A, N^6-threonylcarbamoyladenosine; m^1A, 1-methyladenosine; m^1G, 1-methylguanosine; m^1I, 1-methylinosine; $m^{2,2}G$, N^2,N^2-dimethylguanosine; m^3C, 3-methylcytidine; yW, wybutosine.

Fig. S7

1. Björk GR, et al. (1987) Transfer RNA modification. *Annu Rev Biochem* 56:263–287.
2. Phizicky EM, Hopper AK (2010) tRNA biology charges to the front. *Genes Dev* 24(17):1832–1860.
3. Beier H, Lee MC, Sekiya T, Kuchino Y, Nishimura S (1992) Two nucleotides next to the anticodon of cytoplasmic rat tRNA(^{Asp}) are likely generated by RNA editing. *Nucleic Acids Res* 20(11):2679–2683.

Fig. S8. IFIT5-bound tRNA pseudogene transcripts. Pseudo-tRNA read count and normalized count were rank-ordered in abundance. Total number of reads in each sample: cross-linked = 2,935,906; native no IFN-β = 2,647,164; and native with IFN-β = 3,175,669.

Fig. S8

Fig. S9. Examples of tRNA loci read alignments showing atypically long 5′ or 3′ extensions from the positions of mature tRNA ends or reads ending at a modified tRNA nucleotide position. Screen shots of IGV alignments are labeled as described in Fig. S7. The CysGCA reads begin from the tRNA 3′ CCA and end at the position of a modified nucleotide. The ThrCGT read alignment provides an example of a tRNA locus with two potential Pol III termination sites.

Fig. S9

Fig. S10. Lack of 5′ bias for read distribution across mRNAs. Reads from WT and mutant IFIT5 cDNA size category c (Fig. S5) were mapped to the transgene mRNA expressing the coding region of 3xF-IFIT5 flanked by untranslated regions from pcDNA3.1 (*A*) or to the endogenous histone H4 mRNA from locus HIST2H4A (*B*). Less coverage of these mRNAs was attained by mapping reads from the smaller cDNA size categories. Each *x* axis begins at the first nucleotide of the mRNA and ends at the polyadenylation site. The *y* axis scores the count of *x* axis read 5′ positions.

Fig. S10