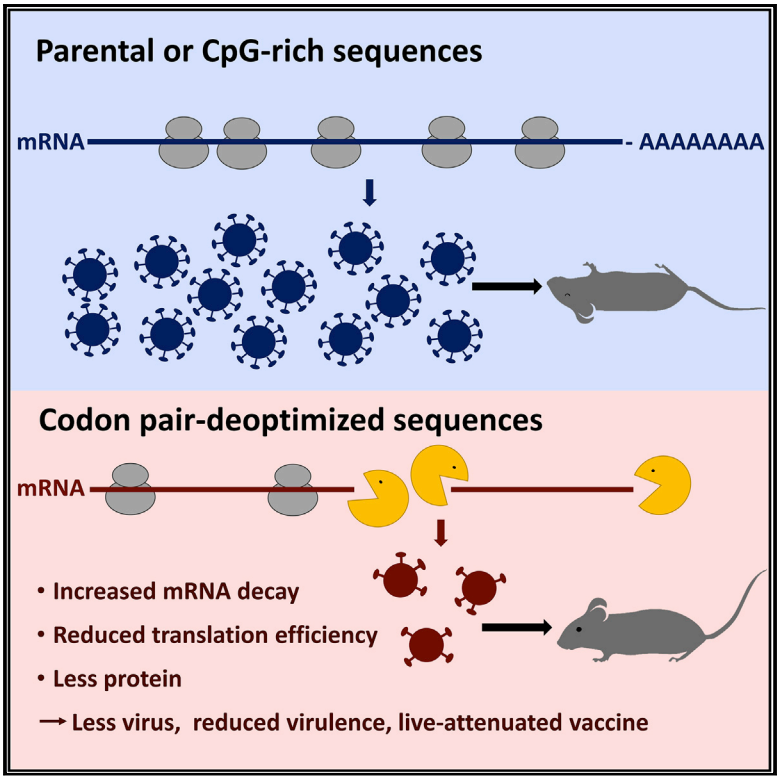
Cell Reports

Article

# Mechanism of Virus Attenuation by Codon Pair Deoptimization

## Graphical Abstract




## Authors

Nicole Groenke, Jakob Trimpert, Sophie Merz, ..., Markus Landthaler, Nikolaus Osterrieder, Dusan Kunec

## Correspondence

no.34@fu-berlin.de (N.O.), dusan.kunec@fu-berlin.de (D.K.)

## In Brief

Codon pair deoptimization is a highly efficient virus attenuation strategy. It involves the recoding of viral genomes using underrepresented codon pairs. Groenke et al. show that suboptimal codon pairs are the primary cause of attenuation because they reduce mRNA stability, translation efficiency, and thus also protein production of recoded genes.

## Highlights

- Attenuation by codon pair deoptimization is determined by suboptimal codon pairs
- Suboptimal codon pairs reduce mRNA stability and throttle or abrogate translation
- CpG dinucleotides are dispensable for attenuation by codon pair deoptimization
- Codon pairs are important determinants of mRNA stability

Groenke et al., 2020, Cell Reports *31*, 107586
April 28, 2020 
https://doi.org/10.1016/j.celrep.2020.107586

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

# Mechanism of Virus Attenuation by Codon Pair Deoptimization

Nicole Groenke,[1] Jakob Trimpert,[1] Sophie Merz,[2] Andelé M. Conradie,[1] Emanuel Wyler,[3] Hongwei Zhang,[4] Orsalia-Georgia Hazapis,[3] Sebastian Rausch,[4] Markus Landthaler,[3,5] Nikolaus Osterrieder,[1,*] and Dusan Kunec[1,6,*]

[1]Institut für Virologie, Freie Universität Berlin, Robert-von-Ostertag-Str. 7-13, 14163 Berlin, Germany
[2]Institut für Veterinärpathologie, Freie Universität Berlin, Robert-von-Ostertag-Str. 15, 14163 Berlin, Germany
[3]Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Hannoversche Str. 28, 10115 Berlin, Germany
[4]Institut für Immunologie, Freie Universität Berlin, Robert-von-Ostertag-Str. 7-13, 14163 Berlin, Germany
[5]IRI Life Sciences, Institute of Biology, Humboldt-Universität zu Berlin, Philippstraße 13, 10115 Berlin, Germany
[6]Lead Contact
*Correspondence: no.34@fu-berlin.de (N.O.), dusan.kunec@fu-berlin.de (D.K.)
https://doi.org/10.1016/j.celrep.2020.107586

## SUMMARY

Codon pair deoptimization is an efficient virus attenuation strategy, but the mechanism that leads to attenuation is unknown. The strategy involves synthetic recoding of viral genomes that alters the positions of synonymous codons, thereby increasing the number of suboptimal codon pairs and CpG dinucleotides in recoded genomes. Here we identify the molecular mechanism of codon pair deoptimization-based attenuation by studying recoded influenza A viruses. We show that suboptimal codon pairs cause attenuation, whereas the increase of CpG dinucleotides has no effect. Furthermore, we show that suboptimal codon pairs reduce both mRNA stability and translation efficiency of codon pair-deoptimized genes. Consequently, reduced protein production directly causes virus attenuation. Our study provides evidence that suboptimal codon pairs are major determinants of mRNA stability. Additionally, it demonstrates that codon pair bias can be used to increase mRNA stability and protein production of synthetic genes in many areas of biotechnology.

## INTRODUCTION

Recent advances in the *de novo* synthesis of DNA enable large-scale recoding of viral genomes (Coleman et al., 2008; Mueller et al., 2010). Viruses with recoded genomes can be used as modified live attenuated vaccines. The goal of the recoding is to modify dinucleotide, codon, or codon pair composition of the recoded viral genes, because all three types of potentially interrelated modifications might produce replication-competent, but attenuated viruses. Importantly, while recoding introduces hundreds of point mutations into viral genomes, the amino acid composition of the encoded proteins remains preserved. Consequently, the recoded viruses are antigenically identical to their pathogenic parents (Broadbent et al., 2016; Coleman et al., 2008; Mueller et al., 2010). The antigenic identity and replicative potential enable attenuated viruses to induce immune responses that are similar to those of virulent strains.

Most of the amino acids are encoded by several codons. Organisms use synonymous codons with varying frequencies—a phenomenon known as codon bias. In addition to codon bias, protein coding sequences also show codon pair bias, which is bias of two adjacent codons (Gutman and Hatfield, 1989; Kunec and Osterrieder, 2016). Some codon pairs are found in protein coding genes significantly more or less frequently than would be expected based on the overall frequencies of two codons that form a particular codon pair (Coleman et al., 2008; Gutman and Hatfield, 1989; Mueller et al., 2006).

Codon pair deoptimization (CPD) is a highly efficient virus attenuation strategy that utilizes suboptimal codon pairs to achieve attenuation of recoded viruses. CPD rearranges the positions of existing synonymous codons in a viral gene, without changing the codon bias or amino acid composition of the encoded protein (Coleman et al., 2008; Mueller et al., 2010). The goal of the codon swapping is to increase the frequency of codon pairs that are statistically underrepresented (suboptimal) in the host genes. CPD has been widely used for the development of viral vaccines, because it allows the generation of efficacious, non-reverting attenuated viruses with the desired degree of attenuation (Coleman et al., 2008; Mueller et al., 2010). It is broadly applicable for rapid and efficient attenuation of a wide range of viruses including poliovirus (Coleman et al., 2008), influenza A virus (IAV) (Broadbent et al., 2016; Mueller et al., 2010; Yang et al., 2013), human immunodeficiency virus (Martrus et al., 2013), and Dengue virus (Shen et al., 2015).

Despite its great potential, the molecular mechanisms that underlie the attenuation by CPD have remained unknown (Broadbent et al., 2016; Coleman et al., 2008; Futcher et al., 2015; Le Nouën et al., 2014; Shen et al., 2015; Simmonds et al., 2015;

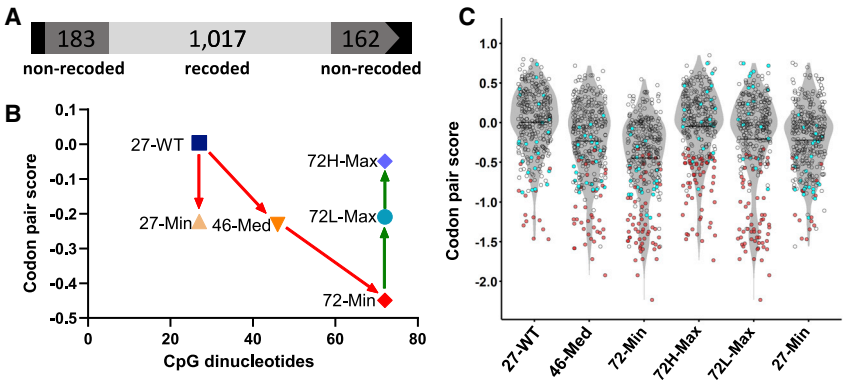



**Figure 1. *In Silico* Design and Properties of Recoded IAV NA Genes**

(A) Structure of the IAV NA segment and recoding design. Only the central part of the NA gene was recoded (light gray), because terminal non-coding (black) and coding regions (gray) contain vRNA packaging signals. Sequence length is in nucleotides.

(B) Average codon pair score (CPS) and number of CpG dinucleotides in recoded sequences of NA genes. Red and green arrows represent CPD and codon pair optimization, respectively.

(C) Distribution of CPS in recoded sequences of NA genes. Each circle represents the CPS of a single codon pair. Empty circles represent codon pairs that do not contain CpG dinucleotides. Blue filled circles represent codon pairs that contain CpG dinucleotides within the codons (CGN, or NCG). Red filled circles represent codon pairs of the NNC-GNN type. Note that all codon pairs of the NNC-GNN type have negative CPS. Because each codon forms two different codon pairs, there are approximately twice as many codon pairs (24) as there are codons with a CpG dinucleotide (12) in the recoded NA sequence. The gray density plots show the shape of the CPS distribution. The black lines indicate average CPS of the recoded NA gene sequence.

Yang et al., 2013). While different mechanisms have been proposed, the prevailing hypothesis is that the large number of underrepresented codon pairs interferes with efficient protein production or processing (Coleman et al., 2008; Shen et al., 2015). It is speculated that physical properties of specific tRNAs, such as their 3D-structures, hamper their efficient fit into adjacent aminoacyl- and peptidyl-sites in the translating ribosome (Coleman et al., 2008). Consequently, codon pair-deoptimized sequences may be prone to mistranslation, stalled translation, or premature termination (Broadbent et al., 2016). It was discovered, however, that CPD of vertebrate viruses not only affects codon pair bias, but also dramatically increases the number of CpG and, to a lesser degree, TpA (UpA) dinucleotides in recoded sequences (Kunec and Osterrieder, 2016; Tulloch et al., 2014). The increase of CpG dinucleotides through recoding is inadvertent and almost inevitable, as codon pairs that contain CpG dinucleotides at the codon pair boundary (NNC-p-GNN) are among the most underrepresented in vertebrates (Atkinson et al., 2014; Kunec and Osterrieder, 2016). Interestingly, most vertebrate RNA and small DNA viruses also show suppression of CpG dinucleotides in their genomes. Thus, an alternate hypothesis proposes that CpG dinucleotides are the sole cause of attenuation because they act as triggers of antiviral immune responses (Greenbaum et al., 2009). Recent studies showed that the host zinc-finger antiviral protein (ZAP) is a long-suspected single-stranded RNA (ssRNA) CpG receptor because it binds specifically to CpG-rich RNAs and targets them for degradation by the RNA exosome (Ficarelli et al., 2019; Takata et al., 2017). In addition, it was shown that CPD can reduce stability of recoded mRNA (Broadbent et al., 2016; Yang et al., 2013) and induce temperature-sensitivity in mutant viruses (Le Nouën et al., 2014; Yang et al., 2013).

As codon pair preferences and dinucleotide frequencies are tightly linked, disentangling the effects of two phenomena is exquisitely difficult (Kunec and Osterrieder, 2016; Tulloch et al., 2014). We here dissected the contribution of underrepresented codon pairs from that of CpG dinucleotides on attenuation of recoded viruses. To separate the effect of the two features, we designed and characterized a series of IAV mutants in which the two parameters were altered independently. We show that underrepresented codon pairs were solely responsible for attenuation of recoded viruses, whereas CpG dinucleotide frequencies had a negligible effect. We show that decreased mRNA stability and reduced translation efficiency underlie the attenuation by CPD. Our work refutes the hypothesis that the effect of attenuation can be explained solely by the increased number of CpG dinucleotides in recoded sequences. In addition, several recent studies provided evidence that codon optimality is a major determinant of mRNA stability. Here we demonstrate that the identity of codon pairs is another critical determinant of mRNA stability, which profoundly affects protein output and, ultimately, fitness of recoded viruses.

## RESULTS

### Design of Recoded Viruses

To separate the effects of CpG dinucleotides from the codon pair bias on viral attenuation, we designed a series of IAV mutants carrying recoded variants of the neuraminidase (NA) gene. As an experimental model we used the mouse-adapted IAV strain A/WSN/1933 (H1N1), which causes acute respiratory disease in immunocompetent mice (Stuart-Harris, 1939). NA is a virion envelope glycoprotein that facilitates the release of virus particles from the surface of infected host cells and their spread in the respiratory tract. We chose the NA gene because it is amenable to recoding and does not have alternative reading frames (Mueller et al., 2010; Yang et al., 2013). All genomic RNA segments of IAV contain *cis*-acting signals at their termini, which are essential for the packaging of viral RNA (vRNA) segments into virions (Eisfeld et al., 2015). To ensure efficient packaging of recoded NA segments, we have altered only 339 codons (1,017 nt) from the central part of the gene and left the first 61 (183 nt) and the last 54 codons (162 nt) of the NA opening reading frame (ORF) unchanged (Figure 1A).

The degeneracy of the genetic code allows encoding a protein by a myriad of synonymous RNA or DNA sequences. The targeted part of the NA gene could be encoded by $10^{218}$ different

**Table 1. Properties of Recoded IAV H1N1 Strain A/WSN/1933 NA Genes**

| Name | Description | CpG | UpA | CPS | CAI | Codon Identity (%) | Nucleotide Identity (%) |
|---|---|---|---|---|---|---|---|
| 27-WT | original CPS | 27 | 54 | 0.00 | 0.74 | 100.0 | 100.0 |
| 72-Min | minimal CPS | 72 | 100 | −0.45 | 0.74 | 37.8 | 75.6 |
| 46-Med | reduced CPS | 46 | 77 | −0.23 | 0.74 | 61.4 | 83.6 |
| 72H-Max | same number of NNC-GNN as 72-Min, maximal CPS | 72 | 57 | −0.05 | 0.74 | 39.2 | 76.3 |
| 72L-Max | same NNC-GNN as 72-Min, maximal CPS | 72 | 71 | −0.21 | 0.74 | 38.9 | 76.4 |
| 27-Min | same NNC-GNN as 27-WT, minimal CPS | 27 | 82 | −0.22 | 0.74 | 52.5 | 80.2 |

CpG, number of CpG dinucleotides in recoded sequences; UpA, number of UpA dinucleotides in recoded sequences; CPS, average codon pair score; CAI, codon adaptation index (a measure of codon bias); Codon identity, percentage of codons that occupy the same position in the parental and recoded genes; Nucleotide identity, percentage of nucleotides that occupy the same position in the parental and recoded genes.

sequences. We designed five different NA variants by reshuffling the naturally used codons. Therefore, all variants had the same codon bias as the parental sequence. In contrast, the number of CpG dinucleotides at the codon boundaries and the nature of codon pairs varied considerably among the recoded variants (Table 1). Codon pair bias can be quantified using the codon pair score (CPS) statistics (Coleman et al., 2008), where statistically underrepresented codon pairs have negative and overrepresented codon pairs have positive CPS values. The average CPS of a gene is calculated as the arithmetic mean of individual CPS values, which is shown alongside CpG dinucleotide frequencies in Figure 1B. The first part of our naming scheme indicates the number of CpG nucleotides that are present in the recoded part of the NA gene. The second part indicates whether the recoding lead to an increased (Max) or decreased (Med or Min) average CPS of the recoded sequence, given the constraints (Figure 1B).

The NA gene variants were produced by three different recoding strategies. The first set was produced by "regular" CPD (46-Med and 72-Min) of the parental NA sequence (27-WT), without exerting any constraint on the number of CpG dinucleotides in recoded sequences. Because CPD inadvertently increases the number of CpG dinucleotides, both 46-Med and 72-Min have more CpG dinucleotides than 27-WT. The 72-Min variant has the lowest attainable CPS, while 46-Med is moderately ("medium") codon pair deoptimized (Table 1). The number of CpG dinucleotides correlates linearly with the CPS (Figure 1B).

In the second recoding strategy, an NA gene variant was designed to contain the same number of CpG dinucleotides as 27-WT but a lower CPS. The 27-Min was derived from the 27-WT by CPD, but codon pairs of the NNC-GNN type were omitted from the recoding. The remaining synonymous codons of the 27-WT sequence were reshuffled to create as many underrepresented codon pairs as possible, but generation of additional codon pairs of the NNC-GNN type was disallowed (Figures 1B and 1C; Table 1). As a consequence, 27-WT and 27-Min have the same number of CpG dinucleotides (27), but 27-Min has the minimal attainable CPS (Table 1). Note that it is impossible to attain a lower CPS than that of the 27-Min (−0.22) without increasing the number of CpG dinucleotides in the recoded sequence.

In the third set, the 72L-Max and 72H-Max mutants were designed to contain 72 CpG dinucleotides, exactly as many as the 72-Min, which has the lowest attainable CPS and a large number of CpG dinucleotides, whereas the average CPS of 72L-Max and 72H-Max (L stands for low, H stands for high) is increased (Figure 1B). The 72L-Max was designed to have the exact same codon pairs of the NNC-GNN type as 72-Min (60 NNC-GNN codon pairs were omitted from the recoding) (Figure 1C). In contrast, codon pairs of the NNC-GNN type in the 72H-Max gene were designed to have the highest possible CPS (Figure 1C). As a result, both gene variants contain completely different codon pairs of the NNC-GNN type. NNC-GNN codon pairs have the lowest possible CPS values in the 72L-Min, and the highest possible CPS values in the 72H-Max variant. The remaining codon pairs in 72L-Max and 72H-Max were chosen to have the highest possible CPS values.

Codon pair preferences and dinucleotide frequencies are intimately related because the most underrepresented codon pairs contain CpG dinucleotides at the codon pair boundary. Hence, uncoupling the effects of the two phenomena is difficult (Broadbent et al., 2016; Futcher et al., 2015; Simmonds et al., 2015). Previous studies attempted to separate the effects of the two sequence features on attenuation of viruses. The separation remained incomplete, however, because the analyzed viruses not only differed in codon pair bias and CpG dinucleotide frequencies but also in codon bias (Futcher et al., 2015; Tulloch et al., 2014). The unique design of our recoded NA gene variants enabled us to compare biological properties of mutant viruses that either had the same number of CpG dinucleotides, or nearly identical CPS (Figure 1B), but all contained exactly the same codons. This in turn allowed us to cleanly separate the effects of underrepresented codons, codon pairs, and CpG dinucleotides on gene expression, and replication fitness, of recoded viruses.

## CPD Viruses, but Not CpG-High Viruses, Are Strongly Attenuated in Tissue Culture

The recoded NA segments were produced synthetically and infectious IAV mutants were recovered using a reverse genetics system of A/WSN/1933 (H1N1) (Hoffmann et al., 2000). The replication kinetics of all mutant viruses were compared to parental 27-WT virus in multi-step growth analyses in Madin-Darby canine kidney (MDCK) epithelial (Figure 2A) and human basal alveolar epithelial A549 cells (Figure 2B). In MDCK cells, only the codon pair-deoptimized viruses 46-Med and 72-Min replicated less efficiently than the 27-WT virus, and their viral titers

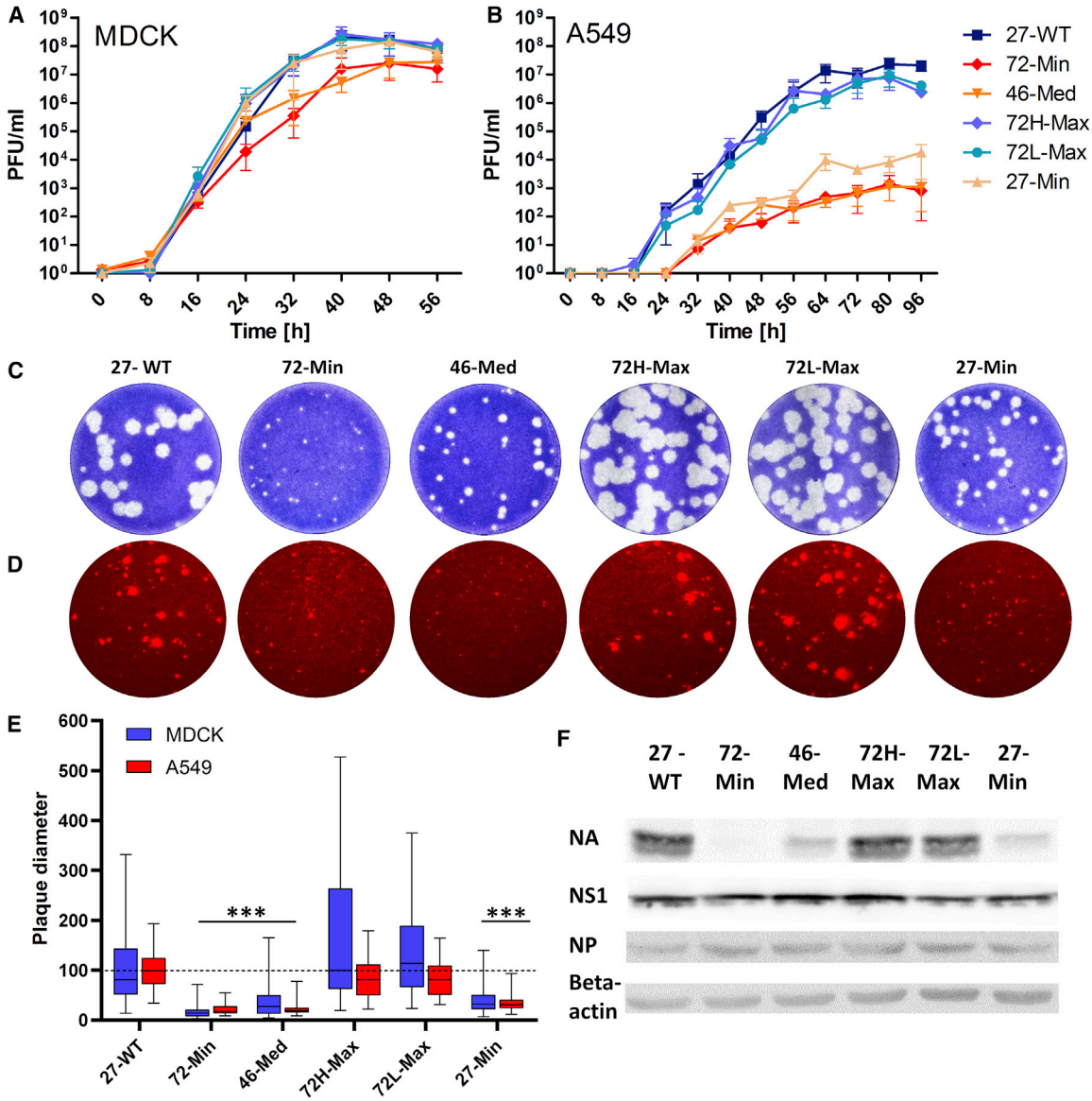



**Figure 2. Biological Properties of Recoded Viruses in Tissue Culture**

(A and B) Multiple-step growth kinetics of parental 27-WT and recoded viruses in (A) MDCK and (B) A549 cells. Data are shown the means of three biological replicates. Error bars show SD.

(C and D) Representative images of plaques formed by 27-WT and recoded viruses on MDCK (C) and A549 cells (D).

(E) Relative plaque diameters of 27-WT and recoded viruses on MDCK and A549 monolayers three days post infection. p values were calculated using an ANOVA Bonferroni's multiple comparison test (***$p < 0.001$).

(F) Western blot analysis of NA, NP, and NS1 protein production in MDCK cells infected with the parental and recoded viruses 5 h after infection. Beta-actin was used as loading control. The images are representative of three independent biological experiments. Original, uncropped images from western blots are shown in Figure S5.

were typically ~10 times lower than those of the 27-WT virus. In A549 cells, the growth kinetics of the viruses 72L-Max and 72H-Max were indistinguishable from those of the 27-WT virus. In contrast, three codon pair-deoptimized viruses 27-Min, 46-Med, and 72-Min replicated with severely reduced kinetics when compared to the 27-WT virus, and we recorded an ~1,000–10,000 times lower output of infectious virus particles by the codon pair-deoptimized viruses (Figure 2B).

The growth of the recoded viruses was also investigated by analyzing plaque formation on MDCK and A549 monolayers (Figures 2C–2E). Three codon pair-deoptimized viruses 27-Min, 46-Med, and 72-Min formed plaques that were significantly smaller than those of the parental 27-WT virus (Figures 2C–2E). The diameter of the plaques was 3–5 times smaller than those of the 27-WT virus on MDCK and A549 cells. On the other hand, the viruses with a high number of CpG dinucleotides, 72L-Max

and 72H-Max, produced plaques that were indistinguishable from those of the 27-WT virus (Figures 2C–2E). In general, sizes of virus plaques on MDCK and A549 cells correlated well with the ability of the viruses to replicate in A549 cells (Figure 2B).

We hypothesized that the replication of mutant viruses in MDCK and A549 cells might be a function of NA expression. Western blot analyses revealed that viruses that form smaller plaques than the parental 27-WT virus also produced less NA protein in MDCK cells at 5-h post infection, indicating that reduced levels of NA are most likely responsible for attenuation of viral growth in cultured cells (Figure 2F). The codon pair-deoptimized viruses 27-Min, 46-Med, and 72-Min produced less NA protein than the parental virus. NA levels were the lowest and almost undetectable in cells infected with the 72-Min virus. In contrast, NA production was unaffected in cells infected with 72L-Max and 72H-Max viruses. As expected, all viruses produced comparable amounts of nonstructural protein 1 (NS1) and nucleoprotein (NP), as these genes had not been recoded (Figure 2F).

### CPD Viruses Are Significantly Attenuated *In Vivo*

The studied viruses can be classified into two groups based on their characteristics in tissue culture: the attenuated (27-Min, 46-Med, and 72-Min) and the non-attenuated (27-WT, 72L-Max, and 72H-Max). Next, we examined their pathogenicity in a mouse model of infection. Seven groups of nine 6-week-old BALB/c mice were intranasally inoculated with $5 \times 10^4$ plaque-forming units (PFUs) of the respective viruses. Animal health and body weight were continuously monitored for 18 days (Figure 3A; Figure S1). Three mice from each group were euthanized at 2, 5, and 18 days post infection (dpi). As expected, the parental 27-WT virus caused severe clinical disease in the infected mice. Similarly, almost all mice infected with the mutant 72H-Max virus, and most mice infected with the mutant 72L-Max virus, developed severe pneumonia within the first few days after infection. In contrast, mice that were infected with codon pair-deoptimized viruses and mice of the mock-infected control group showed no apparent clinical signs of respiratory disease (Figure 3A; Figure S1).

To obtain a more accurate picture of the virulence of the engineered viruses, we determined the lung virus titers by plaque assays and qPCR, and examined pathological changes in the lungs of infected animals. The parental 27-WT, 72L-Max, and 72H-Max viruses replicated very efficiently and to comparable levels in the lungs of infected mice. On the other hand, the number of infectious viruses (Figure 3B) as well as NP (Figure 3C) and NA RNA copy numbers (Figure S2A) were 10–1,000 times lower in lungs of mice infected with the codon pair-deoptimized viruses 27-Min, 46-Med, and 72-Min when compared to the 27-WT control group. The 72-Min virus showed strongest attenuation and no infectious virus could be detected in lungs of infected animals at 5 dpi.

No infectious virus could be detected in any of the mice at 18 dpi. To evaluate the immune response to infection, we quantified the level of antibodies to IAV present in sera by microneutralization and hemagglutination inhibition (HAI) assays. Mice infected with the codon pair-deoptimized viruses 27-Min, 46-Med, and 72-Min and the single surviving mouse infected with 72L-Max showed high antibody titers (Figures S2B and S2C), indicating that these viruses induced robust immune responses.

We also quantified the degree of necrosis in bronchi, bronchioli, and alveoli (Figures S2D and S2E, 3D, and S3, respectively), inflammatory infiltration of leukocytes into interstitial and perivascular spaces (Figures 3E and 3F; Figure S3), and occurrence of the bronchus-associated lymphoid tissue (BALT) (Figures S2F and S3). Mice infected with the 27-WT, 72L-Max, and 72H-Max viruses developed similar lung histopathologies. Necrosis was most prominent in bronchioli but was also present in the bronchi and alveoli. In contrast, mice infected with the 27-Min and 46-Med viruses developed only mild pneumonia with limited necrosis and inflammation. The 72-Min virus was virtually apathogenic and lungs of mice infected with the 72-Min virus resembled those of the control group, since no histopathological signs of lung disease were present. Nevertheless, robust antiviral T cell responses were induced in all infected groups, as evidenced by in the production of high IFN-$\gamma$ levels by spleen cells in response to re-stimulation with the 27-WT virus 18 days after infection (Figure 3G).

Histological examination of the diseased mice confirmed that mice displayed moderate to severe, multifocal to coalescing, and acute necrotizing bronchointerstitial pneumonia. Immunohistochemical staining of lung tissues confirmed that IAV was the cause of pneumonia, because IAV antigen was widely distributed in the lungs of severely diseased mice. Virus antigen was detected in intact and necrotic bronchiolar and bronchial epithelial cells, within alveolar macrophages, and in necrotic cellular debris in airway lumina (Figure 3H).

### CPD Virus Vaccine Candidates Protect Mice against Challenge Infection

We next conducted a vaccination experiment to test whether viruses that showed no overt disease symptoms in the first experiment (27-Min, 46-Med, and 72-Min) could induce protective immunity against lethal challenge with the parental 27-WT virus. Eight 6-week-old mice were immunized by intranasal inoculation with $5 \times 10^4$ PFUs of the recoded viruses and body weight was monitored (Figure S4A). After 28 days, the mice were challenged with $5 \times 10^5$ PFUs of 27-WT virus. All attenuated viruses induced complete protection against the challenge with the 27-WT virus. Immunized mice did not show any body weight loss upon challenge (Figure 4A), and virus was cleared 10 days after challenge as no RNA copies could be detected. All three attenuated viruses induced strong adaptive immune responses, detected as high anti-IAV antibody titers (Figure 4B) and prominent virus-specific interferon-gamma (IFN-$\gamma$) responses in immunized mice (Figure S4B). There was no detectable necrosis (Figure 4C; Figure S4) and only mild inflammation (Figures 4D and 4F; Figure S4) in the lungs of protected animals.

In contrast, all mice of the control group lost considerable weight. Lungs of the mock-vaccinated mice developed lesions (Figures 4C–4F; Figure S4) and high viral titers were recorded on days 3 and 4 after challenge (Figure S4F). Two out of six mice recovered from severe weight loss (Figure 4A), possibly due to age-dependent immunity, as the mice were at the day of challenge 4 weeks older than in the first experiment. Nevertheless, the lungs of these two mice were severely damaged and

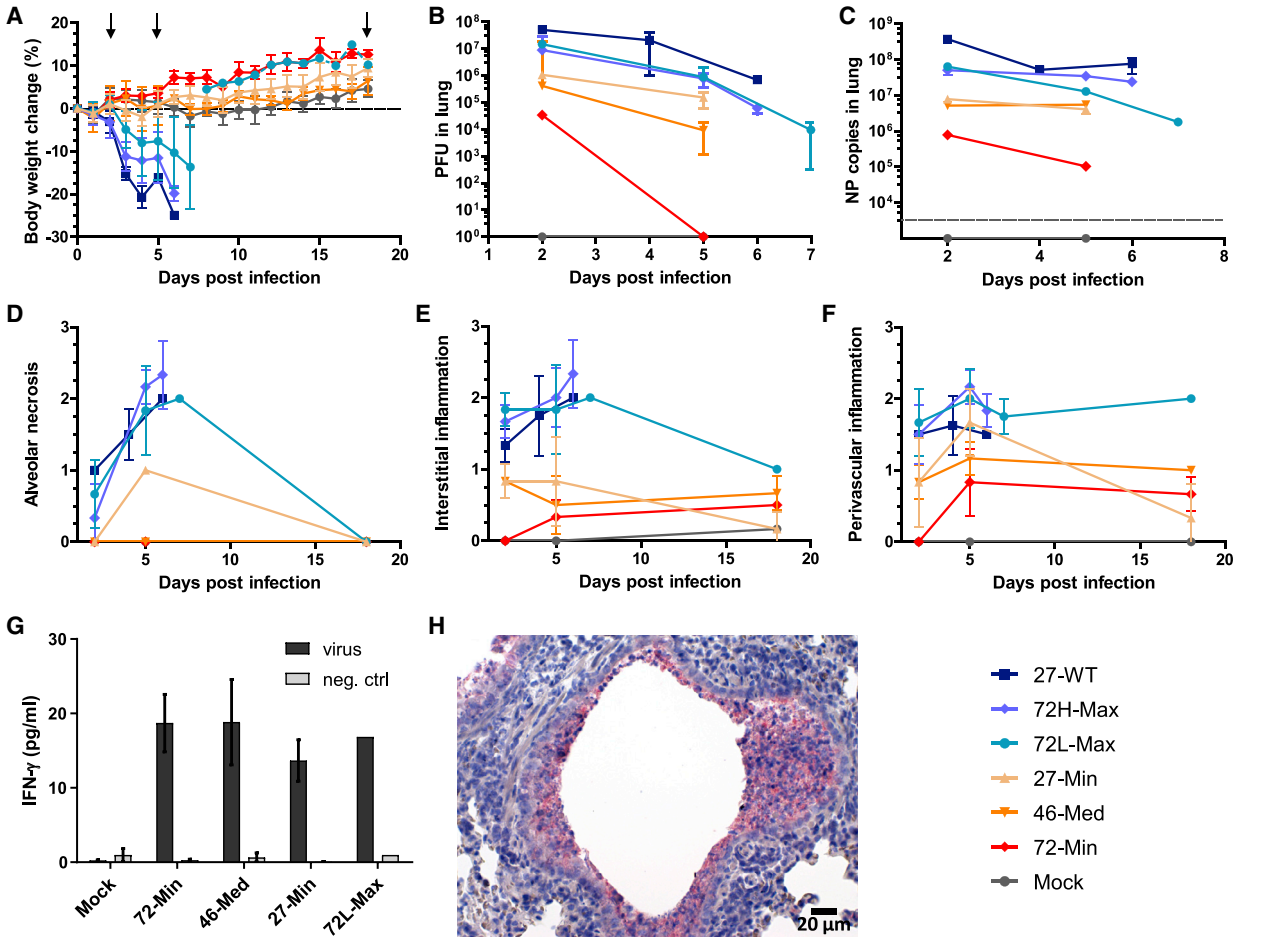



**Figure 3. Pulmonary Histopathology and Immune Responses in Infected Mice**

(A) The relative body weight of the mice of the indicated groups after infection. Nine 6-week-old BALB/c mice were infected with the parental or recoded viruses. On days 2, 5, and 18 after infection, three animals of each group were euthanized (arrows), or when they lost more than 20% of their body weight. None of the codon pair-deoptimized viruses (72-Min, 46-Med, and 27-Min) caused outward disease symptoms. Data show the mean ± SD.

(B and C) Quantification of virus replication in lungs of infected mice. The right lung lobes were homogenized and viral titers determined by plaque assay in MDCK cells (B) and by quantification of NP RNA molecules by qPCR (C). Data show the mean ± SD.

(D–F) Histopathological analysis of the mice lungs. The severity and distribution of alveolar necrosis (D), interstitial inflammation (E), and perivascular lymphoplasmacytic inflammation (F) in lungs of infected animals were assessed and scored according the following scheme: 0, no lesions; 1, mild; 2, moderate; 3, severe lesions. Data show the mean ± SD.

(G) Production of IFN-γ by splenocytes. Mouse splenocytes were isolated 18-days post infection, stimulated with purified IAV virus or control protein for 3 days, and production of IFN-γ was measured by ELISA. Data show the mean ± SD.

(H) Immunohistochemical staining confirmed presence of IAV antigens in the necrotic bronchiolar epithelium of infected mice (red color). Scale bar, 20 μm.

showed signs of regeneration with prominent pneumocyte type II hyperplasia (Figure 4E; Figure S4) and moderate inflammation (Figures 4D and 4F; Figure S4).

## CPD Causes Unstable RNA and Inefficient Translation

While it is known that CPD causes attenuation of recoded viruses because recoded genes produce less protein, it is unknown which sequence features that arise through recoding are primarily responsible for the reduction of protein production. Protein production is primarily controlled by the quantity and translatability of cytoplasmic mRNA. The amount of mRNA in the cytoplasm is mainly influenced by the RNA synthesis and degradation rates (Paulsen et al., 2014). Yet, transcription and translation of mRNA derived from genes manipulated by CPD have received only cursory examination and, importantly, mRNA stability has never been studied (Broadbent et al., 2016; Coleman et al., 2008; Eschke et al., 2018; Futcher et al., 2015; Le Nouën et al., 2014; Mueller et al., 2010; Shen et al., 2015; Wang et al., 2015). We aimed to determine whether the primary cause of attenuation by CPD lies at the level of mRNA production, decay, or translation.

We first assessed RNA kinetics and protein production from the parental and recoded NA genes in transiently transfected HEK293T cells. To quantify expression, we constructed dual

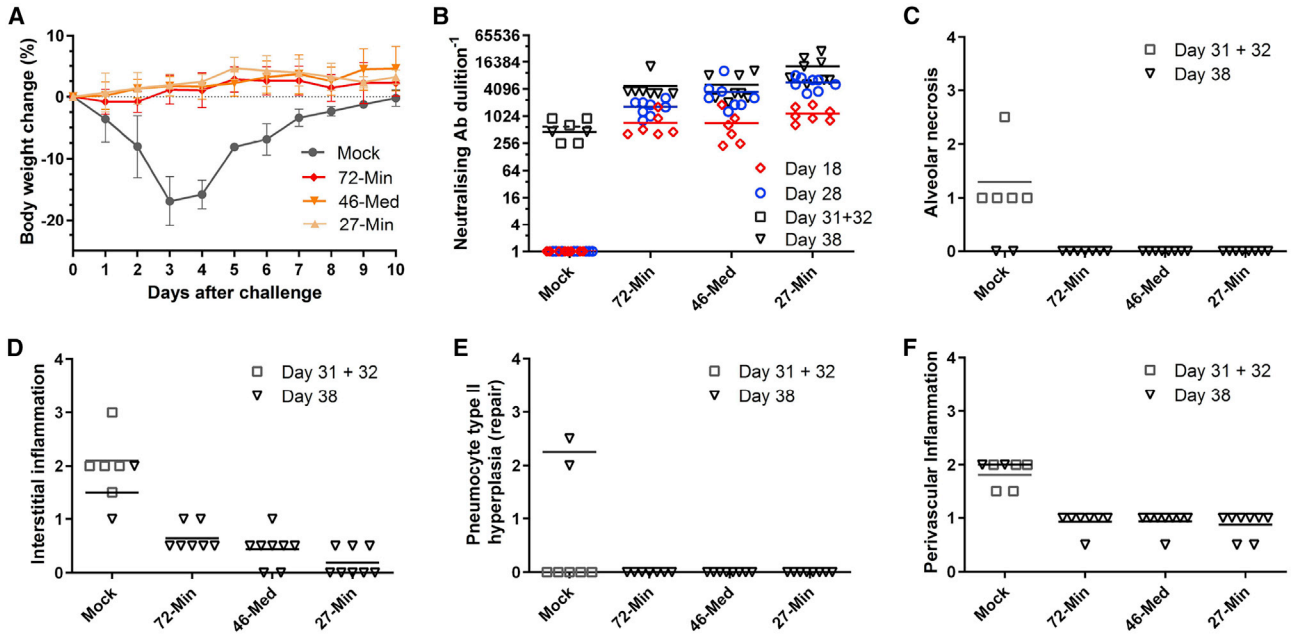



**Figure 4. Immune Protection of Vaccinated Mice**

(A) The relative body weight of mice of the indicated groups after challenge with a parental 27-WT virus. Mice from the control group that lost more than 20% of their body weight were euthanized. Two mice from the control group survived the challenge infection and started to gain weight on day 4 after challenge. Data show as the mean ± SD.

(B) Anti-IAV neutralizing antibody titers in serum at days 18, 28 (day of challenge), and 38 after vaccination. Virus neutralization titers were determined in MDCK cells and are expressed as the reciprocals of the highest serum dilution that showed complete inhibition of infection with 200 PFUs of 27-WT virus after 48 h. 72-Min, 46-Med, and 27-Min show similar high antibody titers. Mock control group shows high titers after 27-WT challenge.

(C–F) Histopathological analysis of the (mock-)immunized mice lung after challenge with the parental IAV virus. The severity and distribution of alveolar necrosis (C), interstitial inflammation (D), pneumocyte type II hyperplasia (repair) (E), and perivascular inflammation (F) were assessed and scored on a scale from 0 (no lesions) to 3 (severe lesions).

expression plasmids, in which the expression of EGFP and NA genes was driven by two different promoters. The plasmids were transfected into HEK293T cells, and NA mRNA and protein production was analyzed 24 h after transfection. NA mRNA levels were significantly lower in cells transfected with plasmids containing the codon pair-deoptimized genes 27-Min, 46-Med, and 72-Min than in cells transfected with plasmid expressing parental the 27-WT gene (Figure 5A). In contrast, NA mRNA levels were only moderately decreased in the case of 72L-Max (75% of the 27-WT) and increased in the case of 72H-Max (170% of the 27-WT).

Protein amounts of the recoded NA genes were compared to that of the 27-WT gene by western blotting (Figure 5B). The codon pair-deoptimized NA genes 27-Min, 46-Med, and 72-Min produced less NA protein, while genes with increased number of CpG dinucleotides produced comparable (72L-Max) or increased (72H-Max) amounts of NA protein than the 27-WT gene (Figure 5B). It is important to note that the NA protein levels correlated very well with the levels produced by the viruses in infected cells (Figure 2F), confirming that the capacity of recoded genes to produce proteins is an inherent characteristic of the sequence, and independent of virus replication.

Next, we determined the RNA synthesis and RNA degradation rates of the recoded NA genes. Both experiments were based on the metabolic labeling of RNA with a ribonucleotide analog, bromouridine (BrU) (Paulsen et al., 2014). The quantification of RNA production did not show significant differences between the different recoded NA genes (Figure 5C). In contrast, the rates of RNA decay varied considerably among different NA RNA transcripts as determined by pulse chase labeling (Figure 5D). The RNA levels produced by the 27-WT or 72H-Max, 72L-Max genes remained relatively stable up to 2 h after labeling; however, the RNA produced by the codon pair-deoptimized genes 72-Min, 46-Med, and 27-Min was unstable and was rapidly degraded (Figure 5D).

To confirm that the mRNA transcripts produced by the codon pair-deoptimized genes were degraded faster than those of the 27-WT gene, we studied the stability of the different mRNA NA transcripts within cells by analyzing their half-life following transcription inhibition (Figure 5E). HEK293T cells were transfected with the dual expression plasmids and the synthesis of new mRNA in cells was inhibited with the RNA polymerase II inhibitor flavopiridol after 24 h. The abundance of EGFP and NA transcripts was quantified at different time points post inhibition by qRT-PCR. The degradation rates of mRNA NA transcripts were almost identical to those obtained in the BrU chase experiment (Figure 5D), confirming that mRNA species made by codon pair-deoptimized genes are less stable than those produced by the parental 27-WT gene.

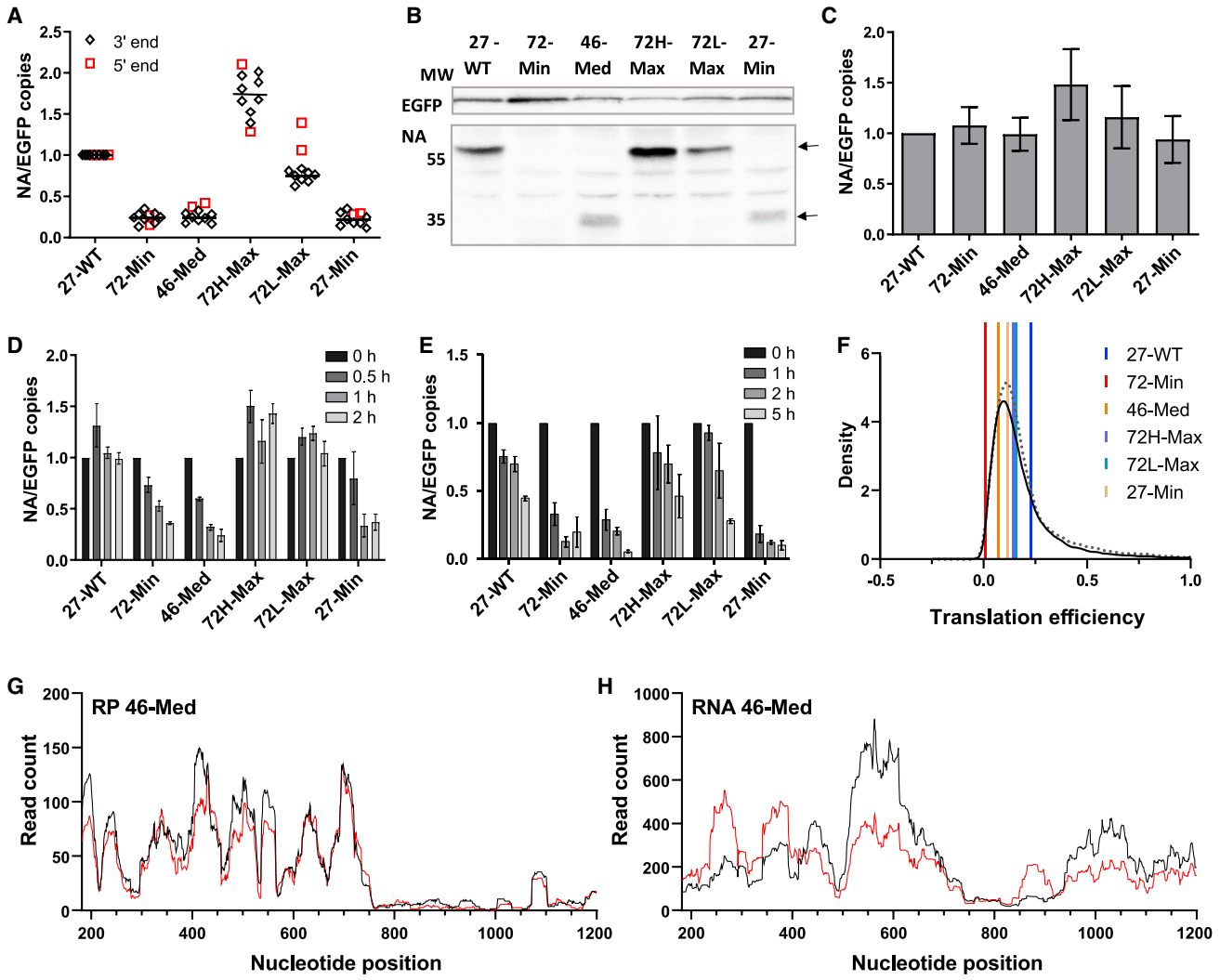



**Figure 5. RNA Kinetics and Protein Production by the Recoded NA Genes in Transiently Transfected HEK293T Cells**

(A) Steady-state RNA levels produced by recoded NA genes in transiently transfected cells. Measurements of eight independent experiments are shown with qPCR probe at the 3′ end of the NA fragment (black) and two independent experiments with qPCR probe at the 5′ end (red).

(B) Western blot analysis of NA protein production from recoded genes in transiently transfected cells from a plasmid background. EGFP protein levels were used for normalization of the transfection efficiency.

(C) Transcription rates of recoded NA genes in HEK293T cells. NA RNA copy numbers were normalized to EGFP copy numbers to account for differences in transfection efficiency. Data represent the means of four biological replicates ± SD.

(D and E) Stability of NA RNA transcripts in transfected cells determined by BrU pulse-chase labeling experiment (D) or transcription inhibitor flavopiridol (E). Data are the means of three biological replicates ± SD.

(F) Translation efficiency of 29-nt footprints of recoded NA genes in transiently transfected cells from ribosome profiling experiments. Translation efficiency of all cellular genes is given as a density plot with dotted line as SD for different experiments.

(G and H) Read density of sequencing (G) ribosome footprints and (H) total RNA from recoded part of 46-Med in transiently transfected cells. Red and black read densities represent two replicates.

Translation efficiency was determined by ribosome profiling experiments (Figures 5F–5H; Figure S6). Numerous studies have shown that 5′ untranslated regions (UTRs) and N-terminal protein-coding sequences can modulate translation efficiency of genes in prokaryotes and eukaryotes because they determine the structural stability of RNA at its 5′ end, which in turn influences translation initiation rates (Goodman et al., 2013; Gu et al., 2010; Tuller et al., 2010). However, because all recoded NA genes have identical 5′ and 3′ UTRs and N- and C-terminal protein-coding sequences, we could exclude that potential differences in NA protein production are due to cis-regulatory elements or the secondary structure of the 5′-terminal region of NA mRNA.

We observed a strong reduction of translation efficiency in the codon pair-deoptimized constructs 72-Min, 46-Med, and 27-Min (Figure 5F). The 72-Min gene had the lowest translation

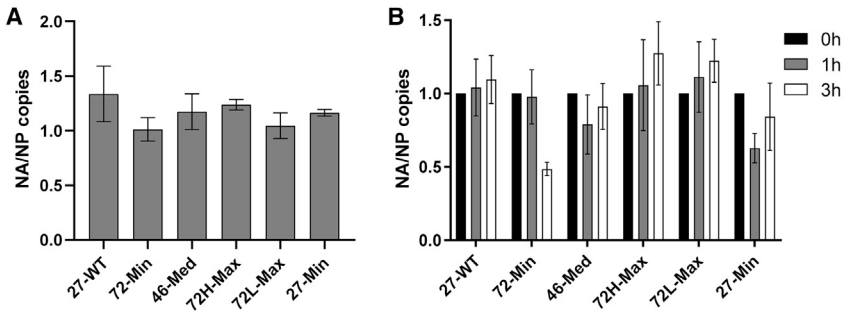



**Figure 6. RNA Kinetics of Recoded Viruses**

(A) Relative NA to NP vRNA levels detected from supernatants of infected MDCK cells. Quantitative analysis of three independent experiments are shown with qPCR probe at the 3′ end of the NA and NP fragment.

(B) Stability of NA mRNA transcripts in infected MDCK cells determined by BrU pulse-chase labeling experiment. Data are the means of three biological replicates ± SD.

efficiency, which is in good agreement with the strong attenuation of the corresponding mutant *in vitro* and *in vivo*. While 46-Med and 27-Min also exhibited reduced translation efficiency, we also identified an almost complete stop of translation after 760 nucleotides (Figure 5G; Figure S6). We confirmed presence of full-length mRNA transcripts by qPCR with probes at the 3′ and 5′ ends (Figure 5A). The truncated protein could also be detected by western blotting (Figure 5B). Mass spectrometric analysis showed that the truncated NA polypeptide contained only the first 250 N-terminal amino acid residues. We did not identify a similar premature cessation of translation in the third codon pair-deoptimized construct, 72-Min. The translation efficiency in this case was extremely low in this case and we obtained only a few ribosome footprints, although the RNA levels were comparable to other samples. In contrast, 72L-Max and 72H-Max showed high translation efficiencies. This corresponded well with the results showing that underrepresented codon pairs are the cause of reduced translation efficiency and lead to reduced protein production.

The analysis of RNA kinetics in transiently transfected cells showed that the stability and translatability of RNA transcripts produced by codon pair-deoptimized genes is reduced (Figures 5D–5F). To examine whether the same molecular mechanisms might also be responsible for reduced protein production in the virus context, we analyzed the stability of RNA produced by the recoded viruses in infected cells. We first examined the genetic composition of virus particles to rule out the possibility that recoded viruses contain a skewed ratio of NA to NP vRNA in their virions and found that the NA/NP vRNA ratio was unaffected (Figure 6A). The analysis of RNA stability in infected MDCK cells showed that NA RNA transcripts produced by codon pair-optimized viruses decay faster than those of the parental virus (Figure 6B). The stability of viral NA RNA was slightly reduced in 46-Med, and markedly reduced in 72-Min and 27-Min viruses (Figure 6B).

## DISCUSSION

Several studies have shown that increasing the frequency of CpG dinucleotides in the RNA virus genome reduces the virus fitness (Atkinson et al., 2014; Burns et al., 2009; Tulloch et al., 2014). Since CPD unintentionally increases the frequency of CpG dinucleotides in recoded sequences, it was hypothesized that the attenuation of mammalian codon pair-deoptimized viruses can be explained by the increased frequency of CpG dinucleotides in their genomes (Tulloch et al., 2014). In this study, we have provided several lines of evidence that prove the opposite—namely that underrepresented codon pairs rather than CpG dinucleotides are primary determinants of CPD-based attenuation. We achieved virus attenuation by CPD without increasing the frequency of CpG dinucleotides in a recoded sequence (27-Min), as evidenced by the fact that codon pair-deoptimized viruses form smaller plaques on MDCK and A549 cells (Figures 2C–2E) and produce less NA protein than the parental virus (Figure 2F). Similarly, we restored protein production and reversed the attenuated phenotype of the 72-Min mutant virus by replacing the non-optimal codon pairs with optimal codon pairs, and thereby producing the 72L-Max and 72H-Max genes. Mutant viruses 72L-Max and 72H-Max show no signs of attenuation *in vitro* (Figure 2) or *in vivo* (Figure 3), although they still contain a large number of CpG dinucleotides.

We observed that the parental and all mutant viruses replicated with similar kinetics and produced similar titers on MDCK cells (Figure 2A), but three codon pair-deoptimized viruses—27-Min, 46-Med, and 72-Min—showed a strongly attenuated phenotype on A549 cells (Figure 2B), growing to 100–1,000 times lower titer than the parental virus. While the relatively unimpeded growth of viruses that produce only small amounts of NA in MDCK cells may seem surprising, a similar observation has been made previously in a study that analyzed replication of a codon pair-deoptimized NA IAV mutant in MDCK and A549 cells (Yang et al., 2013). Our results are consistent with observations suggesting that dissimilar amounts of mucus are responsible for differences in the replication of IAV on the two cell lines. MDCK cells only produce a relatively thin layer of mucus (Barnard et al., 2019; Harush-Frenkel et al., 2008). In contrast, human alveolar carcinoma A549 cells are secretory cells and produce substantial amounts of mucin (Barnard et al., 2019; Berger et al., 1999). In addition, A549 cells, like epithelial cells of the respiratory tract, harbor a layer of mucus that is rich in strongly sialylated glycoproteins. It appears that IAV mutants do not need high NA levels to overcome the relatively modest physical barriers to spread efficiently in MDCK cells. Thus, single-step growth kinetics on MDCK cells have only limited discriminatory power to reveal compromised fitness of codon pair-deoptimized viruses.

In addition to CpG dinucleotides, CPD of vertebrate viruses also increases the frequency of TpA/UpA dinucleotides in recoded sequences, albeit to a lesser degree (Kunec and Osterrieder, 2016). Indeed, all of our codon pair-deoptimized viruses contain more UpA dinucleotides than the parental 27-WT

sequence (Table 1). However, the number of UpA dinucleotides in genomes of analyzed viruses did not correlate with their replication or fitness levels either. For example, the 72L-Max mutant contains a high number of UpA and CpG dinucleotides but did not show any attenuation of replication (Figure 2; Table 1). Thus, we conclude that neither UpA nor CpG dinucleotides play a prominent role in the attenuation of recoded IAV mutants. The results presented in this study refute the hypothesis, which suggests that the effect of attenuation by CPD can be explained by the number of CpG and UpA dinucleotides in codon pair-deoptimized sequences. Two different studies evaluated the effect of CPD, and CpG/UpA dinucleotide frequencies on replication and pathogenicity of recoded influenza A virus mutants (strain A/PR8/34 [H1N1]) (Gaunt et al., 2016; Yang et al., 2013). Although these studies are not directly comparable, the level of attenuation achieved by the elevation of CpG/UpA dinucleotide frequencies (Gaunt et al., 2016) appears to be much weaker than that by CPD (Yang et al., 2013). This observation is in line with our results and indicates that the attenuation strategy based on artificial elevation of CpG/UpA dinucleotide frequencies is much less efficient than CPD. Several recent studies have also shown that codon optimality is a major determinant of mRNA stability (Boël et al., 2016; Presnyak et al., 2015). Our results demonstrate that codon pair choice is another major determinant of RNA stability. Introduction of underrepresented codon pairs into recoded NA genes did not have a measurable effect on RNA transcription (Figure 5C), but it severely impaired the stability of mRNA transcripts and their translation efficiency (Figures 5D–5F).

In a similar study, IAV double mutants with codon pair deoptimized NA and hemagglutinin genes were completely attenuated in BALB/c mice (Yang et al., 2013) and ferrets (Broadbent et al., 2016). Similar to our results, the recoded NA RNA levels in infected cells were significantly reduced. Our study explains this observation as we show that transcripts produced by codon pair-deoptimized genes decay much faster than parental transcripts. We showed that the stability of RNA transcripts produced by codon pair-deoptimized genes is reduced in both transfected (Figures 5D and 5E) and infected cells (Figure 6B). While the increase of CpG (and UpA) dinucleotides in recoded sequences might contribute to virus attenuation, our data show that the increase of CpG dinucleotides in codon pair-deoptimized sequences had negligible effects on protein production and virus attenuation. Thus, it is plausible that CPD leads to virus attenuation by different mechanisms. We argue that the principal role in attenuation by CPD is exerted by underrepresented codon pairs, which reduce the stability of recoded mRNA and inhibit efficient translation.

We showed that neither CpG dinucleotides nor codon pairs of the NNC-GNN type play an important role in the attenuation of recoded viruses (72L-Max and 72H-Max). Since codon pairs of the NNC-GNN type have extremely low CPS, CPD of vertebrate virus genomes leads to an accumulation of these codon pairs in recoded sequences. We speculated that codon pairs of the NNC-GNN type have low CPS, but they do not correlate proportionally with their role in attenuation by CPD. For example, the low average CPS of 72L-Max is solely due to NNC-GNN codon pairs (Figures 1B and 1C), but the virus mutant 72L-Max is not attenuated. It is possible that a modification of the current recoding approach, which would not lead to the accumulation of underrepresented codon pairs of the NNC-GNN type, could lead to even stronger attenuation of recoded viruses.

As a final implication, our work suggests that inhibitory effects of individual non-optimal codon pairs are unequal. The identification of a clear premature translation termination point in different recoded genes (Figure 5G; Figure S6) suggests that reduction of protein production is not the sum of small negative effects exerted by individual underrepresented, non-optimal codon pairs (Coleman et al., 2008). Instead, some codon pairs might have strong inhibitory effects on translation and associated mRNA decay. This emphasizes the need for future research aimed at determining the contribution of individual codon pairs to mRNA destabilization and translation inhibition. Such knowledge could ultimately lead to the development of safer, non-reverting attenuated live attenuated virus vaccines for animal and human use.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell culture
  - Mouse lines
  - Animal experiments
- METHOD DETAILS
  - Virus design
  - Cell culture and virus recovery
  - Virus titration and plaque size measurement
  - Multi-step growth kinetics
  - Western blotting
  - Mass spectrometry
  - Transient transfections
  - RNA extraction and quantitative RT-PCR
  - Relative number of NA and NP copies in virions
  - RNA stability assay based on transcription inhibition
  - RNA synthesis and degradation assays based on RNA labeling with 5-bromouridine
  - Ribosome profiling
  - Viral load and copy number in the lungs
  - Antibody titers
  - Cytokine detection by ELISA
  - Histopathology and immunohistochemistry
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Histopathology of mouse lungs
  - Statistical analysis

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.107586.

## ACKNOWLEDGMENTS

We thank Florian Heyd for his help with the RNA assays and Fan Liu and Heike Stephanowitz for mass-spectometrical analysis of proteins. Annett Neubert, Mathias Kaiser, and Michaela Scholz provided expert technical assistance. We are grateful to Kristina Dietert and Robert Klopfleisch for support in histopathological evaluation. We thank the animal caretakers for excellent assistance during our animal studies. This work was supported by the Deutsche Forschungsgemeinschaft (OS 143/9-1 and 143/12-1).

## AUTHOR CONTRIBUTIONS

Conceptualization, N.G., E.W., N.O., and D.K.; Investigation, N.G., J.T., S.M., A.M.C., and H.Z.; Data Curation, N.G., J.T., S.M., A.M.C., E.W., H.Z., O.-G.H., and S.R.; Formal Analysis, N.G., E.W., and D.K.; Validation, N.G.; Software: N.G., E.W., O.-G.H., and D.K.; Visualization: N.G.; Writing – Original Draft, N.G.; Writing – Review & Editing, E.W., M.L., N.O., and D.K.; Funding Acquisition, M.L., N.O., and D.K.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 22, 2019
Revised: March 6, 2020
Accepted: April 8, 2020
Published: April 28, 2020

## REFERENCES

Atkinson, N.J., Witteveldt, J., Evans, D.J., and Simmonds, P. (2014). The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. Nucleic Acids Res. *42*, 4527–4545.

Barnard, K.N., Wasik, B.R., LaClair, J.R., Buchholz, D.W., Weichert, W.S., Alford-Lawrence, B.K., Aguilar, H.C., and Parrish, C.R. (2019). Expression of 9-*O*- and 7,9-*O*-acetyl modified sialic acid in cells and their effects on influenza viruses. MBio *10*, e02490-19.

Berger, J.T., Voynow, J.A., Peters, K.W., and Rose, M.C. (1999). Respiratory carcinoma cell lines. MUC genes and glycoconjugates. Am. J. Respir. Cell Mol. Biol. *20*, 500–510.

Boël, G., Letso, R., Neely, H., Price, W.N., Wong, K.H., Su, M., Luff, J., Valecha, M., Everett, J.K., Acton, T.B., et al. (2016). Codon influence on protein expression in E. coli correlates with mRNA levels. Nature *529*, 358–363.

Broadbent, A.J., Santos, C.P., Anafu, A., Wimmer, E., Mueller, S., and Subbarao, K. (2016). Evaluation of the attenuation, immunogenicity, and efficacy of a live virus vaccine generated by codon-pair bias de-optimization of the 2009 pandemic H1N1 influenza virus, in ferrets. Vaccine *34*, 563–570.

Burns, C.C., Campagnoli, R., Shaw, J., Vincent, A., Jorba, J., and Kew, O. (2009). Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. J. Virol. *83*, 9957–9969.

Calviello, L., Mukherjee, N., Wyler, E., Zauber, H., Hirsekorn, A., Selbach, M., Landthaler, M., Obermayer, B., and Ohler, U. (2016). Detecting actively translated open reading frames in ribosome profiling data. Nat. Methods *13*, 165–170.

Coleman, J.R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E., and Mueller, S. (2008). Virus attenuation by genome-scale changes in codon pair bias. Science *320*, 1784–1787.

Dalby, B., Cates, S., Harris, A., Ohki, E.C., Tilkins, M.L., Price, P.J., and Ciccarone, V.C. (2004). Advanced transfection with Lipofectamine 2000 reagent: primary neurons, siRNA, and high-throughput applications. Methods *33*, 95–103.

Dietert, K., Gutbier, B., Wienhold, S.M., Reppe, K., Jiang, X., Yao, L., Chaput, C., Naujoks, J., Brack, M., Kupke, A., et al. (2017). Spectrum of pathogen- and model-specific histopathologies in mouse models of acute pneumonia. PLoS ONE *12*, e0188251.

Dodt, M., Roehr, J.T., Ahmed, R., and Dieterich, C. (2012). FLEXBAR—flexible barcode and adapter processing for next-generation sequencing platforms. Biology (Basel) *1*, 895–905.

Eisfeld, A.J., Neumann, G., and Kawaoka, Y. (2015). At the centre: influenza A virus ribonucleoproteins. Nat. Rev. Microbiol. *13*, 28–41.

Eschke, K., Trimpert, J., Osterrieder, N., and Kunec, D. (2018). Attenuation of a very virulent Marek's disease herpesvirus (MDV) by codon pair bias deoptimization. PLoS Pathog. *14*, e1006857.

Ficarelli, M., Wilson, H., Pedro Galão, R., Mazzon, M., Antzin-Anduetza, I., Marsh, M., Neil, S.J., and Swanson, C.M. (2019). KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides. eLife *8*, e46767.

Futcher, B., Gorbatsevych, O., Shen, S.H., Stauft, C.B., Song, Y., Wang, B., Leatherwood, J., Gardin, J., Yurovsky, A., Mueller, S., and Wimmer, E. (2015). Reply to Simmonds et al.: codon pair and dinucleotide bias have not been functionally distinguished. Proc. Natl. Acad. Sci. USA *112*, E3635–E3636.

Gaidatzis, D., Lerch, A., Hahne, F., and Stadler, M.B. (2015). QuasR: quantification and annotation of short reads in R. Bioinformatics *31*, 1130–1132.

Gaunt, E., Wise, H.M., Zhang, H., Lee, L.N., Atkinson, N.J., Nicol, M.Q., Highton, A.J., Klenerman, P., Beard, P.M., Dutia, B.M., et al. (2016). Elevation of CpG frequencies in influenza A genome attenuates pathogenicity but enhances host response to infection. eLife *5*, e12735.

Goodman, D.B., Church, G.M., and Kosuri, S. (2013). Causes and effects of N-terminal codon bias in bacterial genes. Science *342*, 475–479.

Greenbaum, B.D., Rabadan, R., and Levine, A.J. (2009). Patterns of oligonucleotide sequences in viral and host cell RNA identify mediators of the host innate immune system. PLoS ONE *4*, e5969.

Gu, W., Zhou, T., and Wilke, C.O. (2010). A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. PLoS Comput. Biol. *6*, e1000664.

Gutman, G.A., and Hatfield, G.W. (1989). Nonrandom utilization of codon pairs in Escherichia coli. Proc. Natl. Acad. Sci. USA *86*, 3699–3703.

Harush-Frenkel, O., Rozentur, E., Benita, S., and Altschuler, Y. (2008). Surface charge of nanoparticles determines their endocytic and transcytotic pathway in polarized MDCK cells. Biomacromolecules *9*, 435–443.

Hoffmann, E., Neumann, G., Kawaoka, Y., Hobom, G., and Webster, R.G. (2000). A DNA transfection system for generation of influenza A virus from eight plasmids. Proc. Natl. Acad. Sci. USA *97*, 6108–6113.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kunec, D., and Osterrieder, N. (2016). Codon pair bias is a direct consequence of dinucleotide bias. Cell Rep. *14*, 55–67.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Le Nouën, C., Brock, L.G., Luongo, C., McCarty, T., Yang, L., Mehedi, M., Wimmer, E., Mueller, S., Collins, P.L., Buchholz, U.J., and DiNapoli, J.M. (2014). Attenuation of human respiratory syncytial virus by genome-scale codon-pair deoptimization. Proc. Natl. Acad. Sci. USA *111*, 13169–13174.

Martrus, G., Nevot, M., Andres, C., Clotet, B., and Martinez, M.A. (2013). Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture. Retrovirology *10*, 78.

Morrison, E., Kuropka, B., Kliche, S., Brügger, B., Krause, E., and Freund, C. (2015). Quantitative analysis of the human T cell palmitome. Sci. Rep. *5*, 11598.

Mueller, S., Papamichail, D., Coleman, J.R., Skiena, S., and Wimmer, E. (2006). Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. J. Virol. *80*, 9687–9696.

Mueller, S., Coleman, J.R., Papamichail, D., Ward, C.B., Nimnual, A., Futcher, B., Skiena, S., and Wimmer, E. (2010). Live attenuated influenza virus vaccines by computer-aided rational design. Nat. Biotechnol. *28*, 723–726.

Paulsen, M.T., Veloso, A., Prasad, J., Bedi, K., Ljungman, E.A., Magnuson, B., Wilson, T.E., and Ljungman, M. (2014). Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. Methods *67*, 45–54.

Plumb, D.C. (2015). Plumb's Veterinary Drug Handbook (PharmaVet).

Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., and Coller, J. (2015). Codon optimality is a major determinant of mRNA stability. Cell *160*, 1111–1124.

Schierhorn, K.L., Jolmes, F., Bespalowa, J., Saenger, S., Peteranderl, C., Dzieciolowski, J., Mielke, M., Budt, M., Pleschka, S., Herrmann, A., et al. (2017). Influenza A virus virulence depends on two amino acids in the N-terminal domain of its NS1 protein to facilitate inhibition of the RNA-dependent protein kinase PKR. J. Virol. *91*, e00198-17.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Shen, S.H., Stauft, C.B., Gorbatsevych, O., Song, Y., Ward, C.B., Yurovsky, A., Mueller, S., Futcher, B., and Wimmer, E. (2015). Large-scale recoding of an arbovirus genome to rebalance its insect versus mammalian preference. Proc. Natl. Acad. Sci. USA *112*, 4749–4754.

Simmonds, P., Tulloch, F., Evans, D.J., and Ryan, M.D. (2015). Attenuation of dengue (and other RNA viruses) with codon pair recoding can be explained by increased CpG/UpA dinucleotide frequencies. Proc. Natl. Acad. Sci. USA *112*, E3633–E3634.

Stuart-Harris, C. (1939). A neurotropic strain of human influenza virus. Lancet *233*, 497–499.

Takata, M.A., Gonçalves-Carneiro, D., Zang, T.M., Soll, S.J., York, A., Blanco-Melo, D., and Bieniasz, P.D. (2017). CG dinucleotide suppression enables antiviral defence targeting non-self RNA. Nature *550*, 124–127.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA *76*, 4350–4354.

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell *141*, 344–354.

Tulloch, F., Atkinson, N.J., Evans, D.J., Ryan, M.D., and Simmonds, P. (2014). RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. eLife *3*, e04531.

Wang, B., Yang, C., Tekes, G., Mueller, S., Paul, A., Whelan, S.P., and Wimmer, E. (2015). Recoding of the vesicular stomatitis virus L gene by computer-aided design provides a live, attenuated vaccine candidate. MBio *6*, e00237-15.

WHO (2002). WHO Manual on Animal Influenza Diagnosis and Surveillance, Available at. http://www.who.int/csr/resources/publications/influenza/whocdscsrncs20025rev.pdf. (Accessed 16 April 2019).

Yang, C., Skiena, S., Futcher, B., Mueller, S., and Wimmer, E. (2013). Deliberate reduction of hemagglutinin and neuraminidase expression of influenza virus leads to an ultraprotective live vaccine in mice. Proc. Natl. Acad. Sci. USA *110*, 9481–9486.

Zhong, Y., Karaletsos, T., Drewe, P., Sreedharan, V.T., Kuo, D., Singh, K., Wendel, H.G., and Rätsch, G. (2017). RiboDiff: detecting changes of mRNA translation efficiency from ribosome footprints. Bioinformatics *33*, 139–141.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| anti-NA | GeneTex | Cat# GTX629696; RRID:AB_2728684 |
| anti-NP | GeneTex | Cat#GTX629544 |
| anti-NS1 | Schierhorn et al., 2017 | PMID: 28250123 |
| anti-GFP | Cell Signaling | Cat#2956; RRID:AB_1196614 |
| anti-BrU | BD PharMingen | Cat#555627; RRID:AB_395993 |
| anti-FLAG | Sigma-Aldrich | Cat#F2555; RRID:AB_796202 |
| anti-Influenza A H1N1 | BIO RAD | Cat#5315-0064; RRID:AB_620739 |
| anti-beta-actin | Cell Signaling | Cat#4970; RRID:AB_2223172 |
| anti-Influenza A/WSN/1933 (H1N1) | This paper | N/A |
| Bacterial and Virus Strains | | |
| A/WSN/1933 (H1N1), 8 plasmid system | Hoffmann et al., 2000 | PMID: 10801978 |
| Chemicals, Peptides, and Recombinant Proteins | | |
| Flavopiridol | MedChem Express | Cat#HY-10005 |
| 5-Bromouridine (BrU) | Sigma-Aldrich | Cat#850187 |
| Uridine | Sigma-Aldrich | Cat#U3750 |
| Protein G Dynabeads | Thermo Fisher | Cat#10004D |
| Critical Commercial Assays | | |
| RNeasy Plus Mini Kit | QIAGEN | Cat#74136 |
| RTP DNA/RNA Virus Mini Kit | Stratec | Cat#1040100300 |
| TruSeq Stranded mRNA Kit | Illumina | Cat#20020595 |
| Deposited Data | | |
| Neuraminidase gene sequences | This paper | GenBank: MN176602, MN176603, MN176604, MN176605, MN176606, MN176607 |
| Raw and analyzed ribosome profiling data | This paper | GEO: GSE134752 |
| Experimental Models: Cell Lines | | |
| HEK293T cell | ATCC | CRL-3216 |
| MDCK II cells | ATCC | CCL- 2936 |
| A549 cells | ATCC | CCL-185 |
| Experimental Models: Organisms/Strains | | |
| BALB/cJRj | Janvier | SC-BALBJ-F |
| Oligonucleotides | | |
| Primers and probes for RT-qPCR, see Table S1 | This paper | N/A |
| Recombinant DNA | | |
| Recoded neuraminidase genes | BioBasic | N/A |
| pVITRO2-MCS | InvivoGen | N/A |
| A/WSN/1933 (H1N1) plasmids | Hoffmann et al., 2000 | PMID: 10801978 |
| Software and Algorithms | | |
| GraphPad Prism 8.1.2 | GraphPad Software | https://www.graphpad.com/ |
| R, version 3.5.3 | R Foundation for Statistical Computing | https://www.r-project.org/ |
| RiboDiff | Zhong et al., 2017 | https://github.com/ratschlab/ribodiff |
| ImageJ | Schneider et al., 2012 | https://imagej.nih.gov/ij/ |
| Flexbar | Dodt et al., 2012 | https://github.com/seqan/flexbar |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| tophat2 | Kim et al., 2013 | http://ccb.jhu.edu/software/tophat/index.shtml |
| quasR | Gaidatzis et al., 2015 | https://github.com/fmicompbio/QuasR |
| bowtie2 | Langmead and Salzberg, 2012 | http://bowtie-bio.sourceforge.net/bowtie2/index.shtml |

## RESOURCE AVAILABILITY

### Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dusan Kunec (dusan.kunec@fu-berlin.de).

### Materials Availability

Plasmids generated in this study are available upon request.

### Data and Code Availability

Unprocessed datasets generated in this study are available through Mendeley Data: https://doi.org/10.17632/rsxsyy5wrd.1. Sequences of the recoded IAV NA genes have been deposited in GenBank: MN176602, MN176603, MN176604, MN176605, MN176606, MN176607. Raw and analyzed ribosome profiling data have been deposited in the NCBI's Gene Expression Omnibus database (GEO: GSE134752).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Cell culture

Madin-Darby canine kidney (MDCK.2, ATCC CCL- 2936), human adenocarcinomic alveolar basal epithelial A549 (ATCC CCL-185) and human embryonic kidney HEK293T (ATCC CRL-3216) cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS, penicillin (100 U/ml) and streptomycin (100 μg/ml) at 37°C and 5% $CO_2$.

### Mouse lines

Female 6 -week-old BALB/c mice (BALB/cJRj) were purchased from Janvier and distributed randomly. All the mice were maintained in a pathogen-free animal facility under standard 12 h light/12 h dark cycle at 21°C with access to normal chow and water *ad libitum*.

### Animal experiments

Animal experiments were conducted in accordance with the EU legislation for the use of animals for scientific purposes (Directive 2010/63/EU) and German law (paragraph 8 Tierschutzgesetz). Animal experiments were approved by the Landesamt für Gesundheit und Soziales in Berlin, Germany (approval G0291/17). Female 6-week-old BALB/c mice were purchased from Janvier, distributed randomly to 7 groups of 9 animals, and infected intranasally with $5 \times 10^4$ PFU of IAV in 30 μl DMEM medium under ketamine and xylazine anesthesia (Plumb, 2015). Mice were infected with 27-WT, 72-Min, 46-Med, 72H-Max, 72L-Max, 27-Min virus, or mock-infected with DMEM medium. The body weight was measured every day over the whole duration of the experiment. Mice which showed signs of severe disease (ruffled fur, tremors, respiratory distress, and lack of activity), or lost more than 20% of their body weight were euthanized. On day 2, 5 and 18, three mice from each group were killed and blood, lungs and spleens were collected and analyzed for viral titers and histology (see below). In a second experiment, mice were vaccinated intranasally with 30 μL inoculum containing $5 \times 10^4$ PFU of attenuated mutant viruses 72-Min, 46-Med, 27-Min, or mock-vaccinated with DMEM. On day 28 post vaccination mice were challenged with $5 \times 10^5$ PFU of parental 27-WT virus. Blood was taken on day 18, 28 and 38 after initial infection.

## METHOD DETAILS

### Virus design

The viral mutants were based on the prototype IAV strain A/WSN/1933 (H1N1) (Hoffmann et al., 2000). Viral mutants differed from each other in how the neuraminidase (NA) gene was encoded. To ensure unaltered packaging and replication of the recoded NA segments, only the central part of the NA ORF (1,017 nt, 339 codons) was recoded, leaving the first 183 nt (61 codons) and the last 162 nt (54 codons) of the ORF unmodified. The NA gene was recoded using custom perl scripts, utilizing codon pair bias observed in human protein coding sequences (Kunec and Osterrieder, 2016). Recoding preserved the amino acid sequence but modified the number of CpG dinucleotides and/or codon pair bias of the recoded genes. The recoded NA segments were synthesized (BioBasic, Inc) and

cloned into a pHW vector of the standard 8-plasmid A/WSN/1933 reverse genetics system (Hoffmann et al., 2000). Gene sequences can be retrieved from the NCBI sequence database Genbank: MN176602, MN176603, MN176604, MN176605, MN176606, MN176607.

### Cell culture and virus recovery

Mutant viruses were recovered after transfection of the A/WSN/1933 plasmids into susceptible HEK293T cells (Hoffmann et al., 2000). Infectious medium was harvested and virus was propagated in MDCK.2 cells grown in infection medium (DMEM supplemented with 0.2% BSA, 0.1% FBS, 2 mM Na-pyruvat, 2 μg/mL TPCK-treated trypsin). Cell culture medium was harvested 2-3 days after inoculation, when at least 75% of the cell monolayer exhibited CPE. Virus titers of the cell supernatant were determined by titration on MDCK.2 cells. To confirm sequences of recoded NA genes, viral RNA was isolated using RTP DNA/RNA Virus Mini Kit (Stratec SE), reverse transcribed with M-MLV reverse transcriptase (Promega) and sequenced by Sanger sequencing.

### Virus titration and plaque size measurement

To determine virus titers and virus-induced plaque sizes, MDCK.2 or A549 cells were grown in 6-well plates and infected with 500 μl of serial 10-fold dilutions of the virus. After a 1 h adsorption period, the medium was removed and cells were overlaid with Eagle's minimum essential medium (EMEM) containing 0.6% microcrystalline cellulose Avicel (FMC BioPolymer), 0.1% FBS, 2 μg/ml TPCK-Trypsin, 0.2% BSA, 0.05% $NaHCO_3$, 0.01% diethylaminoethyl-dextran, penicillin (100 U/ml) and streptomycin (100 μg/ml). After 2 to 3 days, cells were washed with 1 × PBS, fixed with 2% paraformaldehyde. MDKC.2 cells were stained with 0.1% crystal violet (in 25% methanol) for 30 min. Immunohistochemistry was performed for A549 cells. Cells were fixed with 2% paraformaldehyde, permeabilized with 0.1% Triton X-100, and blocked with 3% bovine serum albumin (BSA) in PBS. Cells were then incubated with anti-IAV mouse serum obtained from mice used in this study (dilution 1:1000) for 1 h, and then with goat anti-mouse IgG-Alexa Fluor 568 secondary antibody (dilution 1:2000, Invitrogen) for 45 min to visualize plaques. Images of an entire well of a 6-well plate were taken with a plate imager (Bioreader®-6000 FZβ, BIO-SYS), and images of individual plaques were taken at 50- and 100-fold magnification using an inverted fluorescence microscope (Axiovert S100, Zeiss). The plaque diameters of 100 randomly selected plaques of each virus were determined with ImageJ software (Schneider et al., 2012).

### Multi-step growth kinetics

To assess virus growth kinetics, confluent MDCK.2 or A549 cells grown in 12-well plates were infected with 500 μl of recombinant virus at a multiplicity of infection (MOI) of 0.0001. After a 1 h adsorption period, the medium was changed to infection medium (containing 1 μg/mL TPCK-treated trypsin for A549 cells or 2 μg/mL for MDCK.2 cells). At the indicated times after infection, cell culture medium was taken and viral titers were determined by plaque assays.

### Western blotting

Protein production by mutant viruses was assessed in MDCK.2 cells. Confluent cells grown in 6-well plates were infected at an MOI of 5. After 1 h, medium was changed to infection medium and cells were collected and analyzed after an additional 4 h. To assess NA production by transfection experiments, 70% confluent HEK293T cells were transfected with pVITRO2-EGFP-NA expression plasmids that contained the various recoded genes and collected 24 h post transfection. Cells were lysed in radioimmunoprecipitation assay (RIPA) buffer (20 mM Tris, pH 7.5, 150 mM NaCl, 1% Nonidet P-40; 0.5% sodium deoxycholate, 0.1% SDS) containing protease inhibitors (Roche) and benzonase (Novagen). Proteins in cell lysates were separated by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) under reducing conditions, and proteins were transferred to PVDF membranes using a semi-dry blotting system (Towbin et al., 1979). The membranes were blocked for 1 h at room temperature in 3% milk in PBS, with 0.05% Tween-20. Membranes were incubated with primary antibodies for 16 h at 4°C. To detect NA, NP, NS-1, EGFP and beta-actin, monoclonal anti-NA (GT288, GeneTex), anti-NP (GT778, GeneTex), or polyclonal anti-NS1 (Schierhorn et al., 2017), anti-EGFP (Cell Signaling), or anti-beta-actin (Cell Signaling Technology) antibodies, respectively, were used. As secondary antibodies, a horseradish peroxidase-conjugated anti-mouse IgG (1:2000, Sigma) or anti-rabbit IgG antibody (1:2000, Cell Signaling Technology) was used. Antibody binding was visualized by chemiluminescent ECL prime western blotting detection reagent (GE Healthcare).

### Mass spectrometry

The identity of truncated NA protein was determined by mass spectrometry. After separation by SDS-PAGE, gel bands containing the truncated NA protein were cut out, proteins in the gel slices were reduced with dithiothreitol, ankylated with chloroacetamide, and digested *in situ* with trypsin, chymotrypsin, elastase or AspN (Morrison et al., 2015). The digested peptides were extracted from the gel matrix and analyzed by nano-liquid chromatography tandem mass spectrometry on a LTQ Orbitrap Elite mass spectrometer (Thermo Fischer Scientific) (Morrison et al., 2015). Amino acid sequence of NA peptides was determined by searching the UniProt virus database with Mascot search engine.

### Transient transfections

To quantify NA expression, the NA and EGFP genes were cloned into a bicistronic reporter plasmid pVITRO2 (InvivoGen), which allows constitutive co-expression of two genes of interest (Eschke et al., 2018). The obtained plasmids pVITRO2-EGFP-NA were

transfected into HEK293T cells using Lipofectamine (Thermo Fisher Scientific) (Dalby et al., 2004), and NA mRNA and protein production was analyzed 24 h after transfection.

### RNA extraction and quantitative RT-PCR

Total RNA from cells was isolated using the RNeasy Plus Mini kit (QIAGEN) and converted into cDNA with M-MLV reverse transcriptase (Promega) and random primers according to the manufacturer's instructions. cDNA was quantified on the StepOnePlus Real-Time PCR System (Thermo Fisher Scientific) using Hot FIREPol Polymerase (Solis Biodyne). Primers and TaqMan probes are listed in Table S1. Gene-specific primer and probe sets for NA, NP and EGFP were designed using Primer Express 3.0.1 software (Thermo Fisher Scientific). Expression of different NA gene variants was quantified using a single primer and probe set because they were designed to amplify non-recoded 3′ or 5′ end of the NA gene. The copy numbers were determined from the calibration curves generated based on known concentrations of pVITRO2-EGFP-NA, or pHW WSN/33 IAV plasmids.

### Relative number of NA and NP copies in virions

MDCK cell were infected with a MOI of 0.001. After 48 h supernatant was collected and viral RNA was isolated using RTP DNA/RNA Virus Mini Kit (Stratec SE), reverse transcribed with M-MLV reverse transcriptase (Promega) and RT-qPCR from NA and NP fragments was performed as described above.

### RNA stability assay based on transcription inhibition

Stability of transcripts was measured indirectly by analyzing the half-life of RNA transcripts following transcription inhibition with flavopiridol (Hycultec). HEK293T cells were transfected with pVITRO2-EGFP-NA plasmids. Flavopiridol was added to cultured cells 24 h after transfection at a final concentration of 1 μM. Cells were then cultured for additional 0, 1, 2, or 5 h, and RNA was isolated and quantified as described above.

### RNA synthesis and degradation assays based on RNA labeling with 5-bromouridine

The rate of RNA synthesis and degradation was determined in HEK293T cells by metabolic labeling of nascent RNA with a ribonucleotide analog, bromouridine (BrU, Sigma-Aldrich) (Paulsen et al., 2014). Cells grown in 6-well plates were transfected with expression plasmids pVITRO2-EGFP-NA (1 μg/well) and the BrU pulse (-chase) experiments were done 24 h after transfection.

To determine the rate of RNA synthesis, nascent RNA was pulse-labeled with BrU 800 μl of 2 mM BrU (Sigma-Aldrich, #850187) in DMEM containing 10% FBS was added to infected 100% confluent MDCK cells or transfected 80% confluent 293Ts. After 2 h for infection and 1 h for transfection expresiments, excess of BrU was removed. For 0h time point RLT buffer (QIAGEN) was directly added, all other cells were washed 3 times with 20 mM PBS. 1.2 mL Uridine (Sigma-Aldrich, #U3750) in DMEM containing 10% FBS was added for indicated time points (0.5 h to 3 h) and incubated under normal growth conditions. At indicated time points cells were washed with PBS, RLT buffer (QIAGEN) was added and cells were harvested. RNA was isolated with RNeasy Plus kit (QIAGEN).

For the preparation of the purification beads, 15 μL of anti-mouse IgG magnetic Dynabeads (Thermo Fisher Scientific, #10004D) per sample are transferred to a 1.5 mL microfuge Protein Low binding tube. Beads are captured with magnetic stand and storage buffer was removed. For 6 samples beads were resuspend in 600 μL BrU-IP buffer (0.05% Tween in RNase free PBS), captured on the magnetic stand and supernatant was aspirated. After one wash with BrU-IP buffer, beads are resuspended in 300μl (6 samples) BrU-IP buffer supplemented with 1 μl RNase inhibitor (RNasin, Promega). For each sample 1 μl (0.5 μg) BrdU antibody (Becton Dickinson PharMingen, #000000005556270000000000000000000000000) was added. Samples are incubated for 30 min at room temperature with gentle rotation. After three washes of beads as described above, beads are resuspend in 100 μl BrU-IP buffer per sample supplemented with RNase inhibitor.

For isolation of BrU labeled RNA, 100 μl of the magnetic beads conjugated with the BrdU antibody are added to the isolated RNA and 135 μl 1 × BrU-IP buffer was added. Samples are incubated for 60 min at room temperature with gentle rotation in Protein low binding tubes. Beads are washed twice with BrU-IP buffer for 5 min on the rotator. Afterward beads are captured on the magnetic stand and supernatant was aspirated. Three additional washes with BrU-IP buffer were performed and the bead pellet was resuspended in 40 μl DEPC-water and incubate for 10 min in a 95°C heat block to elute BrU-RNA from the beads. Beads are captured on the magnetic stand and supernatant was collected into a clean 1.5 mL microfuge tube. Afterward NA, NP or GFP copies of isolated BrU RNA was reverse transcribed with oligo dT and quantified by RT-qPCR as described above.

For the comparison of the RNA synthesis rates transfected HEK293T cells are labeled for 30 min with BrU and labeled RNA is measured by qPCR following the procedure described above.

### Ribosome profiling

HEK293T cells were transfected with expression plasmids pVITRO2-EGFP-NA for 24 h. Ribosome profiling was performed as described in Calviello et al. (2016), except that the mRNA fragments were excised between 18-nt and 35-nt marker oligos, instead of 27-nt and 30-nt. Unfragmented RNA extracted from cell lysates was converted into sequencing libraries using the Illumina TruSeq Stranded mRNA kit. Samples were pooled and sequenced on a HiSeq 4000 (Illumina) device to generate 51-nt single-end reads. Read sequences and analyses have been deposited into the Gene Expression Omnibus (GEO) database and are accessible through GEO Series accession number GSE134752. Sequencing reads were demultiplexed and adaptor-trimmed using FLEXBAR (Dodt

et al., 2012) and mapped to the hg19 version of the human genome using tophat2 (Kim et al., 2013) or to the NA construct sequences using bowtie2 (Langmead and Salzberg, 2012). Only reads with mapping quality 24 or higher were kept, which also filtered out reads mapping to the constant regions at the beginning and end of the NA constructs. Sequencing reads per gene as annotated in the RefSeq database were counted using quasR (Gaidatzis et al., 2015). Translation efficiencies from readcounts were calculated using RiboDiff (Zhong et al., 2017).

### Viral load and copy number in the lungs

The right lung lobes were homogenized in 1 mL DMEM and virus titer was determined by plaque assay as described earlier. Total RNA was isolated from 200 μl of the homogenized lung and the RNA copy numbers of IAV NA and NP fragments were determined by RT-qPCR as described above.

### Antibody titers

Anti-IAV antibody titers were assessed by hemagglutination inhibition (HAI), and virus neutralization assays in blood serum collected on day 18 post infection from first animal experiment exactly as described in the World Health Organization manual on animal influenza diagnosis and surveillance (WHO, 2002). In the vaccination experiment blood serum was taken on day 18, 28 and 38 and anti IAV antibody titers were determined with virus neutralization assay (WHO, 2002).

### Cytokine detection by ELISA

Spleens were removed from mice, homogenized and filtered through 70 μm cell strainers (Becton Dickinson) to obtain single cell suspensions. Erythrocytes in spleen cell suspension were lysed using ACK buffer containing 150 mM $NH_4Cl$, 0.1 mM $KHCO_3$ and 0.1 mM $Na_2EDTA$, pH 7.2 for 3 min. Cells were washed and re-suspended in complete RPMI 1640 medium (cRPMI) and counted using a CASY automated cell counter (Roche). A total of $1 \times 10^6$ splenocytes were stimulated with 10 μg/ml purified 27-WT virus or 10 μg/ml control protein purified from cell supernatant for 3 days. Cells were cultured in cRPMI containing 10% FCS, and Penicillin [100 U/ml]/ Streptomycin [100 μg/ml]. Supernatants were analyzed for IFN-γ using Ready-Set-Go Elisa Kits (eBioscience) according to the manufacturer's instructions.

For purification of virus antigen, 27-WT was grown in MDCK.2 cells for 48 h. Supernatants were collected and cleared from cell debris by centrifugation for 10 min at 5,000 × g. Virus was purified using a 20% sucrose cushion in TNE buffer (10 mM Tris, 1 mM EDTA, 100 mM NaCl) by centrifugation for 2 h at 100,000 × g. The resulting virus pellet was dissolved in TNE buffer and protein concentration was determined with BCA assay. As a control protein supernatant of uninfected MDCK.2 cells was used.

### Histopathology and immunohistochemistry

The left lung of mice was carefully removed, fixed in formalin pH 7.0 for 48 h, embedded in paraffin, and cut into 2 μm sections. Lung sections stained with hematoxylin and eosin were evaluated by microscopy in a blinded fashion (Dietert et al., 2017). Histological examination revealed a moderate to severe, multifocal to coalescing, acute necrotizing bronchointerstitial pneumonia. The pathological features assessed were damage to the airway epithelium (necrosis of bronchi, bronchioli and alveoli) and inflammation (perivascular, interstitial, bronchus associated lymphoid tissue (BALT)) as well as pneumocyte type II hyperplasia (repair). Each feature was scored: 0 (no lesion), 1 (mild), 2 (moderate), 3 (severe). Viral antigens in lung sections were visualized by immunohistochemical staining (Dietert et al., 2017). Briefly, slides were deparaffinized and IAV antigens were retrieved by boiling in citrate buffer for 12 min in a microwave oven. Slides were first incubated with goat polyclonal anti-IAV H1N1 antibody (dilution 1:250; Bio-Rad Laboratories; 5315-0064) for 1 h, and then with alkaline phosphatase-conjugated rabbit anti-goat antibody (1:200, Vector Laboratories) for 30 min at RT. The alkaline chromogen triamino-tritolyl-methanechloride (neufuchsin) served as phosphatase substrate for color development. Sections were counterstained with hematoxylin, dehydrated through alcohol washes of increasing concentrations, cleared in xylene and mounted on coverslips. Incubation with a purified goat antibody at the same dilution instead of primary antibody served as a negative control.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Histopathology of mouse lungs

Lung sections were evaluated by microscopy in a blinded fashion. The pathological features assessed were scored: 0 (no lesion), 1 (mild), 2 (moderate), 3 (severe).

### Statistical analysis

Statistical tests were performed using GraphPad/ Prism. Statistical details of experiments can be found in the figure legends. Exact values of n and statistical tests are indicated in figure legends. P values are indicated in Figures. Data are presented as mean ± SD. P values of plaque diameters were calculated using one-way ANOVA Bonferroni's multiple comparison test (*** $P < 0.001$).

**Cell Reports, Volume *31***

## Supplemental Information

# Mechanism of Virus Attenuation by Codon Pair Deoptimization

**Nicole Groenke, Jakob Trimpert, Sophie Merz, Andelé M. Conradie, Emanuel Wyler, Hongwei Zhang, Orsalia-Georgia Hazapis, Sebastian Rausch, Markus Landthaler, Nikolaus Osterrieder, and Dusan Kunec**

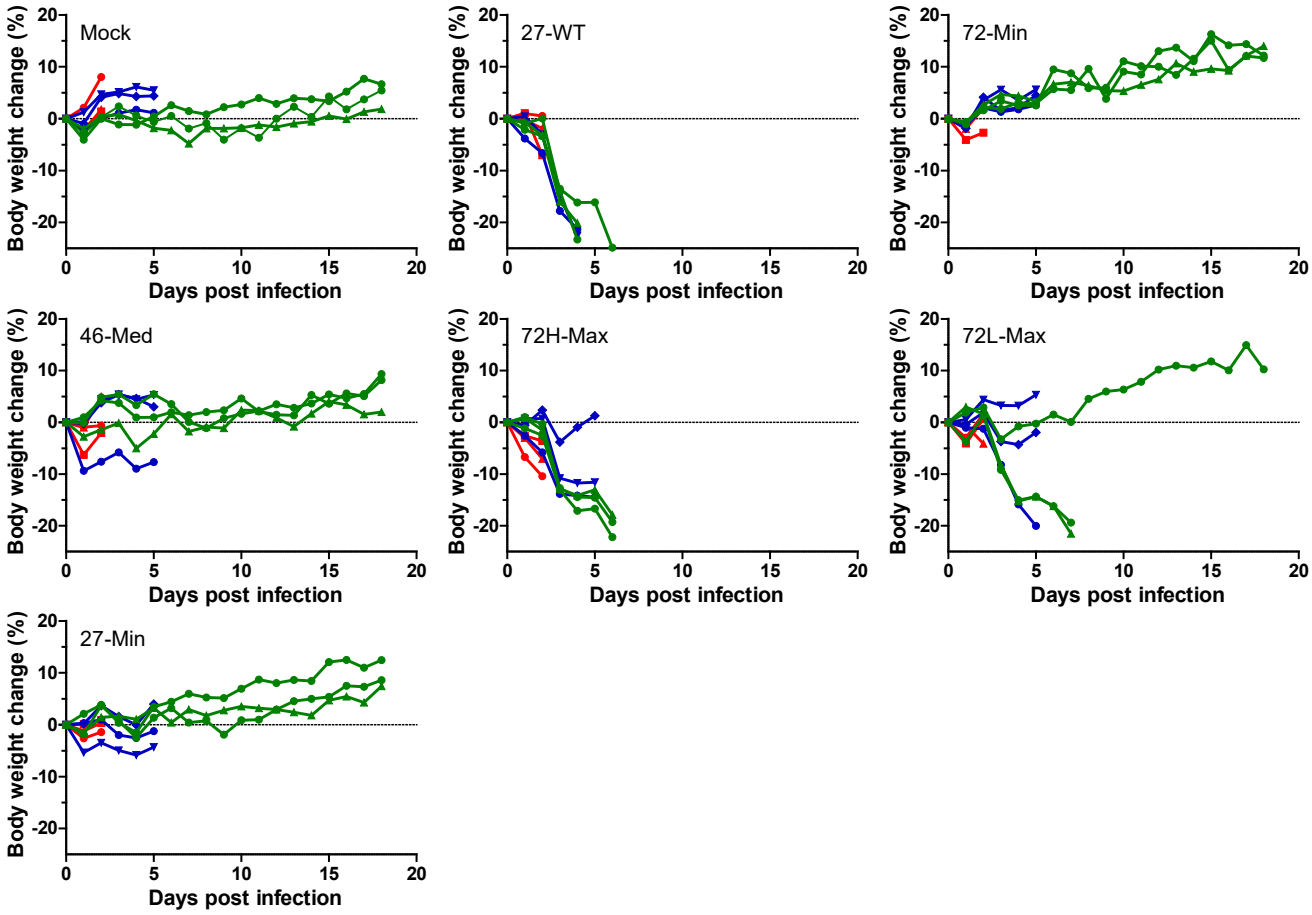



**Figure S1. Survival and body weight changes of individual mice after infection with IAV, related to Figure 3A.** Nine 6-week-old BALB/c mice were infected with the parental or recoded IAV or DMEM as control. On day 2 (red), 5 (blue) and 18 (green), three animals of each group were euthanized. Mice in moribund conditions, or those which lost more than 20 % body weight were euthanized earlier than scheduled.

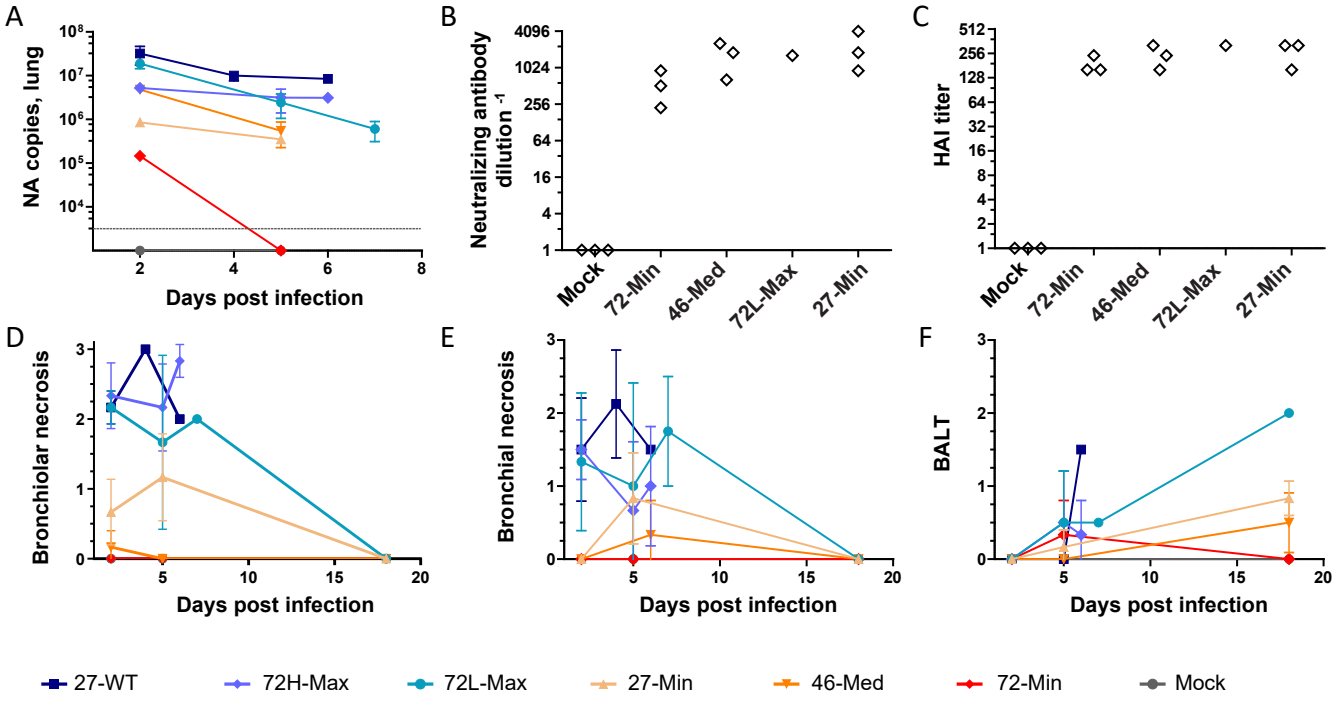



**Figure S2. Histopathological changes in the lungs of infected mice, related to Figure 3.** Nine 6-week-old BALB/c mice were infected with different recoded influenza viruses. On day 2, 5 and 18 three animals were euthanized or earlier if they lost more than 20 % body weight. (**A**) The right lung lobes were homogenized and copy numbers of NA segment were determined. Data are represented as mean ± SD. No virus could be detected in samples on day 18. (**B**) Anti-IAV neutralizing antibody titers were measured and (**C**) an HAI test was performed at 18 dpi with serum of mice. (**D-F**) Left lung lobes were embedded in paraffin, cut into sections, and lesions were scored from 0 (no lesion) to 3 (severe lesions). 27-WT, 72H-Max and 72L-Max show marked (**D**) bronchiolar necrosis, (**E**) bronchial necrosis and (**F**) mild increase in bronchus-associated lymphoid tissue (BALT), 72-Min, 46-Med and 27-Min only moderate or none. Data are represented as mean ± SD.

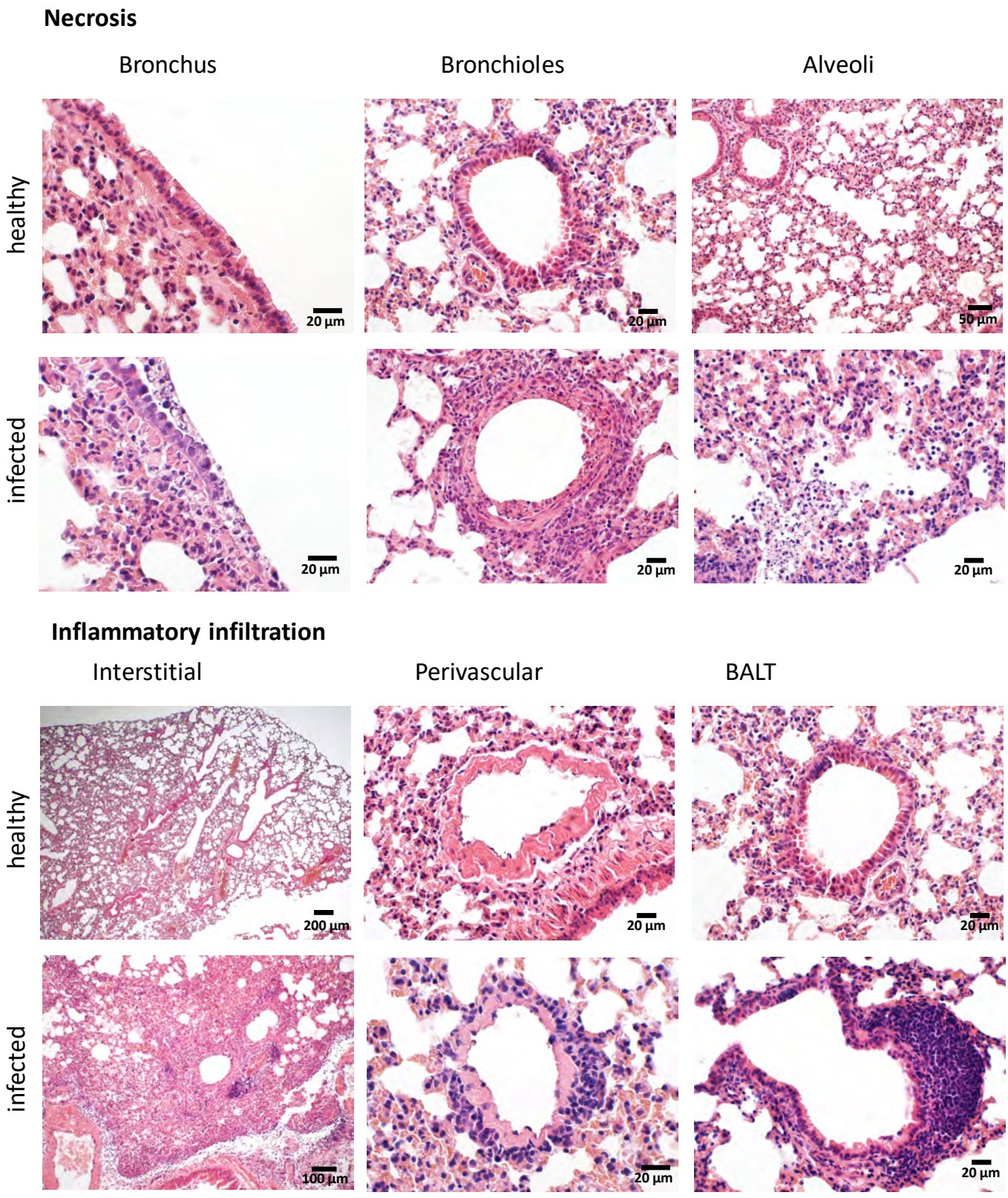



**Figure S3. Representative images of histopathological changes that were assessed in the lungs of infected mice, related to Figure 3.** To assess the distribution and character of pathologic changes in the lungs of infected mice, lung sections were stained with hematoxylin and eosin and microscopically evaluated for necrosis (bronchi, bronchioli and alveoli) and inflammation (perivascular, interstitial, bronchus-associated lymphoid tissue (BALT)). Representative tissue sections of healthy and infected lungs are shown.

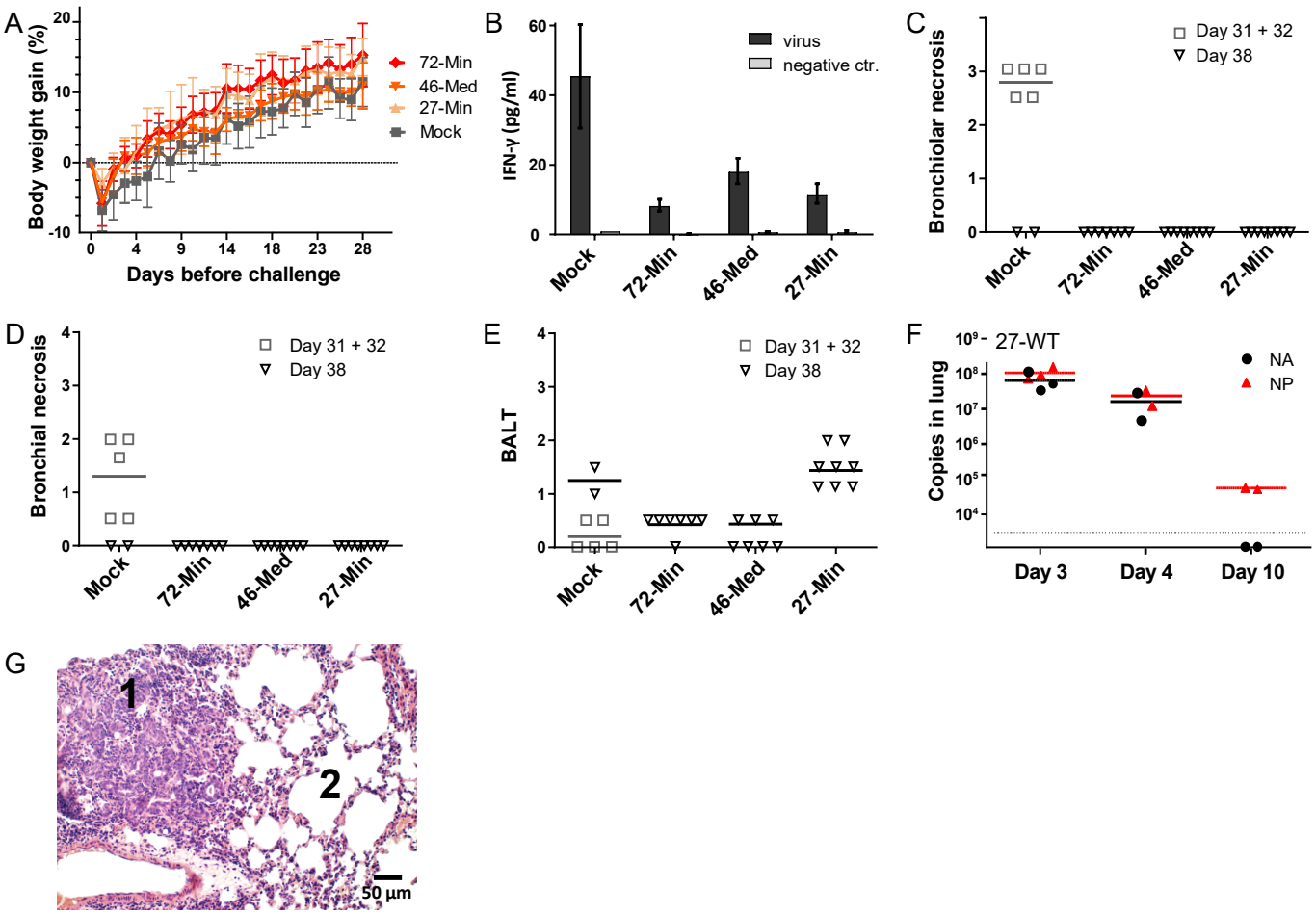



**Figure S4. Histopathological assessment of the severity and distribution of pneumonic lesions in the lungs of immunized mice, related to Figure 4.** Eight six-week-old BALB/c mice were vaccinated with attenuated viruses and one control group was inoculated with DMEM. (**A**) All groups showed similar development of body weight after vaccination. Data are represented as mean ± SD. (**B**) Production of IFN-γ by splenocytes. Mouse splenocytes were isolated 10 days after challenge, stimulated with purified IAV virus or control protein for 3 days, and production of IFN-γ was measured by ELISA. Data are represented as mean ± SD. (**C-E**) Left lung lobes were embedded in paraffin, cut into sections and lesions were scored from 0 (no lesions) to 3 (severe lesions). Shown are (**C**) bronchiolar necrosis, (**D**) bronchial necrosis and (**E**) bronchus-associated lymphoid tissue. (**F**) Number of NA and NP copies of mock-vaccinated mice are shown after challenge. (**G**) Image of pneumocyte type II hyperplasia (repair) of with hematoxylin and eosin stained in paraffin embedded lung. The left side of the image shows affected (1) and the right side unaffected tissue (2).

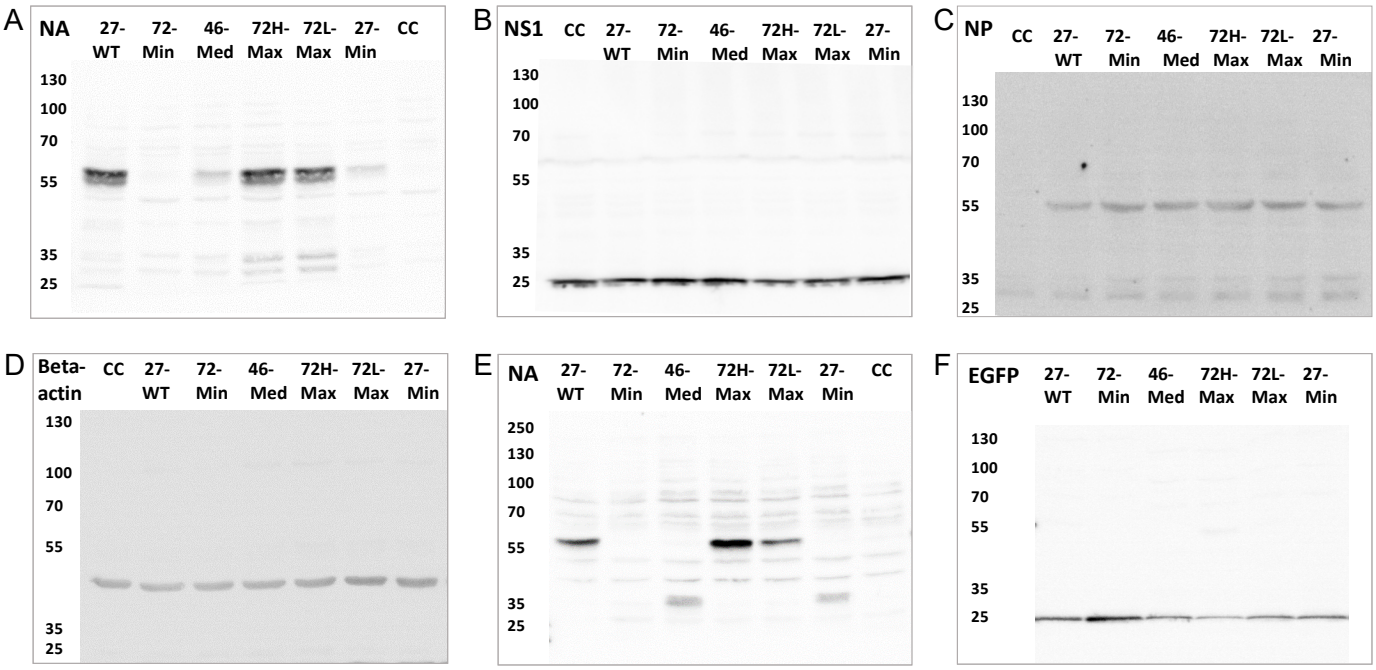



**Figure S5. Uncropped images of western blots shown in the manuscript, related to Figure 2 and Figure 5.** (**A-D**) Images of western blots from Figure 2F. MDCK cells were infected with the parental and recoded viruses for 5 h and the production of NA (**A**) NS1 (**B**), NP (**C**) and beta-actin (**D**) proteins was analyzed by western blotting. (**E, F**) Images of western blots from Figure 5B. HEK 293T cells were transfected with expression plasmids pVITRO2-EGFP-NA for 24 h and the production of NA (**E**) and EGFP (**F**) proteins was detected by western blotting.

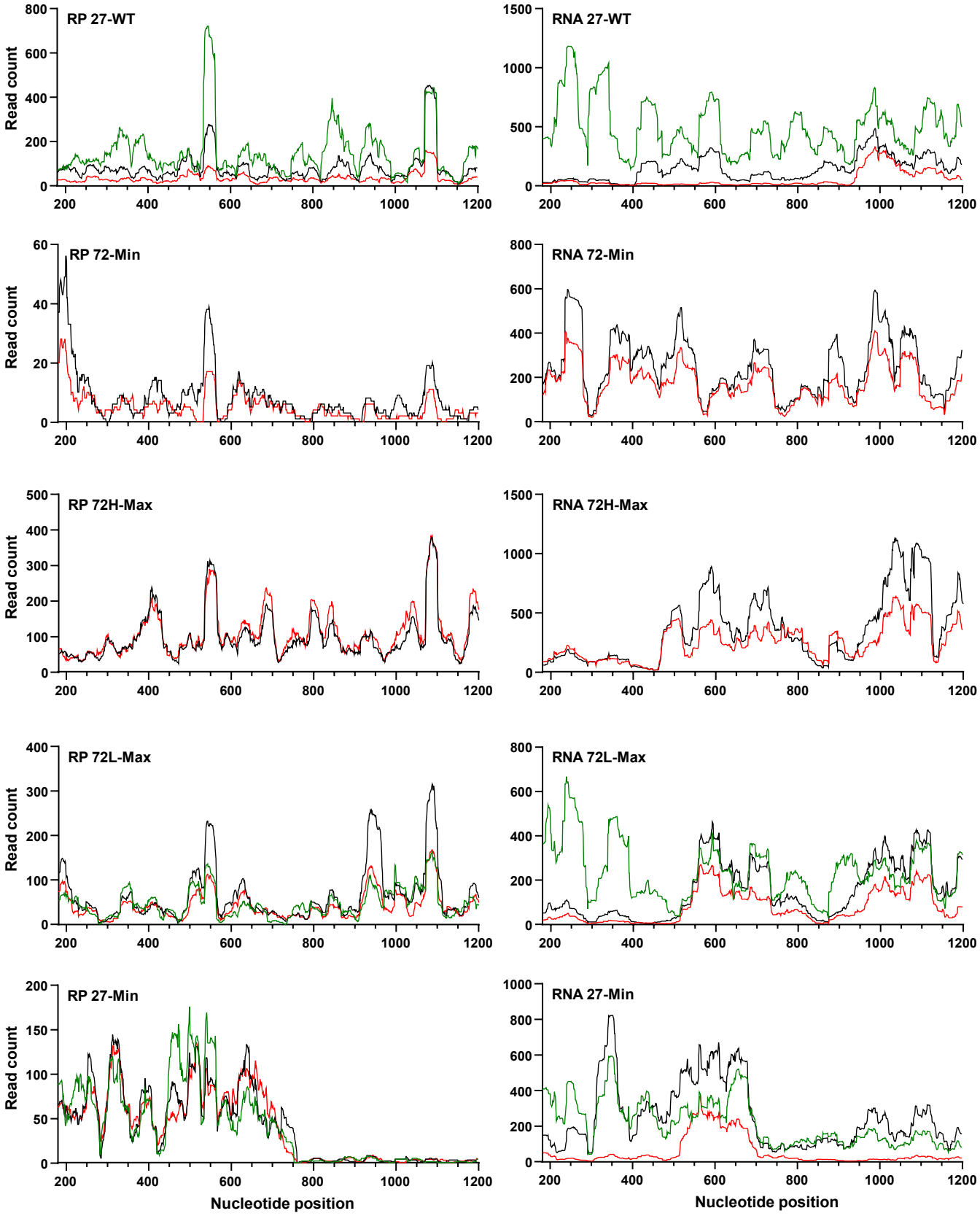



**Figure S6. Read density of ribosome footprints and total RNA by next generation sequencing from recoded parts of NA fragments in transiently transfected cells, related to Figure 5.** Red, black and green read densities represent replicates.

**Table S1. Primers and probes for RT-qPCR, related to Figure 5 and STAR Methods**

| **Name** | **Sequence 5'-3'** |
|---|---|
| NA 3' end forward | CTGGACTAGTGGGAGCATCA |
| NA 3' end reverse | ATGGTGAACGGCAACTCAG |
| NA 3' end probe | 6-FAM-CACCGTCTGGCCAAGACCAATC-TAMRA |
| NA 5' end forward | TGGGTCAATCTGTATGGTAGTC |
| NA 5' end reverse | GCTGCCTTGGTTGCATATT |
| NA 5' end probe | 6-FAM-TGGATTAGCCATTCAATTCAAACCGGA-Tamra |
| EGFP forward | CCACAAGTTCAGCGTGTCC |
| EGFP reverse | GAACTTCAGGGTCAGCTTGC |
| EGFP probe | 6-FAM-TGGCATCGCCCTCGCCCTCG-TAMRA |
| NP forward | CCAAAGGCACCAAACGATCT |
| NP reverse | AGTGGCATTCTGGCGTTCTC |
| NP probe | 6-FAM-ACGAACAGATGGAGACTG-TAMRA |