







ARTICLE

<https://doi.org/10.1038/s41467-019-13520-8>

OPEN

# The dynamic proteome of influenza A virus infection identifies M segment splicing as a host range determinant

Boris Bogdanow<sup>1,2,7</sup>, Xi Wang <sup>1,8</sup>, Katrin Eichelbaum<sup>1</sup>, Anne Sadewasser <sup>2</sup>, Immanuel Husic<sup>1</sup>, Katharina Paki<sup>2</sup>, Matthias Budt<sup>2</sup>, Martha Hergeselle<sup>1</sup>, Barbara Vetter<sup>3</sup>, Jingyi Hou<sup>1</sup>, Wei Chen <sup>1,4</sup>, Lüder Wiebusch <sup>3</sup>, Irmtraud M. Meyer<sup>1,5</sup>, Thorsten Wolff <sup>2</sup> & Matthias Selbach <sup>1,6\*</sup>

Pandemic influenza A virus (IAV) outbreaks occur when strains from animal reservoirs acquire the ability to infect and spread among humans. The molecular basis of this species barrier is incompletely understood. Here we combine metabolic pulse labeling and quantitative proteomics to monitor protein synthesis upon infection of human cells with a human- and a bird-adapted IAV strain and observe striking differences in viral protein synthesis. Most importantly, the matrix protein M1 is inefficiently produced by the bird-adapted strain. We show that impaired production of M1 from bird-adapted strains is caused by increased splicing of the M segment RNA to alternative isoforms. Strain-specific M segment splicing is controlled by the 3' splice site and functionally important for permissive infection. *In silico* and biochemical evidence shows that avian-adapted M segments have evolved different conserved RNA structure features than human-adapted sequences. Thus, we identify M segment RNA splicing as a viral host range determinant.

<sup>1</sup>Max Delbrück Center for Molecular Medicine, Robert-Rössle-Strasse 10, 13125 Berlin, Germany. <sup>2</sup>Unit 17 "Influenza and other Respiratory Viruses", Robert Koch Institut, Seestrass 10, 13353 Berlin, Germany. <sup>3</sup>Labor für Pädiatrische Molekularbiologie, Charité Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany. <sup>4</sup>Department of Biology, Southern University of Science and Technology, Xueyuan Road 1088, 518055 Shenzhen, China. <sup>5</sup>Freie Universität Berlin, Department of Biology, Chemistry, Pharmacy Institute of Chemistry and Biochemistry, Thielallee 63, 14195 Berlin, Germany. <sup>6</sup>Charité Universitätsmedizin Berlin, 10117 Berlin, Germany. <sup>7</sup>Present address: Structural Interactomics, Leibniz-Forschungsinstitut für Molekulare Pharmakologie, Robert-Rössle-Strasse 10, 13125 Berlin, Germany. <sup>8</sup>Present address: Division of Theoretical Systems Biology, German Cancer Research Center, 69120 Heidelberg, Germany. \*email: [matthias.selbach@mdc-berlin.de](mailto:matthias.selbach@mdc-berlin.de)

Influenza A viruses (IAVs) are negative-sense, single-stranded RNA viruses with a segmented genome. IAV infection causes seasonal epidemics and sporadically pandemic outbreaks in the human population with significant morbidity, mortality, and economic burden. IAVs can infect both mammals (e.g., humans, pigs, horses) and birds (e.g., chicken, waterfowl). However, strains that are replicating in birds typically do not infect mammals and vice versa. Pandemics occur when influenza strains of avian origin with novel antigenicity acquire the ability to transmit among humans<sup>1</sup>. Understanding the molecular basis of host specificity is therefore of high medical relevance.

The species barriers that hinder most avian IAVs from successfully infecting humans are effective at several steps in the viral life cycle. For example, the avian virus receptor hemagglutinin (HA) recognizes oligosaccharides containing terminal sialic acid (SA) that are linked to galactose by  $\alpha$ 2,3<sup>2</sup>. In the human upper respiratory airway epithelium, the dominant linkage is of  $\alpha$ 2,6 type, to which human-adapted HA binds. Despite these differences in receptor binding, many avian viruses are internalized by human cells and initiate expression of the viral genome. Such infections typically lead to an abortive, nonproductive outcome in mammalian cell lines. Our understanding of this intracellular restriction is still incomplete. One well-established factor is the influenza RNA-dependent RNA polymerase (RdRP): this enzyme catalyzes replication of the viral genome and transcription of viral messenger RNAs (mRNAs)<sup>3</sup>. Polymerases from avian strains are considerably less active in mammalian cells than their counterparts from mammalian-adapted strains<sup>4</sup>. A wealth of experimental data described adaptive mutations that alter receptor specificity or fusion activity of HA (reviewed in ref. <sup>5</sup>) and polymerase activity (reviewed in ref. <sup>6</sup>). However, relatively little is known about the contribution of other IAV genes for permissive vs. non-permissive infection<sup>7</sup>.

A crucial aspect for permissive infection is the correct timing of viral gene expression: IAV proteins are produced at the specific phase of infection when they are needed<sup>8</sup>. One example is the M gene, which encodes predominantly two polypeptides: the larger protein, M1, is produced from a collinear transcript. The smaller one, M2, is encoded by a differentially spliced transcript<sup>9</sup>. M1 is the matrix protein with multiple functions that encapsulates the viral genome, and also mediates nuclear export<sup>10,11</sup>. M2 is a proton-selective channel that is an integral part of the viral envelope<sup>12,13</sup>. The ratio of spliced to unspliced products increases during infection<sup>14</sup>, which reflects the changing demands required for optimal viral replication.

Systems-level approaches have provided important insights into the molecular details of host–virus interaction<sup>15</sup>. For example, RNA interference (RNAi) screens identified host factors required for IAV replication<sup>16–18</sup>. Also, interaction proteomics experiments identified many cellular binding partners of IAV proteins<sup>19–21</sup>. A number of studies also quantified changes in protein abundance<sup>22–27</sup>. However, these steady-state measurements cannot reveal the dynamic changes in protein synthesis during different phases of infection. Early studies used radioactive pulse labeling to monitor protein synthesis in IAV-infected cells<sup>28,29</sup>. However, radioactive pulse labeling cannot provide kinetic profiles for individual proteins. More recently, stable isotope labeling by amino acids in cell culture (SILAC) emerged as a powerful means to study the dynamic proteome<sup>30</sup>. SILAC-based pulse labeling methods such as pulse SILAC (pSILAC) and dynamic SILAC can quantify protein synthesis and degradation on a proteome-wide scale<sup>31,32</sup>. Moreover, metabolic incorporation of bioorthogonal amino acids such as azidohomoalanine (AHA) provides a means to biochemically enrich for newly synthesized proteins<sup>33</sup>. In combination with SILAC, AHA labeling can be used to quantify proteome dynamics with high temporal resolution<sup>34–36</sup>.

Here, we use metabolic pulse labeling and quantitative mass spectrometry to compare proteome dynamics upon infection of human cells with a human-adapted and a bird-adapted IAV strain. We find that host proteins behave surprisingly similar, but observe striking differences in the production of viral proteins, especially for the matrix protein M1. Follow-up experiments with reporter constructs, in silico studies, and reverse genetics identify an evolutionarily conserved *cis*-regulatory element in the M segment as a host range determinant.

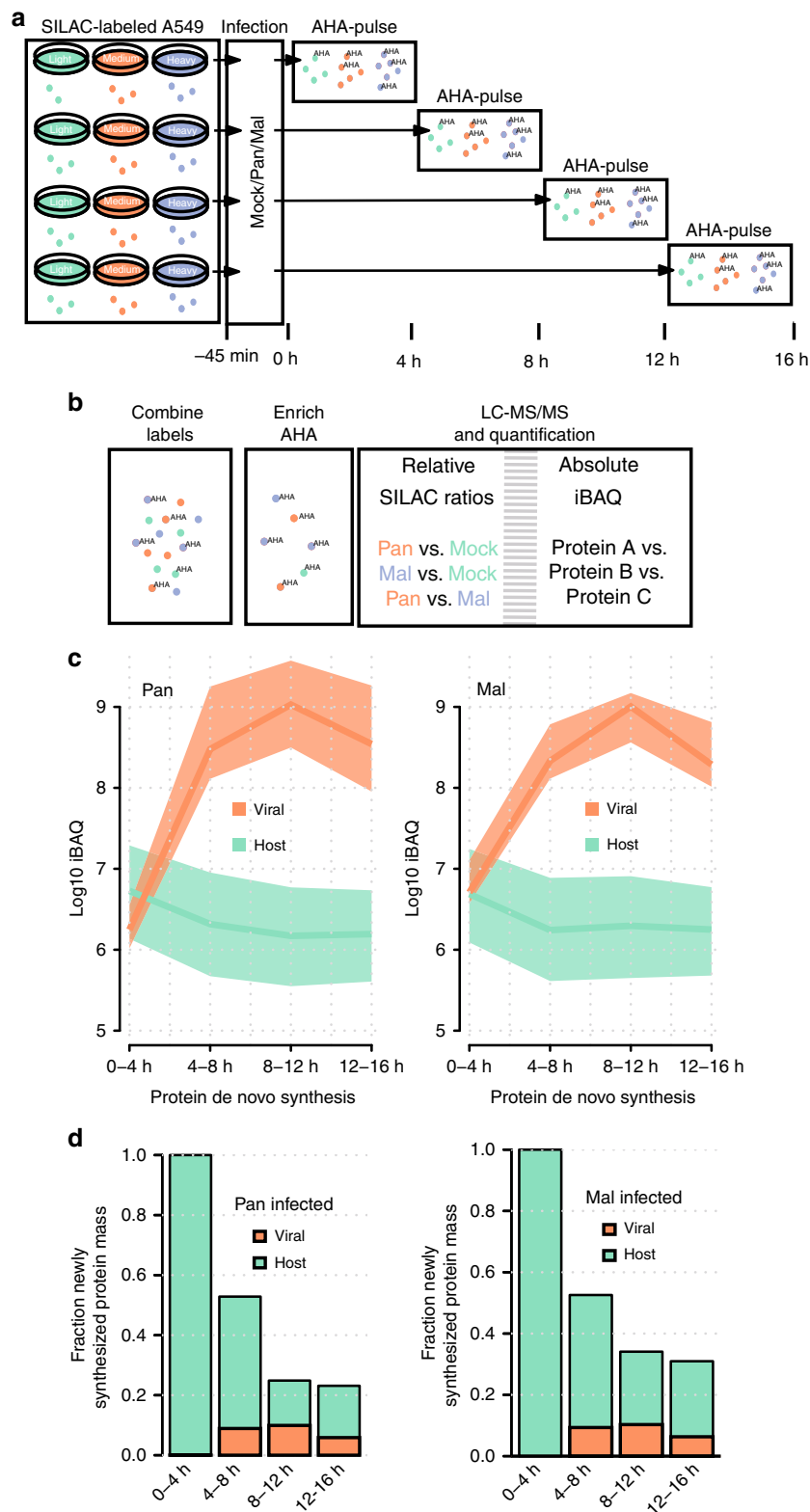
## Results

**Quantifying the dynamic proteome of IAV infection.** To assess species specificity of IAVs, we used a model system comparing a low-pathogenic avian H3N2 IAV (A/Mallard/439/2004—Mal) to a seasonal human IAV isolate of the same subtype (A/Panama/2007/1999—Pan). While the avian virus is not adapted to efficient growth in cultured human cells and causes a non-permissive infection, the seasonal human virus replicates efficiently. We demonstrated previously that the Pan virus produces >1000-fold more infectious viral progeny than the non-adapted virus, even though both strains efficiently enter human cells and initiate their gene expression program<sup>27</sup>.

We reasoned that comparing the kinetics of protein synthesis upon infection with both strains might reveal determinants of species specificity. To this end, we performed proteome-wide comparative pulse-labeling experiments by combining labeling with AHA and SILAC (Fig. 1a): cells incorporate AHA instead of methionine into newly synthesized proteins when the cell culture medium is supplemented with this bioorthogonal amino acid. AHA contains an azido group, which can be used to covalently couple AHA-containing proteins to alkyne beads via click chemistry. In this manner, newly synthesized proteins can be selectively enriched from the total cellular proteome. Combining AHA labeling with SILAC reveals the kinetics of protein synthesis with high temporal resolution<sup>36</sup>. First, we fully labeled human lung adenocarcinoma cells (A549) using SILAC. Second, individual cell populations were infected with either Pan or Mal virus or left uninfected. Third, all cells were pulse labeled with AHA for 4 h during different time intervals post infection (0–4, 4–8, 8–12, and 12–16 h). The three cell populations for every time interval were then lysed, combined, and AHA-containing proteins were enriched from the mixed lysate using click chemistry (Fig. 1b). After on-bead digestion, peptide samples were analyzed by high-resolution shotgun proteomics.

We quantified proteins using two readouts: (i) SILAC-based relative quantification to assess differences in *de novo* protein synthesis and (ii) intensity-based absolute quantification (iBAQ) to quantify absolute amounts of newly synthesized proteins<sup>37</sup>. Our data thus provides kinetic profiles for relative and absolute differences in *de novo* protein synthesis across the course of infection (Supplementary Data 1). In total, we identified 7729 host and 10 viral proteins and quantified 6029 proteins in at least two biological replicates with overall good reproducibility (Supplementary Fig. 1).

**The dynamic host proteome.** It is well established that IAV induces a global reduction in the production of host proteins. This host shutoff was attributed to a plethora of viral effector functions<sup>38</sup>. To assess the host shutoff in our proteomic data, we investigated iBAQ values for viral and host proteins. As expected, viral proteins were potently induced, while the production of host proteins decreased over time (Fig. 1c). The difference between host and viral protein synthesis reached several orders of magnitude and was highest during the 8–12 h pulse interval. Moreover, the total cellular protein output dropped to ~24% (Pan) or



**Fig. 1 A strategy to quantify protein de novo synthesis proteome-wide.** **a** SILAC-L/M/H-labeled A549 cells were infected with the human seasonal H3N2 IAV isolate Panama (Pan) or the avian H3N2 isolate Mallard (Mal) or left uninfected. The methionine analogous AHA was given to the methionine-depleted medium in different 4 h intervals. **b** After lysis and enrichment for proteins that incorporated AHA, samples were subjected to shotgun proteomics. Absolute and relative protein synthesis profiles were quantified for host and viral proteins. **c** iBAQ-based quantification of protein synthesis levels for host and viral proteins in cells infected with either Pan or Mal virus, as indicated. Median, 25th and 75th percentile of the respective populations are given. **d** Quantification of the total newly synthesized protein mass for host and viral proteins with either Pan or Mal infection as indicated. Data were normalized to the 0–4 h time period. All data are based on the mean of  $n = 2$  biological replicates.

~33% (Mal) at later stages of infection, of which ~20–40% was of viral origin (Fig. 1d). At this level of detail, we observed no major differences between both strains. Thus, both strains initiate viral protein synthesis and induce the shutoff of host protein synthesis to an overall similar extent.

Next, we investigated the profiles of individual host proteins across the course of infection. For this, we directly looked at SILAC ratios comparing infected and non-infected cells (Fig. 2a, b). As expected, synthesis of the vast majority of host proteins markedly decreased over time. However, some proteins were less affected by the host shutoff and displayed only a mildly decreased or even increased production. To assess this observation more systematically, we selected the proteins that were least affected by the shutoff at different pulse periods and performed gene ontology (GO) analysis. The heatmap of enriched GO terms provides a global overview of biological processes as the infection progresses (Fig. 2c and Supplementary Data 2). For example, many well-known interferon-induced antiviral defense proteins (e.g., MX1, several IFIT proteins, several oligoadenylate synthase proteins) were relatively strongly produced at late stages of infection. Also, many ribosomal proteins (GO term “peptide chain elongation”) largely escaped the host shutoff. Interestingly, we also observed significant enrichment of proteins involved in steroid metabolism and mitochondrial proteins (mito-ribosomal, respiratory chain proteins) at early and intermediate stages of infection, respectively. Cellular responses to infection with the Pan and Mal strain were overall similar. To assess potential differences between permissive and non-permissive infection, we compared protein log<sub>2</sub> fold changes between both viruses directly (Fig. 2d). Interestingly, type I interferon response proteins were first preferentially produced during non-permissive infection. At later stages, however, infection with the Pan virus elicited a stronger interferon response.

Several different hypotheses were made to explain the IAV-induced host shutoff. This included mechanisms at the transcriptional<sup>39</sup>, post-transcriptional<sup>40,41</sup>, and translational<sup>42</sup> level. To study the relationship of mRNA and protein levels, we quantified mRNA levels at 8 h post infection by RNA-sequencing (RNA-seq). mRNA level differences at this time point showed good correlation with corresponding differences in de novo protein synthesis, particularly during the subsequent 8–12 h period (Fig. 2e–g). Thus, mRNA level changes play an important role for the shutoff of individual mRNAs, corroborating the view that the host shutoff is mainly due to reduced host mRNA levels<sup>41</sup>. The correlation was higher for the Mal than for the Pan strain, suggesting that infection with the latter strain involves additional post-transcriptional processes.

Traditionally, IAV is thought to prioritize the translation of viral over host mRNAs<sup>42</sup>, but more recent experimental and computational analyses challenge this view<sup>41,43</sup>. We investigated this question by calculating protein synthesis efficiencies (i.e., the amount of protein made per mRNA). To this end, we divided iBAQ values by corresponding RPKM (Reads Per Kilobase Million) values (Fig. 2h). Infection with both strains reduced host protein synthesis efficiencies compared to uninfected controls. Importantly, we did not observe preferential translation of viral transcripts. This suggests that mRNAs from human and avian influenza virus strains access the translational machinery with comparable efficiency, which argues against the idea that modulation of translation efficiency affects species specificity.

**Dysregulated synthesis of viral proteins.** Since the observed differences in host protein synthesis were surprisingly subtle, we focused our attention to the dynamics of viral protein synthesis. Production of most viral proteins peaked in the 8–12 h period

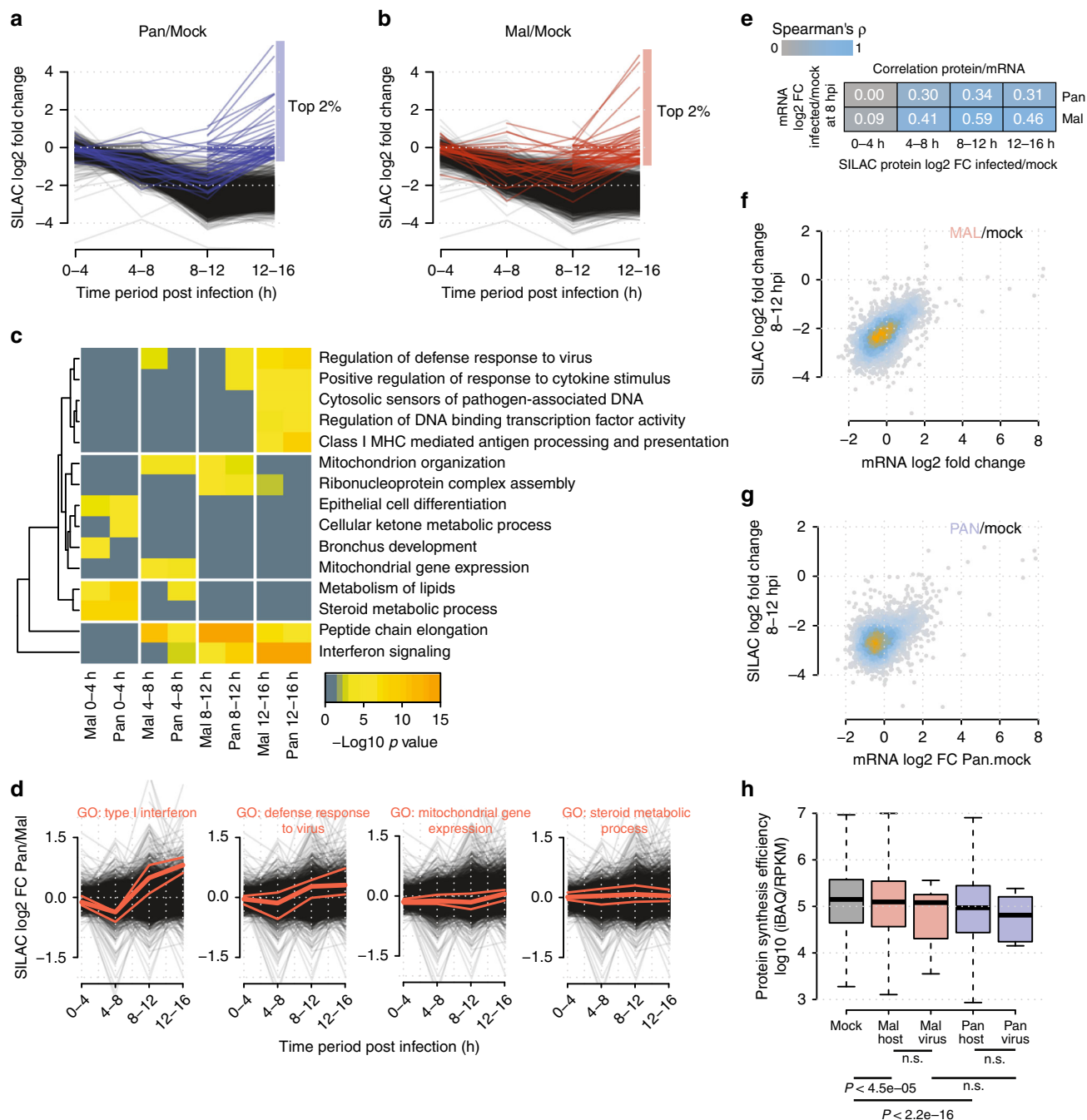
(see Supplementary Fig. 2). The kinetics such as the early production of NS1 and NP and delayed synthesis of M1 is consistent with classical radioactive pulse labeling experiments<sup>29</sup>. We then used SILAC ratios of shared peptides (i.e., peptides with sequence identity between both strains) to precisely compare the kinetics of viral protein synthesis (Fig. 3a, b). We found that the avian strain produced higher amounts of all viral proteins at the beginning, confirming that the Mal virus successfully enters cells and initiates its gene expression program. Later on, during mid to late phases, the human Pan virus produced most proteins more abundantly than the avian strain. Note that NS1 and M2 are excluded in this analysis because no identical peptides were identified.

It is well established that the RdRP from avian-adapted IAV strains is less active in mammalian cells<sup>4,44</sup>. Thus, we would have expected the production of all viral proteins in the bird-adapted strain to be reduced to a similar extent. In contrast, we observed striking differences in the synthesis of individual proteins: HA was more abundantly produced by the avian strain throughout infection. In contrast, neuraminidase (NA) and particularly matrix protein M1 were stronger produced by the human strain at later stages. These differences in the production of individual viral proteins cannot be explained by the global difference in RdRP activity between strains. Thus, the avian strain displays dysregulated protein production relative to its human counterpart.

We focused our attention on the M1 protein since it showed the largest difference between both strains. The protein is highly conserved between Pan and Mal (~96% amino acid identity) and the most abundant protein in virions<sup>45</sup>. Moreover, M1 is known to mediate export of the viral genome across the nuclear membrane—an essential step during permissive infection<sup>10,11</sup>. Thus, accumulation of M1 at late stages of infection is required for the appearance of viral ribonucleoproteins (vRNPs) in the cytoplasm of infected cells. Interestingly, when investigating the subcellular distribution of the viral nucleoprotein (NP) by immunofluorescence microscopy, we observed efficient export during infection with the Pan strain (Fig. 3c). In contrast, NP was inefficiently exported and accumulated in the nucleus upon Mal infection. These microscopy data is also corroborated by the increased interferon response induced by the Pan strain at later stages of infection (Fig. 2d), which is stimulated by cytosolic viral RNA sensors<sup>46</sup>. We conclude that non-permissive infection correlates with reduced M1 production and impaired nuclear export of NP.

**Increased M segment RNA splicing.** We next sought to investigate the mechanism for the impaired M1 production. To this end, we first quantified the levels of viral mRNAs from our RNA-seq data. In total, the avian virus produced ~2/3 of the mRNA of the human strain with the single largest difference observed for M1 (Fig. 4a). The strain-specific differences in M1 mRNA levels were very similar to the observed differences in M1 protein production (Fig. 4b). Hence, the impaired M1 protein production during non-permissive infection can largely be explained by reduced M1 mRNA levels.

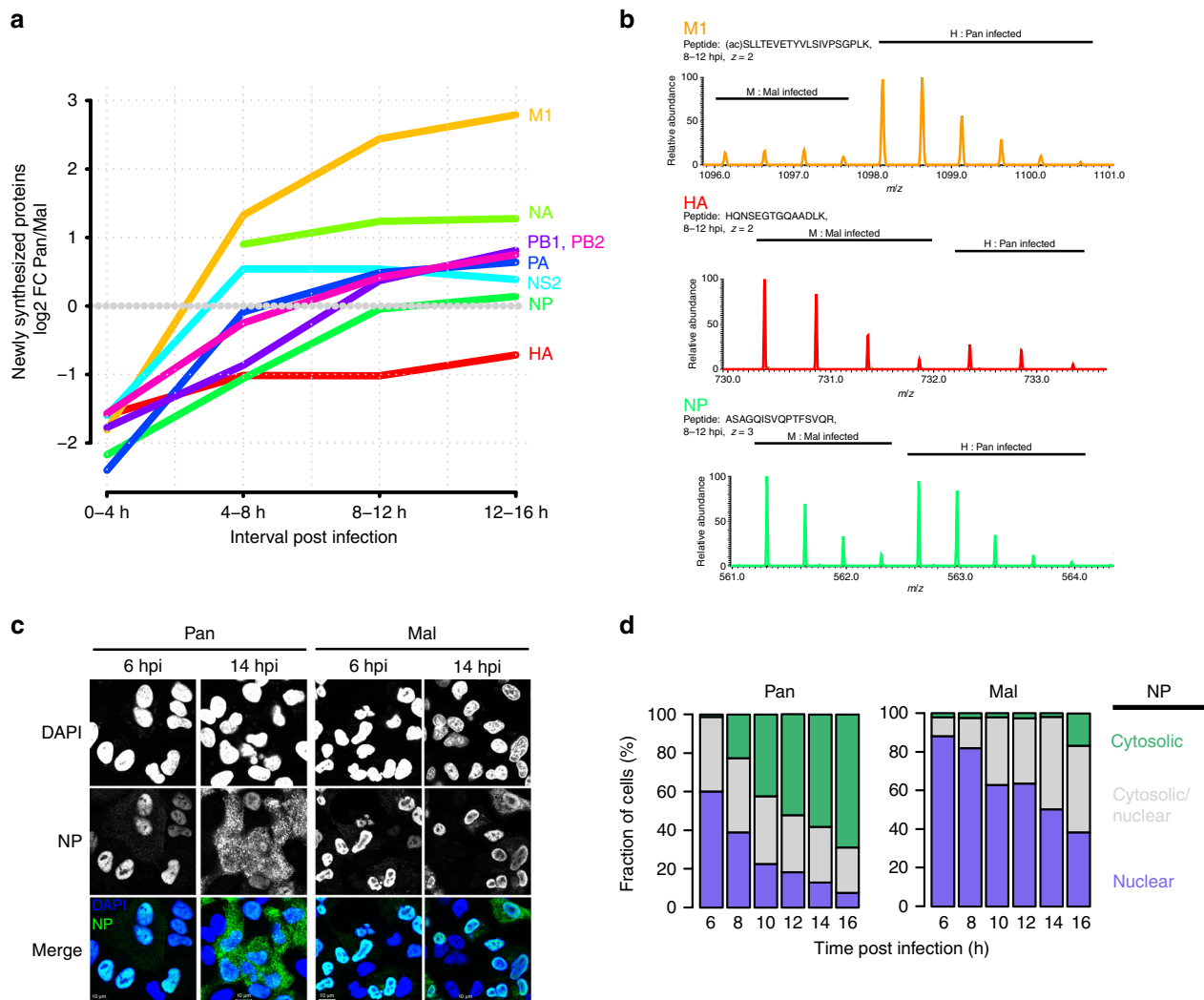
M1 is encoded on segment 7 (i.e., the M segment), which is the most conserved segment between Pan and Mal (~89% nucleotide identity). The M1 protein is produced from a collinear transcript that can be alternatively spliced into three additional isoforms, which all use a common 3' splice site:<sup>47</sup> the M2 mRNA, which encodes the ion channel M2<sup>12</sup>, RNA 3, which is not known to encode a peptide, and M4 mRNA, which is proposed to be translated to an isoform of the M2 ion channel in certain strains<sup>48</sup>. We investigated the relative proportion of these



**Fig. 2** The dynamic host proteome of IAV-infected cells. **a, b** SILAC profiles of host proteins infected with Pan (**a**) or Mal (**b**) compared to mock infection. The top 2% of proteins that are strongest produced during the last time period are highlighted. **c** Functional protein clusters least affected by host shutoff were identified. The 2% proteins with the highest ratios were selected for each time period and multi-list GO enrichment was performed. **d** Median, 25th, and 75th percentile of proteins related to a selected set of GO terms are highlighted in the direct SILAC comparison Pan/Mal. All other protein profiles in black. **e** Spearman's rank correlation coefficients when comparing the indicated RNA and protein level data. **f, g** Scatterplot of mRNA changes vs. protein synthesis level changes in Pan/mock or Mal/mock infection conditions. Blue and orange coloring of data points reflects the density of data points. **h** Protein synthesis efficiencies calculated from absolute mRNA (RPKM, 8 hpi) and protein data (IBAQ, 8-12 hpi) (center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range, outliers (data points outside the whiskers) were removed for visibility). For the assessment of statistical significance two-sided Wilcoxon's rank-sum tests were performed and *p* values are given (ns: non-significant). All data are based on the mean of *n* = 2 biological replicates.

isoforms in the RNA-seq data via splice junction reads. We detected all known isoforms plus a yet undescribed transcript of the avian M segment, which we call RNA 5. This transcript results from splicing at 5' donor GG site (pos 520/521) and the common 3' acceptor site and contains an open reading frame in-frame with M1 with a missing internal region (Fig. 4c).

While only a few percent (2–3%) of the M1 mRNA was alternatively spliced during permissive infection, ~27% was spliced upon infection with the avian strain (Fig. 4d). Thus, the reduced level of M1 mRNA during non-permissive infection is at least partially due to increased splicing of the M1 mRNA to alternative isoforms. In contrast, M2 mRNA levels were rather

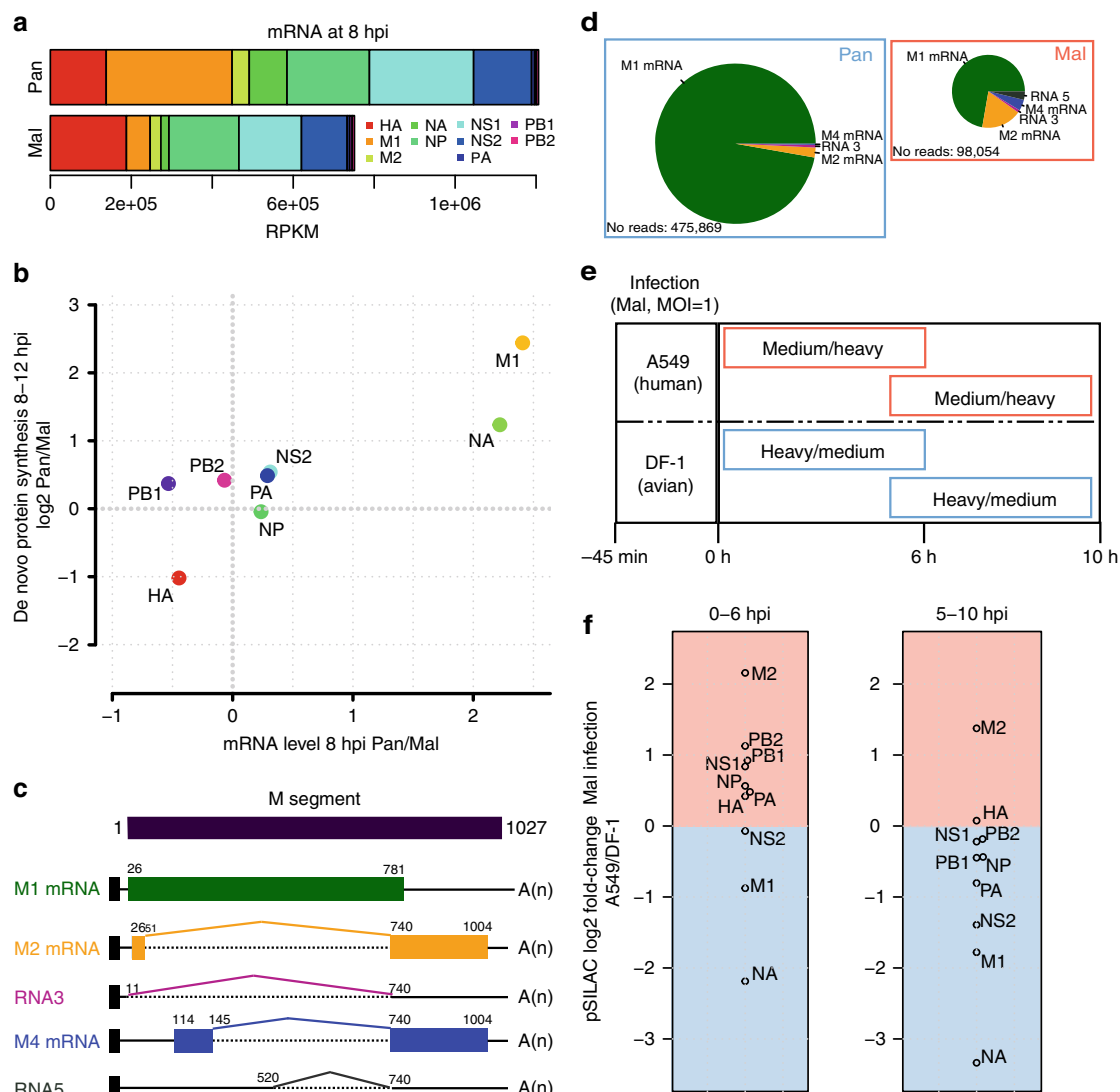


**Fig. 3** Viral protein synthesis is dysregulated during non-permissive infection. **a** SILAC protein synthesis profiles of human vs. avian-adapted IAV proteins based on shared peptides. No shared peptides were identified for NS1 and M2. The average of  $n = 2$  biological replicates is shown. **b** MS1 spectra of individual precursor (M1—top, HA—middle, NP—lower) peptides in M and H SILAC channels. **c** Cells were infected with the indicated viruses at MOI = 1 (FFU/cell) and analyzed by immunofluorescence microscopy for NP trafficking. Nuclei were counterstained with DAPI. Scale bars: 10  $\mu$ m. **d** For quantification, NP staining pattern was categorized as predominantly cytosolic, nuclear or both. At least 140 cells were counted per condition.

similar. We note that the comparable M2 mRNA level during non-permissive infection results from two opposing processes—the increased splicing of the primary transcript to the M2 mRNA, and the global reduction in viral transcripts, which is probably due to the impaired polymerase activity<sup>4,44</sup>. We conclude that M segment RNA splicing is markedly different in permissive vs. non-permissive infection.

In principle, the increased splicing of the M segment RNA by bird-adapted strains may reflect an evolutionary adaptation to different needs of the corresponding viral proteins in avian cells. In this case, avian-adapted strains would produce less M1 in both human and avian cells. Alternatively, the optimal balance of viral proteins could be constant and independent of the host species. In this case, avian-adapted strains would produce more M1 in avian than in human cells. To investigate this experimentally, we infected both human and chicken cells with the Mal virus and monitored the production of viral proteins using pSILAC (Fig. 4e). We found that the Mal strain produced considerably more M1 in chicken cells than in human cells (Fig. 4f). Hence, the reduced M1 production of avian viruses in human cells appears to reflect poor adaptation to the mammalian splicing environment.

**A cis-regulatory element controls M segment RNA splicing.** The differences in M1 mRNA splicing can be due to (i) *cis*-regulatory elements (i.e., specific signals encoded in the M segment), (ii) *trans*-acting factors (i.e., other viral or host factors that interact with M1 mRNA), or (iii) a combination of both. To assess whether *cis*-regulatory elements are involved, we sought to investigate M segment RNA splicing outside the context of infection. We therefore designed a splicing reporter system (Fig. 5a, b). To this end, we cloned the coding region of the M segment (nucleotides 29–1007) into a eukaryotic expression vector and fused it to an N-terminal Flag/HA tag. Importantly, this construct avoids the strong 5' splice site of RNA 3<sup>49</sup> and enabled us to assess the relative levels of M1 to M2 proteins. When we transfected human A549 cells with these reporter constructs, we found that M2 was produced to high levels with the construct containing the Mal M sequence, but was barely detectable when the Pan M sequence was transfected (Fig. 5c). Thus, our reporter system recapitulates splicing differences observed during infection. We conclude that *cis*-regulatory elements in the M segment cause excessive splicing of the avian variant.



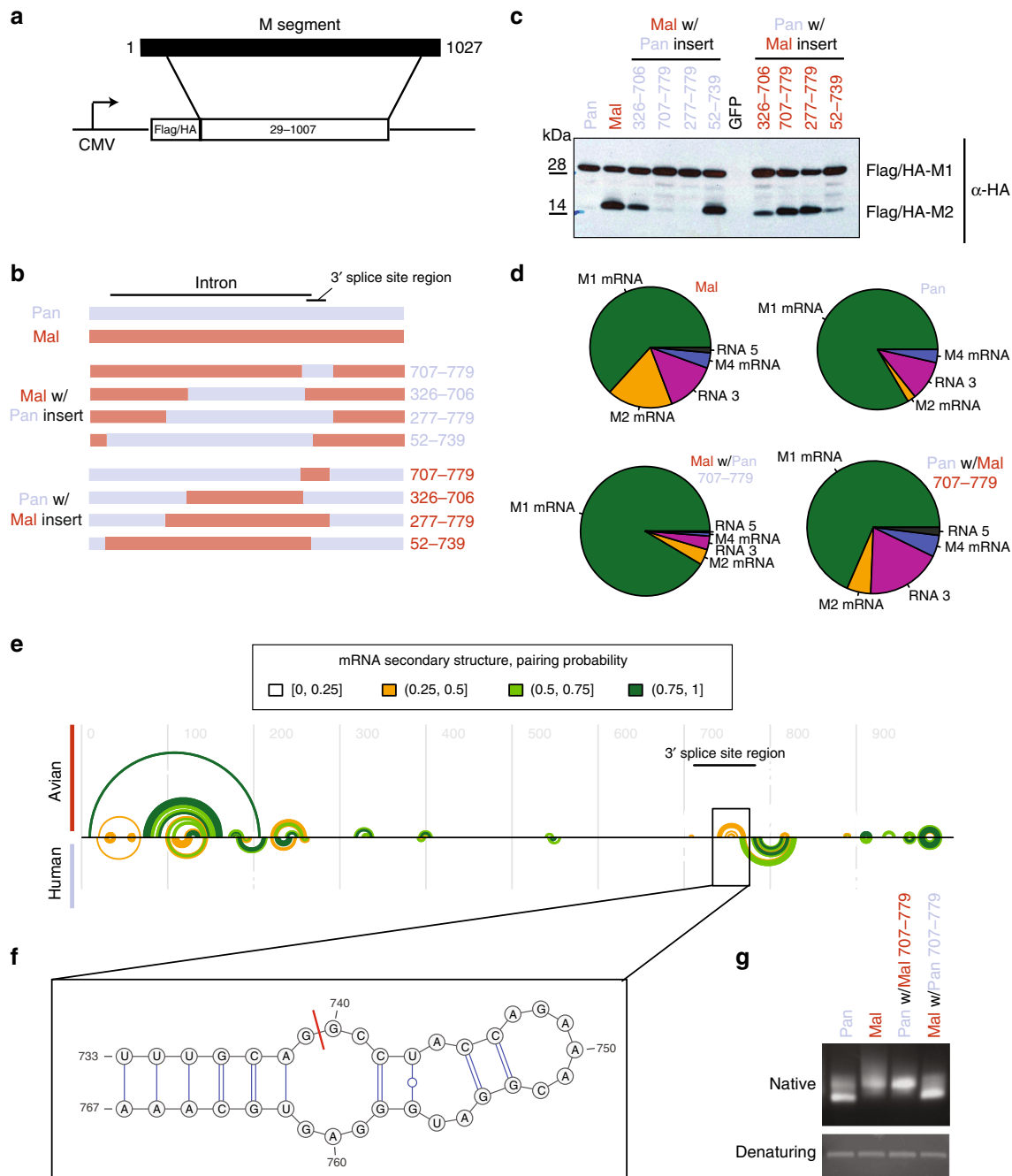
**Fig. 4 Differences in M segment RNA splicing.** **a** Normalized read counts (RPKM, based on splice junction reads and exonic reads) for 10 viral mRNAs from poly-A-enriched and either Pan- or Mal-infected samples at 8 hpi. The sum of  $n = 2$  biological replicates is depicted. **b** Direct comparison of fold changes for viral mRNAs/proteins from RNA level and protein synthesis data. **c** Schematic depiction of the M gene architecture. The M pre-mRNA can be spliced into the indicated isoforms. **d** Relative quantification of the different isoforms based exclusively on splice junction reads from RNA-seq data for both strains at 8 hpi. The area of the pies reflects the absolute number of splice junction reads. The average of  $n = 2$  biological replicates is depicted. **e** SILAC-light-labeled DF-1 and A549 cells were pulse labeled with heavy or medium heavy SILAC amino acids (in label-swap duplicates) after infection with the Mal virus (MOI = 1 PFU/cell) and subjected to quantitative shotgun proteomics. **f** pSILAC fold-changes are depicted for both pulse intervals comparing the production of viral proteins with the Mal strain in human vs. avian cells. The average of  $n = 2$  biological replicates is depicted.

To determine the sequence responsible for the strain-specific splicing, we made chimeric reporter constructs (Fig. 5b). When swapping the entire intron sequence of the M2 splice variant (nucleotides 52–739, corresponding to ~70% of the coding sequence), we did not observe major changes in the relative amount of M1 to M2. In contrast, integrating the human 3' splice site region (nucleotides 707–779, 73 nucleotides) into the avian construct strongly impaired splicing down to the levels of the human wild-type construct. Conversely, when we integrated the avian 3' splice site region into the human construct, we observed a strong increase in splicing, similar to the avian wild-type construct (Fig. 5c). To address splicing in the context of the viral RdRP, we co-transfected HEK293T cells with expression vectors for PB1, PB2, PA, and NP (of WSN strain) and Pan and Mal M wild-type and chimeric segments. In this system, the accumulation of M segment-derived mRNAs depends primarily on the activity of the RdRP. RNA-seq revealed consistent differences in

abundance of the various isoforms with similar overall mRNA levels (Fig. 5d). Thus, the splice site region alone is sufficient to switch the species-specific splicing phenotype in reporter systems using both viral and host RNA polymerases. Interestingly, this region has been reported to contain an RNA secondary structure<sup>50</sup> and a binding site for the splicing factor SRSF1<sup>51</sup>. We conclude that a *cis*-regulatory element in the splice site region determines the strain-specific splicing pattern.

**Evolutionary conserved RNA structure elements in M segments.**

We next wanted to assess whether our findings are also relevant for other human- or bird-adapted IAVs. Specifically, we sought to identify functionally relevant RNA secondary structures that have been conserved during evolution of avian- and human-adapted IAVs. To this end, we analyzed multiple sequence alignments (MSAs) from hundreds of recent human H3N2 and avian isolates using the RNA structure prediction program RNA-Decoder<sup>52</sup>



**Fig. 5** A *cis*-regulatory element in the 3' splice site region controls M segment splicing. **a** Reporter system. The coding sequence of segment 7 was cloned into a eukaryotic expression vector with N-terminal Flag/HA tag. **b** Wild-type Pan and Mal sequences as well as several chimeric Pan/Mal constructs were cloned into the reporter vector. **c** A549 cells were transfected with the indicated reporter constructs, harvested, and subjected to anti-HA immunoblotting. **d** HEK293T cells were transfected with constructs encoding NP, PA, PB1, and PB2 (of strain WSN) and vectors encoding for wild-type or chimeric M segments. At 24 h post transfection, RNA was isolated and subjected to RNA-seq. The pies depict the relative counts of isoforms as quantified by splice junction reads normalized to the levels of NP and polymerase subunits. The area of the pie scales with the total number of normalized splice junction reads. The average of  $n = 2$  biological replicates is depicted. **e** Evolutionary conserved RNA secondary structure along the M mRNA predicted for avian-adapted (top) and human-adapted H3N2 (bottom) IAV strains. Each semi-circle corresponds to 1 bp involving the corresponding two alignment positions. Color coding of base pairs according to their corresponding, estimated base-pairing probability. The 3' splice site region (nt 707–779) is annotated with a bar. **f** The insert shows a detail of the predicted RNA secondary structure at the 3' splice site, predicted for avian but not human sequences in this study. The splice position is indicated by a red bar. **g** The region predicted to be structured with the two species-specific and mutually exclusive elements (707–825) was in vitro transcribed. The RNA was purified and run on native and denaturing agarose gels stained with ethidium bromide.

(see Methods). This program is capable of dis-entangling overlapping evolutionary constraints due to encoded amino acids and RNA structure features and has been shown to successfully identify evolutionarily conserved RNA structures overlapping protein-

coding regions, for example, in viral genomes such as hepatitis C and HIV<sup>52,53</sup>. Importantly, RNA-Decoder captures evidence on conserved RNA structure based on the evolutionary signals encoded in the sequences of the input alignment. This is a key



advantage over computational methods that identify RNA structures based on their thermodynamic stability *in vitro*, as these methods assume that the RNA has no interactions with other molecules (e.g., proteins and other RNAs) *in vivo*. Also, RNA-Decoder employs a probabilistic framework, which is capable of estimating the reliability of its predictions.

The RNA secondary structure that is best supported by the evolutionary signals in the two MSAs (the so-called maximum-likelihood structure) markedly differs between human and avian strains, particularly in the region around the 3' splice site (Fig. 5e): the avian region encodes a hairpin-like structure (Fig. 5f) overlapping the 3' splice site, which is absent from the human-adapted sequences. This structure is similar to a hairpin reported by Moss et al.<sup>50</sup> for four sequences, but differs in details (Supplementary Fig. 3). Most importantly, the evolutionarily conserved structure reported here leaves the GC motif immediately downstream of the AG consensus at the 3' splice site unpaired, making it potentially more accessible to splicing. In human-adapted sequences, this region is involved in a different conserved secondary structure (Fig. 5e) that is mutually exclusive with this hairpin. We conclude that the M segment of avian and human-adapted isolates contain evolutionarily conserved RNA secondary structures that markedly differ in exactly the region that is critical for strain-specific splicing. Importantly, the consensus sequences of avian and human isolates diverge by five nucleotides in the region of interest (pos: 745, 754, 760, 766, and 772), all of which are retained with the two strains that we investigated here (Supplementary Fig. 4A). Thus, the Pan and Mal strains are by and large representative for human and avian strains in this region.

In addition to these computational analyses, we also wanted to assess RNA secondary structures experimentally. To this end, we *in vitro* transcribed the region of the M segment containing evolutionarily conserved secondary structure (nucleotides 707–825) from both strains, purified the RNA, and analyzed it on native and denaturing agarose gels (see Fig. 5g). We observed different migration behavior for Pan and Mal under native, but not denaturing conditions, consistent with different secondary structures (lanes 1 and 2). Importantly, when substituting eight nucleotides in the region 707–779 (termed 3' splice site region in our manuscript) from Mal into the Pan backbone (lane 3), we observed the same migration pattern as for the Mal wild-type RNA. Conversely, when substituting eight nucleotides from the Pan into Mal backbone (lane 4), we observed the same migration pattern as for the Pan wild-type RNA.

To investigate the global significance of these findings, we extended our analyses to other IAV isolates. The M segment of the seasonal Pan strain originates from the M segment of the A/Brevig Mission/1/1918 (p1918) virus, which is at the evolutionary root of human strains and caused the 1918 “Spanish flu” pandemic<sup>54</sup>. Therefore, we cloned the M segment of p1918 into our reporter vector (Supplementary Fig. 4B). Again, we observed inefficient splicing of the p1918 M gene, consistent with our data for the Pan strain and previous reports<sup>55</sup>. Moreover, integration the Mal 3' splice site region into the p1918 gene increased splicing. Thus, inefficient splicing of the M gene in human-adapted IAVs occurs in a seasonal (Pan) and a pandemic (p1918) strain. The M segment of the Mal strain originates from the eastern avian lineage. A/chicken/Rostock/45/1934 (H7N1), an early representative strain of this lineage, depicted even stronger M segment splicing than the Mal strain (Supplementary Fig. 4C, lane 3). Introducing the human-adapted mutations at the splice site region into this construct completely reversed the splicing phenotype (lane 4). Also, the “avian-like” M segment of A/swine/Netherlands/25/1980 (H1N1) showed a splicing phenotype similar to the Mal strain (lane 2). Finally, the human zoonotic

strain A/Vietnam/1203/2004 (H5N1) exhibited a splicing pattern similar to the Pan strain (lane 5). Collectively, these findings support the global significance of the strain-specific splicing patterns with respect to host range.

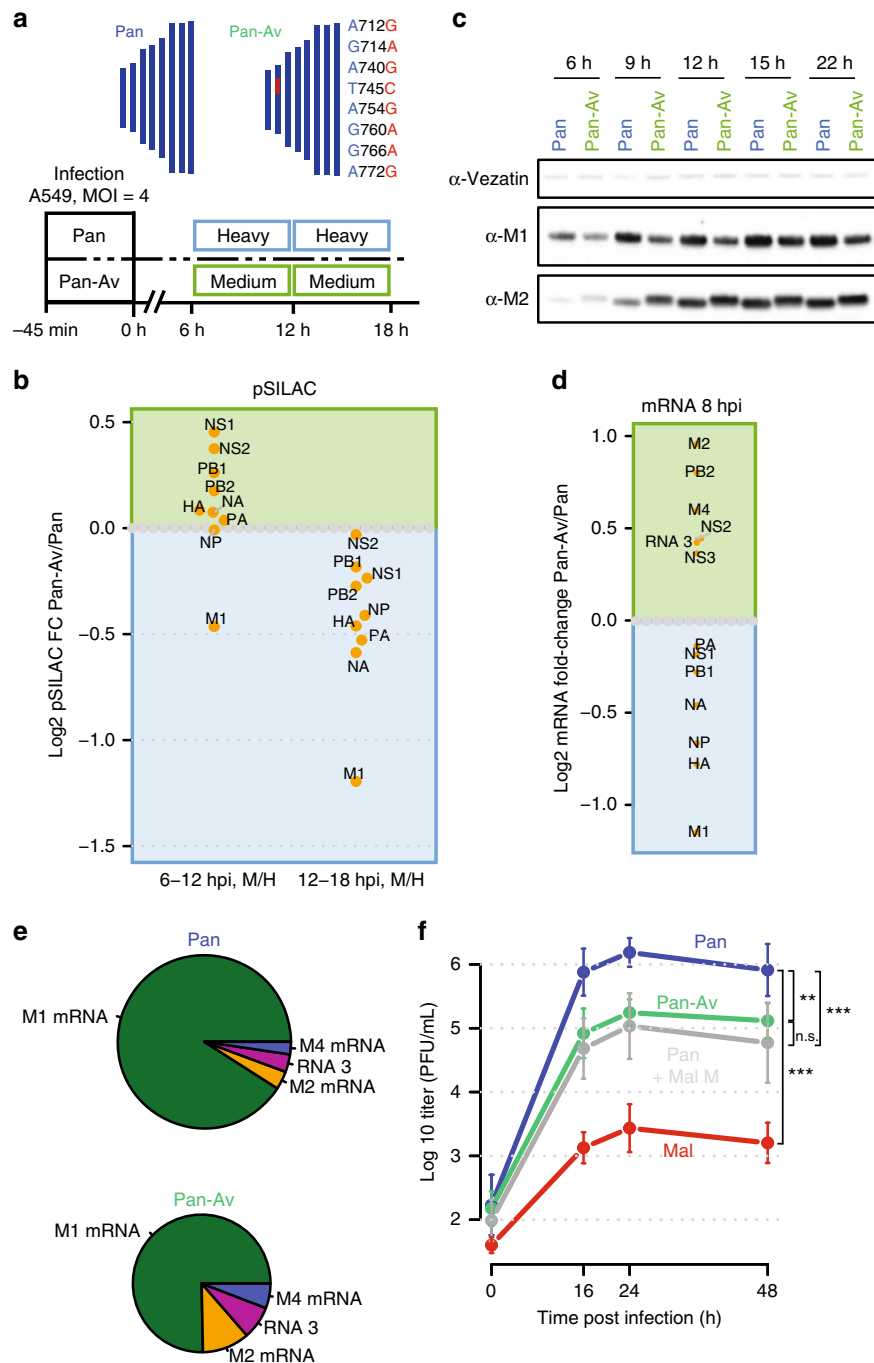
**The 3' splice site region is a host range determinant.** The experiments with reporter constructs described above are advantageous because they allow us to study the impact of M segment sequence features in isolation. Nevertheless, it is also important to assess the relevance of these findings during infection. We therefore mutated eight nucleotides in the splice site region of the Pan wild-type strain to the corresponding nucleotides in the Mal strain using reverse genetics (“Pan-Av” for a Pan strain with an avian splice site region, see Fig. 6a and Supplementary Data 3).

We first compared the kinetics of viral protein synthesis upon infection of A549 cells with both strains using pSILAC<sup>31</sup>. M1 synthesis was selectively impaired during Pan-Av infection during both the 6–12 and 12–18 hours post infection (hpi) time intervals. At later stages, the Pan-Av strain also showed impaired production of other essential viral proteins (Fig. 6b), suggesting that viral replication is also impaired. Next, we quantified M1 and M2 protein (Fig. 6c) and mRNA levels (Fig. 6d, e). The Pan-Av strain displayed decreased M1 protein and mRNA levels, mimicking the behavior of the Mal strain (compare to Fig. 4).

To assess the impact of M segment splicing on IAV replication in human cells, we assessed the growth characteristics of the different viruses (Fig. 6f). As expected, the Pan strain reached ~1000-fold higher titers than the Mal strain. Exchanging the entire M segment of the Pan strain with the M segment of the Mal strain (Pan + Mal M) reduced titers about 10-fold. Importantly, a similar ~10-fold attenuation was also seen in the Pan-Av strain, which only differs from the Pan strain by eight nucleotides. We conclude that the 3' splice site of the IAV M segment is indeed an important host range determinant.

## Discussion

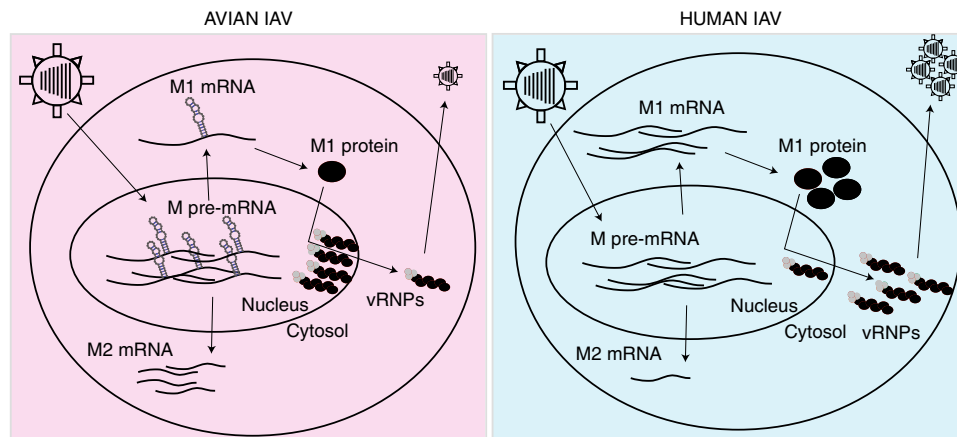
Advances in high-throughput sequencing have provided insights into the extraordinary diversity of viruses and their genomic determinants of host adaptation. However, the mechanism how these adaptive mutations enable replication in a given host is less understood. Our proteomic pulse labeling data allowed us to take an unbiased look at protein synthesis upon permissive and non-permissive infection. This is advantageous since it allows us to investigate changes in protein synthesis with high temporal resolution, which provides complementary information to more classical steady-state measurements<sup>22–27</sup>. We found that the synthesis profiles of host cell proteins were remarkably similar. Hence, the outcome of infection does not appear to depend on a specific host response. In contrast, we observed striking differences in the synthesis profiles of viral proteins. Particularly, the matrix protein M1 was inefficiently produced during non-permissive infection. Our follow-up experiments showed that this depends—at least partially—on excessive splicing of the avian M1 mRNA to alternative transcripts. Systematic computational analysis of the RNA structure of the M segment revealed characteristic and evolutionarily conserved differences in the splice site regions between human and bird-adapted strains. Exchanging eight nucleotides in the 3' splice site region from the human-adapted strain to corresponding sequences in the bird-adapted strain markedly impaired replication. Thus, our proteomic analysis of IAV infection identifies M segment splicing as a host range determinant. A hypothetical model for the influence of M segment splicing on the IAV host range is presented in Fig. 7.



**Fig. 6** The 3' splice site is a host range determinant. **a** (Top) Eight nucleotide polymorphisms of Mal virus at the M segment 3' splice site were integrated into wild-type Pan virus generating "Pan-Av" mutant. **a** (Bottom) Design of pulsed SILAC experiment. A549 cells were infected with wild-type or mutant virus at an MOI of 4 (FFU/cell) and viral gene expression was assessed by pulse labeling during two time intervals. **b** pSILAC-based quantification of viral protein production comparatively for Pan and Pan-Av virus. One replicate ( $n = 1$ ) is depicted. **c** Kinetics of M1/M2 protein levels as detected by immunoblotting. **d** RNA-seq-based quantification of viral mRNA levels at 8 hpi (MOI = 1 FFU/cell) comparing Pan and Pan-Av infection. The average of  $n = 2$  biological replicates is depicted. **e** Quantification of M segment-derived RNAs based on splice junction reads. The area of the pie scales with the total number of splice junction reads. The average of  $n = 2$  biological replicates is depicted. **f** Multicycle replication curve of the indicated viruses at an MOI of 0.05 FFU/cell on A549 cells. Means and standard deviations of biological triplicates ( $n = 3$ ) are shown along with significance estimates based on two-sided paired  $t$  tests for the 16–48 h time points (n.s.: non-significant,  $**p < 0.01$ ,  $***p < 0.001$ ).

The cell culture-based infection model and splicing reporter system employed here are advantageous because they enable experiments under well-controlled conditions. Having said this, it is important to keep in mind that these model systems do not represent the full complexity of IAV infections *in vivo*. Also, the Pan and Mal strain employed here do not represent the full

diversity of human- and bird-adapted IAVs. However, our computational analyses show that differences in RNA secondary structure of the 3' splice site are widely conserved in human- and bird-adapted IAVs. Moreover, recent results from the Steel lab show that avian M segments restrict growth and transmission of mammalian-adapted IAV strains in a guinea pig model<sup>56</sup>.



**Fig. 7 Hypothetical model for the role of M segment splicing for IAV host range.** A *cis*-regulatory secondary structure element (indicated hairpins) in avian but not human IAV M pre-mRNA facilitates splicing. This leads to the underproduction of M1 mRNA and protein in human cells infected with avian-adapted IAVs. The poor availability of the M1 protein may contribute to an impaired nuclear export of viral ribonucleoproteins (vRNPs).

Together, these findings strongly suggest that our results are also relevant outside the specific experimental model system employed here.

Our global assessment of protein *de novo* synthesis upon infection revealed a global reduction in overall protein output during infection with both strains, probably reflecting a global stress response. Also, we observed the well-known shutoff of host protein synthesis<sup>38,40,41,57</sup>. Specific classes such as interferon-related, ribosomal and mitochondrial proteins escaped the shutoff. We observed that the amount of protein synthesis upon infection primarily depends on mRNA levels. Thus, altered translation does not play a major role for the host shutoff, consistent with recent findings<sup>41</sup>. Surprisingly, we also found that viral transcripts were not more efficiently translated than host transcripts. This contrasts with early studies based on reporter systems<sup>57</sup>, but corroborates recent ribosome profiling data<sup>41</sup>. Our finding is also consistent with the fact that the codon usage of IAV genes is not optimized to reflect the codon usage of the host<sup>43</sup>. It is also interesting that the translation efficiency of the bird- and human-adapted strains was similarly poor. Thus, adaptation towards high translational efficiency does not seem to be required for crossing the species barrier.

Our unbiased proteomic analysis indicates that the differences between permissive and non-permissive infection depend on differences in viral rather than host protein synthesis. Hence, the orchestrated synthesis of the viral proteome appears to be critically important for permissive infection. This supports the emerging view that modulation of viral protein synthesis underpins host adaptation<sup>58</sup>. Specifically, we find that the strain-specific differences in M1 protein synthesis critically depend on a conserved *cis*-regulatory element, which controls M-segment mRNA splicing. M1 is particularly important for the nuclear export of the viral genome to the cytoplasm<sup>10,11</sup>. Consistently, we observed that the genome of the bird-adapted strain was inefficiently exported (Fig. 3c). We were not able to directly assess the role of M1 for nuclear export because the level of ectopically expressed M1 is much lower than the level reached during infection. Interestingly, we also observed differences in HA and NA expression between both strains (Fig. 3a), which may be relevant in this context.

We found that exchanging only eight nucleotides of the human-adapted M segment to the bird-adapted sequences markedly impaired viral replication. Hence, the *cis*-regulatory element described here plays an important role for host adaptation. However, it is critical to also emphasize that this is not the only relevant factor for IAV host range. For example, despite the

overall similar host response, we and others have previously described host factors affecting human and avian virus infections<sup>15–18,20,27</sup>. It is also well-established that the RdRP of avian-adapted strains is less active in human cells<sup>4,44</sup>. Moreover, differences in the binding specificity of viral HAs are known to play an important role for host adaptation<sup>5</sup>. Lastly, M-segment splicing does not only depend on *cis*-regulatory elements but also on *trans*-acting factors, such as NS1, RdRP, NS1-BP, or HNRNPK<sup>49,59–62</sup>. Indeed, while M1 production was clearly impaired in our mutant strain (Fig. 6b), the wild-type bird-adapted strain produced even less (Fig. 3a). Also, the increased production of avian M1 in chicken cells relative to human cells indicates that host factors play a role (Fig. 4f). It is therefore important to interpret our findings in the broader context of viral and host factors that jointly determine the success of IAV replication.

We are living in a pandemic era of IAV infections that began at around 1918<sup>63</sup>. At this time, a virus of ultimately avian origin acquired the ability to spread among humans and later on contributed its genetic material to other pandemic viruses until present. The M gene of this p1918 virus is in some regions similar to bird-adapted sequences, but shows important signatures of mammalian adaptation, especially at the 3' splice site region<sup>64</sup>. Our results suggest that mammalian adaptation at the 3' splice site was linked to modulating M segment RNA splicing, which may have been relevant for the emergence of the 1918 pandemics in humans.

## Methods

**Cells and viruses.** A549 (ATCC CCL-185) and HEK293T (ATCC CRL-11268) were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (v/v) fetal bovine serum, 2 mM L-glutamine, and antibiotics. Madin-Darby canine kidney (MDCK) type II cells (ATCC CRL-2936) were grown in minimal essential medium supplemented with 2 mM L-glutamine and 10% (v/v) fetal bovine serum. DF-1 (ATCC CRL-12203) were grown in DMEM supplemented with 10% (v/v) fetal bovine serum and 4 mM L-glutamine. All cells were maintained at 37 °C and 5% CO<sub>2</sub>. Stocks of the avian influenza virus A/Mallard/439/2004 (H3N2) (Mal) (GISAID accession numbers EPI859640–EPI859647) were grown in the allantoic cavities of 10-day-old embryonated chicken eggs for 2 days at 37 °C. A/Panama/2007/1999 (H3N2) (Pan) (NCBI accession numbers: DQ487333–DQ487340), Pan + Mal M reassortant and Pan-Av mutant virus were grown in MDCK type II cells. Virus stocks were titrated on MDCK type II cells by measuring plaque-forming units (PFUs) or fluorescence-forming units (FFUs). For the latter, cells were infected with different dilutions of virus stock for 5 h. Then, cells were harvested by trypsinization, fixed, and permeabilized by incubation in 75% ethanol for at least 12 h at 4 °C and stained with specific antibody against NP antigen (clone AA5H, Bio-Rad/Serotec, 1:1000, #MCA400). An Alexa Fluor 488-conjugated goat anti-mouse immunoglobulin G (IgG) antibody (Invitrogen) was

used as secondary reagent (1:1000). Cells were analyzed on a FACSCanto II flow cytometer (BD Biosciences) using the FACSDiva software package.

**Cloning and mutagenesis.** Gene fragments containing the coding sequence (nucleotides 29–1007) of segment 7 of A/chicken/Rostock/45/1934 (GenBank accession: CY077423), A/Vietnam/1203/2004 (GenBank accession: HM006762.1), A/swine/Netherlands/25/1980 (GenBank accession: Z26862.1), A/BrevigMission/1/1918 (GenBank accession: AY130766, “p1918”), a mutated A/BrevigMission/1/1918 (with the following point mutations: C718T, A725C, A754G, G760A, G766A, A772G, “p1918 w/ Mal 707–779”), and a mutated version of A/chicken/Rostock/45/1934 (with the following point mutations: G712A, A714G, G740A, C745T, G754A, A760G, “A/chicken/Rostock/45/1934 w/ Pan 707–779”) fused to attB1 and attB2 sites were ordered as synthetic double-stranded DNA fragments from Integrated DNA Technologies. The coding sequences of Pan and Mal M segments (nucleotides 29–1007) were amplified from complementary DNA (cDNA) and fused to attB1 and attB2 sites by PCR using the following primers:

fw\_Pan: GGGGACAAGTTTGTACAAAAAAGCAGGCTCCAGCTTCTTAAC  
CGAGGTCG

fw\_Mal: GGGGACAAGTTTGTACAAAAAAGCAGGCTCCAGCTTCTTAAC  
CGAGGTCG

rev\_Pan:  
GGGGACCACTTTGTACAAGAAAGCTGGGTTTACTCCAGCTTCTTAAC  
CAAAATG

rev\_Mal:  
GGGGACCACTTTGTACAAGAAAGCTGGGTTTACTCCAGCTTCTTAAC  
CAAAATG.

Cloning was done using Gateway Technology (Invitrogen) according to the manufacturer’s protocol. Entry clones were generated from pDONR221 vector, expression clones from pDEST26-Flag/HA destination vector. Pan/Mal chimeric constructs were generated by replacing appropriate restriction fragments of wild-type pDEST26-Flag/HA-Pan and pDEST26-Flag/HA-Mal, respectively, with the corresponding synthetic chimeric DNA inserts purchased as gBlocks from Integrated DNA Technologies.

A cDNA copy of the M segment of A/Mallard/439/2004 was amplified by RT-PCR using the following primers:

fw: CGAAGTTGGGGGAGCAAAAAGCAGGTAG  
rev: GGCCGCGGGTTATTAGTAAACAAGGTAG.

This was followed by insertion into pHW2000 resulting in pHW2000-Mal M. The following point mutations were introduced into pHW2000-Pan M plasmid: A712G, G714A, A740G, T745C, A754G, G760A, G766A, and A772G, creating pHW2000-Pan-AV. The pHW2000-Mal M plasmid was mutated at the same positions to the respective sequence of the Pan strain (G712A, A714G, G740A, C745T, G754A, A760G, A766G, G772A, “pHW200-Mal M w/ Pan 707–779”). Mutations were introduced using the Gibson Assembly Cloning Kit (NEB) with pDEST26-Flag/HA-Pan w/ Mal 707–779 and the wild-type pHW2000-Pan M or pDEST26-Flag/HA-Mal w/ Pan 707–779 and the wild-type pHW2000-Mal M as templates for PCR. The following primers were used to amplify the insert:

fw\_pDest: GAGGACTGCAGCGTAGACGCTTTG  
rev\_Malw/Pan\_pDest: CAATGATACTTGGGCAATAACGAGAGG  
rev\_Panw/Mal\_pDest: CAAGATCCCAATGATACTCGCAACA.

The following primers were used to amplify the backbone:

fw\_Pan\_pHW2000: GTTGTCTGCGAGTATCATTGGGATCTTG  
fw\_Mal\_pHW2000: CCTCTCGTTATTGCCGCAAGTATCATTG  
rev\_Mal/Pan\_pHW2000: CAAAGCGTCTACGCTGCAGTCTC.

All constructs were confirmed by cycle sequencing.

**Reverse genetics.** Recombinant IAVs derived from the A/Panama/2007/99 backbone were generated using an eight plasmid system for this strain based on pHW2000 by transfection of human HEK293T cells, followed by passage on MDCK type II cells. For the rescue of the Pan-Mal M reassortant virus, we used pHW2000-Mal M, whereas the Pan-AV virus was generated with pHW2000-Pan-AV together with seven plasmids encoding the other segments of human A/Panama/2007/99 virus.

**In vitro transcription.** First, the M segment region of interest (707–825) that was predicted to be structured was subcloned into pMCh1807\_pBS\_cdc42E7-boxB (Addgene: 118612) using the primers

Mal\_fw: CGATCTCGAGGCTGAAAGATGATCTTCTTG  
Pan\_fw: CGATGCTCGAGTCTAAGAGATGATCTTCTTG  
Mal/Pan Rev: GCTACGATCGAGTGCAAGATCCCAATG,

and fused to a T3 promoter sequence before being used as template for in vitro transcription with the following primers:

fw: CCAAGCTCGAAATTAACCTCACTAAAGG  
rev: GCTAGCTAGCAGTGCAAGATCCCAATG.

In vitro transcription was carried out using the MegaScript Kit (Life Technologies) according to the manufacturer’s instructions with 100 ng of template DNA and subsequent DNA removal. The RNA was purified using phenol/chloroform and resuspended in buffer containing 10 M Tris, 100 mM KCl, 10 mM MgCl<sub>2</sub>, and incubated at 37 °C for 30 min. To determine differences in RNA

secondary structure, RNA was run on native 1% agarose TAE (Tris-Acetate-EDTA) gels, and then supplemented with ethidiumbromide. To denature secondary structures, the RNA was mixed 1:5 with 65% formamide, 22% formalin, 13% 10× MOPS (3-(N-morpholino)propanesulfonic acid) buffer and incubated at 55 °C for 15 min. Then, samples were run on a denaturing MOPS agarose gel containing 3% formalin and ethidium bromide.

**pAHA-SILAC.** A549 cells were fully labeled in SILAC DMEM (PAA) supplemented with glutamine, 10% fetal bovine serum (Life Technologies) and 2 mM L-glutamine, antibiotics, and with either heavy (R10K8, “SILAC-H”), medium (R6K4, “SILAC-M”), or light (R0K0, “SILAC-L”) arginine and lysine (Cambridge Isotope Laboratories). Cells were cultured in SILAC-L/M/H medium for at least six passages. Ten-cm dishes of confluent light-labeled cells were mock-infected, while heavy- and medium-labeled cells were infected with either Pan or Mal strain at a multiplicity of infection (MOI) of 3 PFU (pore-forming unit). Virus was allowed to attach to the cells for 45 min on ice. Cells were washed with pre-warmed phosphate-buffered saline (PBS) before the infection medium was added (SILAC DMEM containing the respective SILAC AA, 0.2% bovine serum albumin (BSA), 2 mM glutamine, antibiotics). Prior to pulse labeling, cells were washed with pre-warmed PBS. Methionine-depleted infection medium additionally containing 100 μM L-AHA (4-azido-L-homoalanine; Anaspec) was added for different 4-h intervals to the cells. Cells were washed in PBS, scraped from the dish, and frozen until further sample processing. Lysis and enrichment for newly synthesized proteins was done using Click-It Protein Enrichment Kit (Invitrogen), with modifications: for the lysis, 283 μl of urea lysis buffer was used per label, samples were incubated for 20 min on ice, sonicated and cell debris was removed before SILAC label were mixed. Ten percent of sample was directly subjected to Wessel-Flügge precipitation and served as the input, and 90% were used for enrichment of newly synthesized proteins. Two hundred microliters of alkyne agarose resin was mixed with 800 μl combined cell lysate and then catalyst solution was added. The enrichment reactions were rotated head over tails for 18 h at room temperature. Enriched proteins were reduced and alkylated as indicated in the manufacturer’s instructions. Beads were then washed sequentially (each 5×) in sodium dodecyl sulfate (SDS) wash buffer (supplied with the kit), 8 M urea in 0.1 M Tris/HCl (pH 8.0), 80% acetonitrile in 0.1 M Tris/HCl pH 8.0, and 5% acetonitrile in 50 mM ammonium bicarbonate. Proteins were then digested in 5% acetonitrile/50 mM ammonium bicarbonate overnight using trypsin (Promega). Peptides were then acidified, desalted, and either directly measured on a nano LC-MS/MS setup (see below) or subjected to isoelectric focusing using an OFFGEL fractionator.

Input samples were reduced by adding dithiothreitol (DTT) to a final concentration of 0.1 M and incubation for 5 min at 95 °C. Sulfhydryl groups were alkylated by adding iodoacetamide to a final concentration of 0.25 M and incubation for 20 min in the dark at room temperature. Proteins were precipitated according to Wessel and Flügge, resuspended in 6 M urea/2 M thiourea and digested into peptides with C-terminal lysine or arginine using Lys-C (3 h) and trypsin (overnight, diluted 4× with 50 mM ABC). Enzyme activity was quenched by acidification of the samples with trifluoroacetic acid. The peptides were desalted with C18 Stage Tips prior to nanoLC-MS/MS analysis.

**pSILAC.** Cells were adapted to SILAC-light medium one day before the experiment and infected as described above using the indicated MOI. Prior to the pulse period, cells were maintained in PBS supplemented with Ca<sup>2+</sup>/Mg<sup>2+</sup> and 0.2% BSA for 30 min. Then, cells were pulse labeled with SILAC-M or SILAC-H medium for 6 h intervals, harvested, and then combined. Lysis was carried out in 125 mM NaCl, 0.1% SDS, 1% NP-40, 5% glycerol, 50 mM Tris-HCl, pH 7.4 for 1 h on a rotating wheel with subsequent centrifugation. The supernatant was precipitated according to Wessel and Flügge and precipitated proteins were subjected to in-solution digest. Proteins were denatured in 6 M urea/2 M thiourea, reduced, alkylated, and digested using Lys-C (3 h at 20 °C). Then, the digest solution was diluted 4× with 50 mM ammonium bicarbonate buffer and incubated with trypsin (Promega) for 16 h at 20 °C. Afterwards, samples were acidified and subjected to stage tip purification.

**Mass spectrometry.** Peptides from input and AHA-enriched samples were separated on a monolithic silica capillary column (MonoCap C18 High-Resolution 2000, GL Sciences), 0.1 mm internal diameter × 2000 mm length, at a flow rate of 300 nl/min with a 5 to 45% acetonitrile gradient on an EASY-nLC II system (Thermo Fisher Scientific) with 480 or 240 min gradient or on a EASY-nLC HPLC (Thermo Fisher) system by 2 or 4 h gradients with a 250 nl/min flow rate on a 15 cm column with an inner diameter of 75 μm packed in-house with ReproSil-Pur C18-AQ material (Dr. Maisch, GmbH). Peptides were ionized using an ESI source on a Q-Exactive, Q-Exactive Plus or a LTQ (linear trap quadrupole) Orbitrap Velos MS (all Thermo Fisher) in data-dependent mode. Q-Exactive and Q-Exactive Plus mass spectrometers were operated in the data-dependent mode with a full scan in the Orbitrap, followed by top 10 MS/MS scans using higher-energy collision dissociation. The full scans were performed with a *m/z* range of 300–1700, a resolution of 70,000, a target value of 3 × 10<sup>6</sup> ions and a maximum injection time of 20 ms. The MS/MS scans were performed with a 17,500 resolution, a 1 × 10<sup>5</sup> target value, and a 60 ms maximum injection time. The LTQ Orbitrap Velos instrument was operated in data-dependent CID top 20 mode. Full scans were performed in

*m/z* range 300–1700 with a resolution of 60,000 and a target value of  $10^6$ . MS/MS scans were performed with an isolation window of  $2m/z$  and a target value of 3000.

Peptides from pSILAC samples were separated by 4 h gradients and ionized with ESI source and analyzed on Q-Exactive HF-X instrument (Thermo Fisher) in data-dependent mode. The full scans were performed with a resolution of 60,000, a target value of  $3 \times 10^6$  ions and a maximum injection time of 10 ms. The MS/MS scans were performed with a 15,000 resolution, a  $1 \times 10^5$  target value and a 22 ms maximum injection time.

**Data analysis.** Raw files of the pAHA-SILAC were analyzed with MaxQuant software version 1.6.0.1. Default settings were kept, except that “requantify” option was turned on. Label-free quantification via iBAQ calculation was enabled. Lys4/Arg6 and Lys8/Arg10 were set as labels and oxidation of methionines, n-terminal acetylation, and deamidation of asparagine and glutamine residues were defined as variable modifications. The in silico digests of the human Uniprot database (downloaded January 2018), the protein sequences of 12 Pan and Mal Influenza virus proteins and a database containing common contaminants were done with Trypsin/P. The false discovery rate was set to 1% at both the peptide and protein level and was assessed by parallel searching a database containing the reverted sequences from the Uniprot database. The resulting text files were filtered to exclude reverse database hits, potential contaminants and proteins only identified by site (i.e., protein identifications that are only explained by a modified peptide). Plotting and statistics were done using R and figures were compiled in Illustrator (Adobe). Raw files for pSILAC were analyzed as described above, except that MaxQuant software version 1.5.2.8 was used. For the pSILAC experiment comparing Mal infection in human and avian cells, spectra were searched against a database additionally containing the uniprot entries of *Gallus gallus*. For the pSILAC experiment comparing Pan and Pan-AV, requantify option was set to off.

**Proteomic data processing.** Two MaxQuant output files were used: proteinGroups.txt and evidence.txt. iBAQ values from infected samples were extracted from proteinGroups.txt. iBAQ values were first normalized by scaling to the iBAQ protein median across all mock-infected samples. This assumes that there are no differences in overall protein synthesis for different mock-infected samples. The iBAQ values were averaged for the corresponding label-swap replicates and proteins were categorized as host or viral. For estimating the newly synthesized protein mass, intensity values of H and M SILAC channels were divided by the summed up intensities of the light channel (mock infected). Data were then averaged for label-swap replicates and summed up for viral and host proteins independently. Finally, data were normalized to the 0–4 h time period.

SILAC ratios of host proteins were processed by first transforming them into log<sub>2</sub> space. The median SILAC-H/L and SILAC-M/L ratios from the input samples were used to estimate the mixing ratio of the input and the H/L and M/L ratios after the enrichment were adjusted correspondingly. SILAC-H/M ratio that relate to the Pan/Mal (or Mal/Pan) infection treatment were normalized to 0. Then, the replicate measurements were averaged. Proteins that were quantified in only one replicate were excluded. Later the top 2% of proteins (highest log<sub>2</sub> fold-change) of either Pan/Mock or Mal/Mock condition were selected and multi-set GO enrichment was performed using Metascape tool (<http://metascape.org>).

Protein level data was matched to RNA level data based on the HGNC official gene symbol. Protein synthesis efficiencies were calculated by subtracting log<sub>10</sub> (RPKM) from log<sub>10</sub>(iBAQ) values.

For quantification of pAHA-SILAC viral protein expression kinetics, we extracted all quantifications of peptide level evidences for each individual replicate from evidence.txt. Median log<sub>2</sub> ratios were then normalized to 0 for individual replicates. Comparative viral protein expression kinetics were based on Pan/Mal shared peptides. For each time period, replicate, and viral protein, the Pan/Mal SILAC protein ratio was calculated as the median of all SILAC peptide level ratios. Replicate SILAC protein ratios were averaged. For pSILAC, ratios for viral proteins were extracted from proteinGroups.txt. Non-normalized and log<sub>2</sub>-transformed ratios were used.

**RNA-sequencing and data processing.** Total RNAs from A549 cells infected with Pan or Mal strain (infection conditions as described in the pAHA-SILAC experiment) were extracted using TRIzol reagent (Life Technologies) following the manufacturer’s protocol. Truseq Stranded mRNA sequencing libraries were prepared with 500 ng total RNA according to the manufacturer’s protocol (Illumina). The libraries were sequenced on HiSeq 2000 platform (Illumina) and yielded in total 186 million 101-nt single-end reads. The sequencing reads were first subjected to adapter removal using flexbar with the following parameters: -x 6 -y 5 -u 2 -m 28 -ae RIGHT -at 2 -ao 1 -n 4 -j -z GZ<sup>65</sup>. Reads mapped to the reference sequences of rRNA, tRNA, snRNA, snoRNA and miscRNAs (available from Ensembl and RepeatMasker annotation) using Bowtie2 (version 2.1.0) with default parameters (in --end-to-end & --sensitive mode) were excluded. The remaining reads were then mapped to the human and Pan/Mal influenza A reference genome using Tophat2 (v2.0.10) with parameters: -N 3 --read-gap-length 2 --read-edit-dist 3 --min-anchor 6 --library-type fr-firststrand --segment-mismatches 2 --segment-length 26, and the guidance of RefSeq/Ensembl human gene structure and known viral gene annotation. Gene expression levels (RPKM) were estimated by Cufflinks

(v2.2.1) with parameters: -u --library-type fr-firststrand --overhang-tolerance 6 --max-bundle-frags 50000000. Splice junction reads for various M transcripts were counted with customized Perl scripts. Splice isoforms were accepted that had >500 read counts in both replicates. The same RNA-seq experimental and computational procedures were applied to samples that were transfected, as well as the Pan-AV and the control wild-type infection samples (Pan), except that HiSeq 4000 platform (Illumina) was used and that libraries were sequenced in the  $2 \times 151$ -nt paired-end mode.

To account for differences in transfection efficiency and sequencing depth from transfected samples, we calculated a scaling factor for NP, PA, PB1, and PB2 across all experiments based on their FPKM expression values. The splice junction read count was then divided by the average of the scaling factors.

**Transfections.** A549 cells were seeded on 6-well plates and transfected with 2.5  $\mu$ g of expression constructs and Lipofectamine 3000 (Thermo Fisher) reagent according to the manufacturer’s instructions. Cells were harvested by trypsinization and subjected to lysis and immunoblotting. For the RdrP reporter experiment, 10 cm dishes of HEK293T cells were transfected with vectors encoding NP (4.5  $\mu$ g), PA (1.5  $\mu$ g), PB1 (1.5  $\mu$ g), or PB2 (1.5  $\mu$ g) of strain WSN and either pHW2000-Mal M, pHW2000-Pan M pHW2000-Pan-AV, or pHW200-Mal M w/ Pan 707-779 (0.75  $\mu$ g) using polyethylenimine. Half of the sample was subjected to RNA isolation and RNA-sequencing 24 h post transfection.

**Lysis, SDS-polyacrylamide gel electrophoresis, and immunoblotting.** Cells were lysed in lysis buffer (125 mM NaCl, 0.1% SDS, 1% NP-40, 5% glycerol, 50 mM Tris-HCl, pH 7.4) for 1 h on a rotating wheel and centrifuged. Supernatant was supplemented with NuPage LDS Sample buffer (Invitrogen), 50 mM DTT, and heated for 10 min at 70 °C. Samples were run on 4–12% Bis-Tris gradient gels (NuPAGE, Invitrogen) before being blotted onto PVDF membrane (Immobilon-P, Millipore) using a wet blotting system (Invitrogen). Specific antibodies against the HA epitope (clone 3F10, Roche, 1:500, #11867423001), vezatin (clone B-1, Santa Cruz, 1:1000, #sc271347), M1 (clone GA2B, Bio-Rad, 1:1000, #MCA401) or M2 (polyclonal, RRID: AB\_2549706, Thermo Fisher, 1:1000, #PA5-32233) and suitable horseradish peroxidase-linked secondary antibodies were used. See source data file for raw blot images.

**Immunofluorescence microscopy.** A549 cells were grown on glass coverslips and infected with the indicated viruses at an MOI of 1 FFU (fluorescence-forming unit)/cell. At the indicated time points post infection, cells were fixed for 15 min in 2.5% formaldehyde, permeabilized in 0.2% Triton X-100, and stained with specific antibody against NP antigen (clone AA5H, Bio-Rad/Serotec, 1:1000, #MCA400) in 3% BSA for 1 h. After washing, cells were stained with Alexa Fluor 488-conjugated goat anti-mouse IgG secondary antibody (1:1000) in 3% BSA. Nuclei were counterstained with 4',6-diamidino-2-phenylindole (DAPI) by adding to the secondary antibody solution. Coverslips were mounted using Mowiol and images were acquired by an Eclipse A1 laser-scanning microscope using the NIS-Elements software package (Nikon). At least 140 cells were counted per condition to quantify the subcellular distribution of NP using the ImageJ software.

**Computational RNA structure analyses.** M segment sequences were obtained from the NIAID Influenza Research Database (IRD) through the website at <http://www.fludb.org>. We used the following settings for human-adapted strains: date range  $\geq 2009$ , subtype H3N2, only complete genomes, host human; exclude laboratory strains and duplicate sequences and geographic grouping: South America, Europe, and Asia; for avian-adapted strains: date range  $\geq 2009$ , only complete genomes, host: avian; exclude duplicate sequences and geographic grouping: Europe and Asia. These two sets of sequences were merged with the respective references sequence, that is, A/Panama/2007/1999—Pan (human) and A/Mallard/439/2004—Mal (mallard), and aligned using the program Muscle<sup>66</sup> resulting in two MSAs, one for the human-adapted strains comprising 403 sequences and one MSA for the avian-adapted strains comprising 199 sequences. Both alignments are straightforward to establish based on the long open reading frame of the M1 isoform covering almost all of the M segment and on the overall high primary sequence conservation of the M segment. Evolutionary trees relating the sequences in either MSA were then derived using PhyML in conjunction with the HKY evolutionary model, which provided the best fit to the data [version v3.0<sup>67</sup>].

These two input alignments (including the combined annotation of the known protein-coding M1 and M2 regions) and the corresponding evolutionary trees were then used as input to RNA-Decoder<sup>52</sup>. We used RNA-Decoder to identify the RNA secondary structure that is best supported by the evolutionary signals contained in the two input alignments (the so-called maximum-likelihood structure). The predictions by RNA-Decoder also included the posterior base-pairing probabilities for each base pair of the predicted RNA structure. Predicted base pairs with a base-pairing probability smaller than 25% were omitted from the RNA structure visualization. Note that each multiple sequence alignment was analyzed by RNA-Decoder in one chunk, that is, without partitioning it artificially into sub-alignments.

Finally, the predicted RNA structure element nucleotides 733–766 was plotted with the sequence of the Mal strain using VARNA tool<sup>68</sup> and the RNA structures predicted for the two alignments of avian- and human-adapted sequences was visualized using R-chie<sup>69</sup>, including information on the pairing probability of each base pair.

**Nucleotide-level conservation.** To retrieve M segment consensus sequences for human and avian-adapted strains, we used the NIAID IRD<sup>70</sup> through the website at <http://www.fludb.org> (downloaded 10 July 2018). The following options were applied to retrieve avian-adapted consensus sequences: date range: 1999–2011, complete genome, minimum length M segment sequence 1027, and exclude duplicated and laboratory strains. The following options were used to retrieve the human consensus: date range: 1999–2011, complete genome, minimum length M segment sequence 1002, and exclude duplicated and laboratory strains.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Proteomic data relating to the pAHA-SILAC and pSILAC experiments was uploaded to ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD011321 (pAHA-SILAC) and PXD015475, PXD015474 (pSILAC) RNA-sequencing data is publicly available under PRJNA495615 [<https://www.ncbi.nlm.nih.gov/sra/PRJNA495615>]. The source data underlying Figs. 1c, d, 2a–d, f–h, 3a, d, 4a, b, d, f, 5c–g, 6b–f and Supplementary Figs. 2 and 4a–c are available as Source Data file.

Received: 8 October 2018; Accepted: 12 November 2019;

Published online: 04 December 2019

## References

- Neumann, G. & Kawaoka, Y. Transmission of influenza A viruses. *Virology* **479–480**, 234–246 (2015).
- Matrosovich, M. N. et al. Avian influenza A viruses differ from human viruses by recognition of sialyloligosaccharides and gangliosides and by a higher conservation of the HA receptor-binding site. *Virology* **233**, 224–234 (1997).
- Te Velthuis, A. J. W. & Fodor, E. Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. *Nat. Rev. Microbiol.* **14**, 479–493 (2016).
- Long, J. S. et al. Species difference in ANP32A underlies influenza A virus polymerase host restriction. *Nature* **529**, 101–104 (2016).
- Mair, C. M., Ludwig, K., Herrmann, A. & Sieben, C. Receptor binding and pH stability—how influenza A virus hemagglutinin affects host-specific virus infection. *Biochimica et Biophysica Acta (BBA)* **1838**, 1153–1168 (2014).
- Mänz, B., Schwemmler, M. & Brunotte, L. Adaptation of avian influenza A virus polymerase in mammals to overcome the host species barrier. *J. Virol.* **87**, 7200–7209 (2013).
- Taubenberger, J. K. & Kash, J. C. Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host Microbe* **7**, 440–451 (2010).
- Chua, M. A., Schmid, S., Perez, J. T., Langlois, R. A. & Tenover, B. R. Influenza A virus utilizes suboptimal splicing to coordinate the timing of infection. *Cell Rep.* **3**, 23–29 (2013).
- Dubois, J., Terrier, O. & Rosa-Calatrava, M. Influenza viruses and mRNA splicing: doing more with less. *MBio* **5**, e00070–14 (2014).
- Martin, K. & Helenius, A. Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (M1) promotes export and inhibits import. *Cell* **67**, 117–130 (1991).
- Bui, M., Wills, E. G., Helenius, A. & Whittaker, G. R. Role of the influenza virus M1 protein in nuclear export of viral ribonucleoproteins. *J. Virol.* **74**, 1781–1786 (2000).
- Pinto, L. H., Holsinger, L. J. & Lamb, R. A. Influenza virus M2 protein has ion channel activity. *Cell* **69**, 517–528 (1992).
- Lamb, R. A., Richardson, C. D. & Zebedee, S. L. Influenza virus M2 protein is an integral membrane protein expressed on the infected-cell surface. *Virus Res.* **3**, 4 (1985).
- Valcárcel, J., Portela, A. & Ortín, J. Regulated M1 mRNA splicing in influenza virus-infected cells. *J. Gen. Virol.* **72**(Part 6), 1301–1308 (1991).
- König, R. & Stertz, S. Recent strategies and progress in identifying host factors involved in virus replication. *Curr. Opin. Microbiol.* **26**, 79–88 (2015).
- Karlas, A., Machuy, N. & Meyer, T. F. Human host cell factors crucial for influenza virus replication identified by genome-wide RNAi screen. *Nat. Biotechnol.* **27**, S84 (2010).
- König, R. et al. Human host factors required for influenza virus replication. *Nature* **463**, 813–817 (2010).
- Brass, A. L. et al. The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. *Cell* **139**, 1243–1254 (2009).
- Watanabe, T. et al. Influenza virus–host interactome screen as a platform for antiviral drug development. *Cell Host Microbe* **16**, 795–805 (2014).
- Wang, L. et al. Comparative influenza protein interactomes identify the role of plakophilin 2 in virus restriction. *Nat. Commun.* **8**, 13876 (2017).
- Pichlmair, A. et al. Viral immune modulators perturb the human molecular network by common and unique strategies. *Nature* **487**, 486–490 (2012).
- Becker, A. C. et al. Influenza A virus induces autophagosomal targeting of ribosomal proteins. *Mol. Cell. Proteomics* mcp.RA117.000364 (2018).
- Simon, P. F. et al. Highly pathogenic H5N1 and novel H7N9 influenza A viruses induce more profound proteomic host responses than seasonal and pandemic H1N1 strains. *J. Proteome Res.* **14**, 4511–4523 (2015).
- Liu, L. et al. Proteome alterations in primary human alveolar macrophages in response to influenza A virus infection. *J. Proteome Res.* **11**, 4091–4101 (2012).
- Kroeker, A. L., Ezzati, P., Halayko, A. J. & Coombs, K. M. Response of primary human airway epithelial cells to influenza infection: a quantitative proteomic study. *J. Proteome Res.* **11**, 4132–4146 (2012).
- Wu, X. et al. Subcellular proteomic analysis of human host cells infected with H3N2 swine influenza virus. *Proteomics* **13**, 3309–3326 (2013).
- Sadewasser, A. et al. Quantitative proteomic approach identifies Vpr binding protein as novel host factor supporting influenza A virus infections in human cells. *Mol. Cell. Proteom.* **16**, 728–742 (2017).
- Inglis, S. C. & Brown, C. M. Differences in the control of virus mRNA splicing during permissive or abortive infection with influenza A (fowl plague) virus. *J. Gen. Virol.* **65**(Part 1), 153–164 (1984).
- Shapiro, G. L., Gurney, T. Jr & Krug, R. M. Influenza virus gene expression: control mechanisms at early and late times of infection and nuclear-cytoplasmic transport of virus-specific RNAs. *J. Virol.* **61**, 764–773 (1987).
- Mann, M. Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell. Biol.* **7**, 952–958 (2006).
- Schwanhäusser, B., Gossen, M., Dittmar, G. & Selbach, M. Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* **9**, 205–209 (2009).
- Doherty, M. K., Hammond, D. E., Clague, M. J., Gaskell, S. J. & Beynon, R. J. Turnover of the human proteome: determination of protein intracellular stability by dynamic SILAC. *J. Proteome Res.* **8**, 104–112 (2009).
- Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A. & Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl Acad. Sci. USA* **103**, 9482–9487 (2006).
- McShane, E. et al. Kinetic analysis of protein stability reveals age-dependent degradation. *Cell* **167**, 803–815.e21 (2016).
- Howden, A. J. M. et al. QuanCAT: quantitating proteome dynamics in primary cells. *Nat. Methods* **10**, 343–346 (2013).
- Eichelbaum, K. & Krijgsvelde, J. Rapid temporal dynamics of transcription, protein synthesis, and secretion during macrophage activation. *Mol. Cell. Proteom.* **13**, 792–810 (2014).
- Schwanhäusser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
- Rivas, H. G., Schmaling, S. K. & Gaglia, M. M. Shutoff of host gene expression in influenza A virus and herpesviruses: similar mechanisms and common themes. *Viruses* **8**, 102 (2016).
- Rodríguez, A., Pérez-González, A. & Nieto, A. Influenza virus infection causes specific degradation of the largest subunit of cellular RNA polymerase II. *J. Virol.* **81**, 5315–5324 (2007).
- Khapersky, D. A., Schmaling, S., Larkins-Ford, J., McCormick, C. & Gaglia, M. M. Selective degradation of host RNA polymerase II transcripts by influenza A virus PA-X host shutoff protein. *PLoS Pathog.* **12**, e1005427 (2016).
- Bercovich-Kinori, A. et al. A systematic view on influenza induced host shutoff. *Elife* **5**, <https://doi.org/10.7554/eLife.18311> (2016).
- Park, Y. W. & Katze, M. G. Translational control by influenza virus. Identification of *cis*-acting sequences and *trans*-acting factors which may regulate selective viral mRNA translation. *J. Biol. Chem.* **270**, 28433–28439 (1995).
- Zhou, T., Gu, W., Ma, J., Sun, X. & Lu, Z. Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses. *Biosystems* **81**, 77–86 (2005).
- Gabriel, G. et al. The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. *Proc. Natl Acad. Sci. USA* **102**, 18590–18595 (2005).
- Hutchinson, E. C. et al. Conserved and host-specific features of influenza virion architecture. *Nat. Commun.* **5**, 4816 (2014).
- Killip, M. J., Fodor, E. & Randall, R. E. Influenza virus activation of the interferon system. *Virus Res.* **209**, 11–22 (2015).
- Shih, S. R., Suen, P. C., Chen, Y. S. & Chang, S. C. A novel spliced transcript of influenza A/WSN/33 virus. *Virus Genes* **17**, 179–183 (1998).
- Wise, H. M. et al. Identification of a novel splice variant form of the influenza A virus M2 ion channel with an antigenically distinct ectodomain. *PLoS Pathog.* **8**, e1002998 (2012).

49. Shih, S. R., Nemeroff, M. E. & Krug, R. M. The choice of alternative 5' splice sites in influenza virus M1 mRNA is regulated by the viral polymerase complex. *Proc. Natl Acad. Sci. USA* **92**, 6324–6328 (1995).
50. Moss, W. N. et al. The 3' splice site of influenza A segment 7 mRNA can exist in two conformations: a pseudoknot and a hairpin. *PLoS ONE* **7**, e38323 (2012).
51. Shih, S. R. & Krug, R. M. Novel exploitation of a nuclear function by influenza virus: the cellular SF2/ASF splicing factor controls the amount of the essential viral M2 ion channel protein in infected cells. *EMBO J.* **15**, 5415–5427 (1996).
52. Pedersen, J. S., Meyer, I. M., Forsberg, R., Simmonds, P. & Hein, J. A comparative method for finding and folding RNA secondary structures within protein-coding regions. *Nucleic Acids Res.* **32**, 4925–4936 (2004).
53. Watts, J. M. et al. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711–716 (2009).
54. Worobey, M., Han, G.-Z. & Rambaut, A. A synchronized global sweep of the internal genes of modern avian influenza virus. *Nature* **508**, 254–257 (2014).
55. Backström Winquist, E. et al. Inefficient splicing of segment 7 and 8 mRNAs is an inherent property of influenza virus A/Brevig Mission/1918/1 (H1N1) that causes elevated expression of NS1 protein. *Virology* **422**, 46–58 (2012).
56. Calderon, B. M. et al. Dysregulation of M segment gene expression contributes to influenza A virus host restriction. *PLoS Pathog.* **15**, e1007892 (2019). <https://doi.org/10.1371/journal.ppat>.
57. Garfinkel, M. S. & Katze, M. G. Translational control by influenza virus. Selective translation is mediated by sequences within the viral mRNA 5'-untranslated region. *J. Biol. Chem.* **268**, 22223–22226 (1993).
58. Huang, X. et al. An NS-segment exonic splicing enhancer regulates influenza A virus replication in mammalian cells. *Nat. Commun.* **8**, 14751 (2017).
59. Mor, A. et al. Influenza virus mRNA trafficking through host nuclear speckles. *Nat. Microbiol.* **1**, 16069 (2016).
60. Thompson, M. G. et al. Co-regulatory activity of hnRNP K and NS1-BP in influenza and human mRNA splicing. *Nat. Commun.* **9**, 2407 (2018).
61. Robb, N. C. & Fodor, E. The accumulation of influenza A virus segment 7 spliced mRNAs is regulated by the NS1 protein. *J. Gen. Virol.* **93**, 113–118 (2012).
62. Tsai, P.-L. et al. Cellular RNA binding proteins NS1-BP and hnRNP K regulate influenza A virus RNA splicing. *PLoS Pathog.* **9**, e1003460 (2013).
63. Morens, D. M., Taubenberger, J. K. & Fauci, A. S. The persistent legacy of the 1918 influenza virus. *N. Engl. J. Med.* **361**, 225–229 (2009).
64. Reid, A. H., Fanning, T. G., Janczewski, T. A., McCall, S. & Taubenberger, J. K. Characterization of the 1918 'Spanish' influenza virus matrix gene segment. *J. Virol.* **76**, 10717–10723 (2002).
65. Dodt, M., Roehr, J. T., Ahmed, R. & Dieterich, C. FLEXBAR—flexible barcode and adapter processing for next-generation sequencing platforms. *Biology* **1**, 895–905 (2012).
66. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
67. Guindon, S., Delsuc, F., Dufayard, J.-F. & Gascuel, O. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol. Biol.* **537**, 113–137 (2009).
68. Darty, K., Denise, A. & Ponty, Y. VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).
69. Lai, D. & Meyer, I. M. e-RNA: a collection of web servers for comparative RNA structure prediction and visualisation. *Nucleic Acids Res.* **42**, W373–W376 (2014).
70. Zhang, Y. et al. Influenza Research Database: an integrated bioinformatics resource for influenza virus research. *Nucleic Acids Res.* **45**, D466–D474 (2017).

## Acknowledgements

This work was supported by the German Ministry of Education and Research (ViroSign grant 0316180B). We would like to acknowledge the excellent technical assistance of Gudrun Heins. We thank Katrina Meyer for providing the pDEST26-Flag/HA plasmid and Koshi Imami for setting up the nLC system using monolithic columns. We also thank Marina Chekulaveva for in vitro transcription constructs and Sascha Sauer of the MDC Scientific Genomics Platform for RNA-sequencing.

## Author contributions

B.B. performed most of the experiments and data analyses. K.E. established the AHA pulse labeling technology and generated pAHA-SILAC data together with B.B. A.S. and K.P. performed the infection experiments for proteomic analysis supervised by T.W. X.W. and J.H. generated and analyzed RNA-seq data supervised by W.C. I.H. performed qRT-PCR analysis supervised by L.W. and B.B. M.H., L.W., and B.V. contributed to cloning and transfection of reporter constructs. I.M.M. performed the RNA structure analyses. M.B. conducted reverse genetic experiments. T.W. and M.S. conceived and supervised the study. B.B., L.W., T.W., and M.S. interpreted the data. B.B. and M.S. wrote the manuscript with input from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-019-13520-8>.

**Correspondence** and requests for materials should be addressed to M.S.

**Peer review information** *Nature Communications* thanks Aleksandra Nita-Lazar and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019