

# Insertions of codons encoding basic amino acids in H7 hemagglutinins of influenza A viruses occur by recombination with RNA at hotspots near snoRNA binding sites

ALEXANDER P. GULTYAEV,<sup>1,2</sup> MONIQUE I. SPRONKEN,<sup>1</sup> MATHIS FUNK,<sup>1</sup> RON A.M. FOUCHIER,<sup>1</sup> and MATHILDE RICHARD<sup>1</sup>

<sup>1</sup>Department of Viroscience, Erasmus Medical Center, 3000 CA Rotterdam, the Netherlands

<sup>2</sup>Group Imaging and Bioinformatics, Leiden Institute of Advanced Computer Science (LIACS), Leiden University, 2300 RA Leiden, the Netherlands

## ABSTRACT

The presence of multiple basic amino acids in the protease cleavage site of the hemagglutinin (HA) protein is the main molecular determinant of virulence of highly pathogenic avian influenza (HPAI) viruses. Recombination of HA RNA with other RNA molecules of host or virus origin is a dominant mechanism of multibasic cleavage site (MBCS) acquisition for H7 subtype HA. Using alignments of HA RNA sequences from documented cases of MBCS insertion due to recombination, we show that such recombination with host RNAs is most likely to occur at particular hotspots in ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), and viral RNAs. The locations of these hotspots in highly abundant RNAs indicate that RNA recombination is facilitated by the binding of small nucleolar RNA (snoRNA) near the recombination points.

**Keywords:** influenza virus; avian influenza; snoRNA; RNA recombination

## INTRODUCTION

Low pathogenic avian influenza (LPAI) viruses, which circulate in wild bird populations, occasionally evolve into highly pathogenic avian influenza (HPAI) viruses in poultry that can cause devastating outbreaks (for review, see Lee et al. 2020). HPAI strains emerge from LPAI viruses upon conversion of a cleavage site in the surface protein hemagglutinin (HA) from the motif recognized by trypsin-like proteases into a so-called multiple basic cleavage site (MBCS) which is recognized by furin-like proteases (Garten et al. 2015). As cleavage of the HA polyprotein precursor HA0 into two chains, HA1 and HA2, is an essential step in the virus replication cycle and furin-like proteases are present in a broader range of host tissues, the motif change facilitates the systemic virus dissemination in chickens, leading to severe disease (Lee et al. 2020).

Influenza A viruses are members of the *Orthomyxoviridae* family. They have a segmented negative-sense RNA genome that consists of eight segments. The viruses are

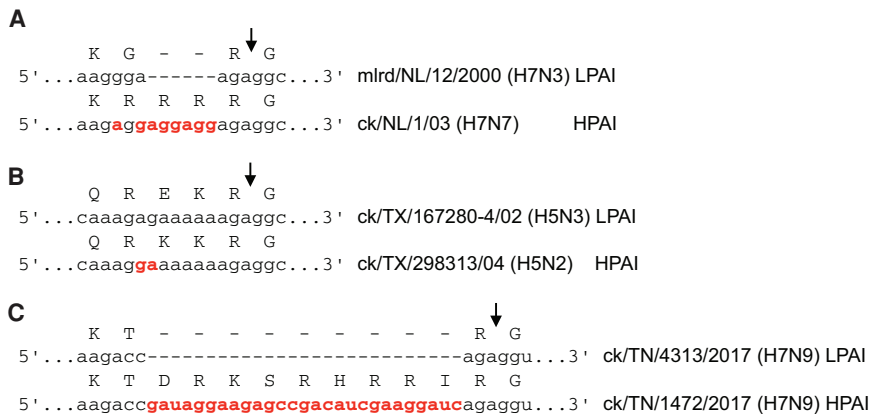
classified into subtypes defined by the HA and neuraminidase (NA) proteins. Although sixteen HA subtypes (H1–H16) are known to circulate in birds, only H5 and H7 subtypes have been observed to evolve toward a HPAI phenotype in poultry (Abdelwhab et al. 2013; Lee et al. 2020). Three molecular mechanisms of the changes leading to MBCS formation have been described (for reviews, see Abdelwhab et al. 2013; Naguib et al. 2019; Lee et al. 2020): (i) duplications of the nucleotides coding for the cleavage site by polymerase stuttering (common for both H5 and H7 subtypes); (ii) point substitutions (usually in combination with duplications, but also observed in their absence in H5 HA segments), and (iii) recombination of H7 HA RNA segment with other RNAs (Fig. 1).

So far, in naturally occurring viruses MBCS formation by recombination has only been observed in H7 HA segments (Banks et al. 2001; Hirst et al. 2004; Suarez et al. 2004; Berhane et al. 2009; Maurer-Stroh et al. 2013; Lee et al. 2017;

**Corresponding authors:** a.goultiaev@erasmusmc.nl, m.richard@erasmusmc.nl

Article is online at <http://www.majournal.org/cgi/doi/10.1261/rna.077495.120>.

© 2021 Gulyaev et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://majournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.



**FIGURE 1.** Examples of MBCS insertions in HPAI viruses. (A) Substitutions and duplications of Arg codons in the HA RNA from HPAI strain A/chicken/Netherlands/1/03 (H7N7) as compared to closely related HA RNA from LPAI virus A/mallard/Netherlands/12/2000(H7N3) (Fouchier et al. 2004). (B) Substitution creating the MBCS site of A/chicken/TX/298313/2004(H5N2) as compared to LPAI strain A/chicken/Texas/167280-4/2002(H5N3) (Lee et al. 2005). (C) Recombination of HA RNA with 28S rRNA leading to the insertion of nine codons in the HPAI virus A/chicken/Tennessee/17-007147-2/2017(H7N9) as compared to related LPAI strain A/chicken/Tennessee/17-007431-3/2017(H7N9) (Lee et al. 2017). Inserted or mutated nucleotides are in red, locations of HA0 cleavage sites are shown by arrows.

Beerens et al. 2020). Also, laboratory H7 strains with MBCS acquired by recombination during adaptation for the growth without trypsin have been produced (Khatchikian et al. 1989; Li et al. 1990; Orlich et al. 1994). Recombina-

tion at the cleavage site region of HA of H5 subtype has been detected for the first time by ultra-deep sequencing upon serial passaging of a H5N2 virus (Laleye and Abolnik 2020). In a few cases with relatively large insertions (at least 21 nt, or seven codons), their origins could be reliably established by sequence database similarity searches. These insertions either originated from other segments of the virus (Table 1; Orlich et al. 1994; Hirst et al. 2004; Suarez et al. 2004; Laleye and Abolnik 2020) or from 28S rRNA (Khatchikian et al. 1989; Maurer-Stroh et al. 2013; Lee et al. 2017). The origins of the shorter insertions are more difficult to identify, in particular due to the low significance of searches with short query sequences which are expected to yield false-positive hits identical to the queries. Nevertheless, some of

these shorter insertions are unlikely to be produced by polymerase stuttering or sequential substitutions and rather could be due to recombination (Banks et al. 2001; Berhane et al. 2009; Monne et al. 2014; Beerens et al. 2020).

**TABLE 1.** MBCS insertions into H7 HA segments with (possible) recombination origin

Strain	Accession/reference	Insertion size (nucleotides)	Insertion origin
A/seal/Mass/1/80(H7N7) (lab strain)	Orlich et al. 1994	60	NP segment (Orlich et al. 1994)
A/turkey/Oregon/1971(H7N3) (lab strain)	Khatchikian et al. 1989	54	28S rRNA (Khatchikian et al. 1989)
A/chicken/Chile/4322/02(H7N3)	AY303631 (Suarez et al. 2004)	30	NP segment (Suarez et al. 2004)
A/chicken/Tennessee/17-007147-2/2017(H7N9)	KY818811 (Lee et al. 2017)	27	28S rRNA (Lee et al. 2017)
A/chicken/Jalisco/CPA1/2012(H7N3)	JX397993 (Maurer-Stroh et al. 2013)	24	28S rRNA (Maurer-Stroh et al. 2013)
A/Canada/rv504/2004(H7N3)	CY015006 (Hirst et al. 2004)	21	M segment (Hirst et al. 2004)
A/chicken/SK/HR-00011/2007(H7N3)	EU500860 (Berhane et al. 2009)	18	28S rRNA (this work)
A/chicken/Spain/6279-2/2009(H7N7)	GU121458 (Iglesias et al. 2010)	18	28S rRNA (this work)
A/chicken/Italy/5093/99(H7N1)	DQ991328 (Banks et al. 2001)	12	tRNA-Val ? (this work)
A/chicken/Netherlands/03008927/2003 (H7N3)	EPI_ISL_391798 (Beerens et al. 2020)	12	tRNA-Val ? (this work)
A/chicken/Guangdong/J1/2017(H7N9)	KY855518 (Chen et al. 2017)	12	?
A/chicken/England/26352/2015(H7N7)	EPI_ISL_191604 (Abdelwhab et al. 2016)	9	?
A/seal/Mass/1/80(H7N7) (lab strain SC35)	Li et al. 1990	9	?

Representative strains of the outbreaks and viruses obtained by passaging in laboratory are listed in the descending order of insertion sizes.

Here, we used various alignment programs and database similarity searches to identify potential sources of insertions that have created MBCS sequences in known cases of H7 HPAI strains. The results led us to hypothesize that the hotspots for rRNA recombination with HA are determined by small nucleolar RNA (snoRNA) binding. Furthermore, recombination of other RNA molecules with HA also occurs at the sites homologous to snoRNA target sequences, yielding new mechanistic insight into emergence of HPAI strains from LPAI virus precursors via recombination.

The main function of snoRNAs is to guide modifications of nucleotides in ribosomal RNAs (rRNAs) and small nuclear RNAs (snRNAs) (Smith and Steitz 1997; Bachellerie et al. 2002), via base-pairing of guide regions, also called antisense elements, to target sequences. SnoRNAs are classified into two major types on the basis of their characteristic motifs and proteins found in the small nucleolar ribonucleoprotein complexes (snoRNPs): C/D box and H/ACA box snoRNAs. The C/D box snoRNPs catalyze 2'-O-methylation of nucleotides located within targets bound by antisense elements of 7–21 nt. The H/ACA box snoRNPs guide pseudouridylation of uridines. Each of the guide elements of H/ACA box snoRNAs forms two relatively short duplexes. Binding of some snoRNAs to their targets is involved in regulation of processes other than nucleotide methylation (Bachellerie et al. 2002; Bratkovič et al. 2019; Bergeron et al. 2020), and snoRNA functions could be exploited by viruses (for review, see Stamm and Lodmell 2019). The structures of multiple snoRNAs, their guide regions and binding sites are known (Lestrade and Weber 2006; Bratkovič et al. 2019) or can be identified by sequence comparison between different species. In particular, we analyzed the binding sites of chicken snoRNAs to 28S rRNA in order to identify the features that could distinguish those located near hotspots of recombination with HA from others.

## RESULTS

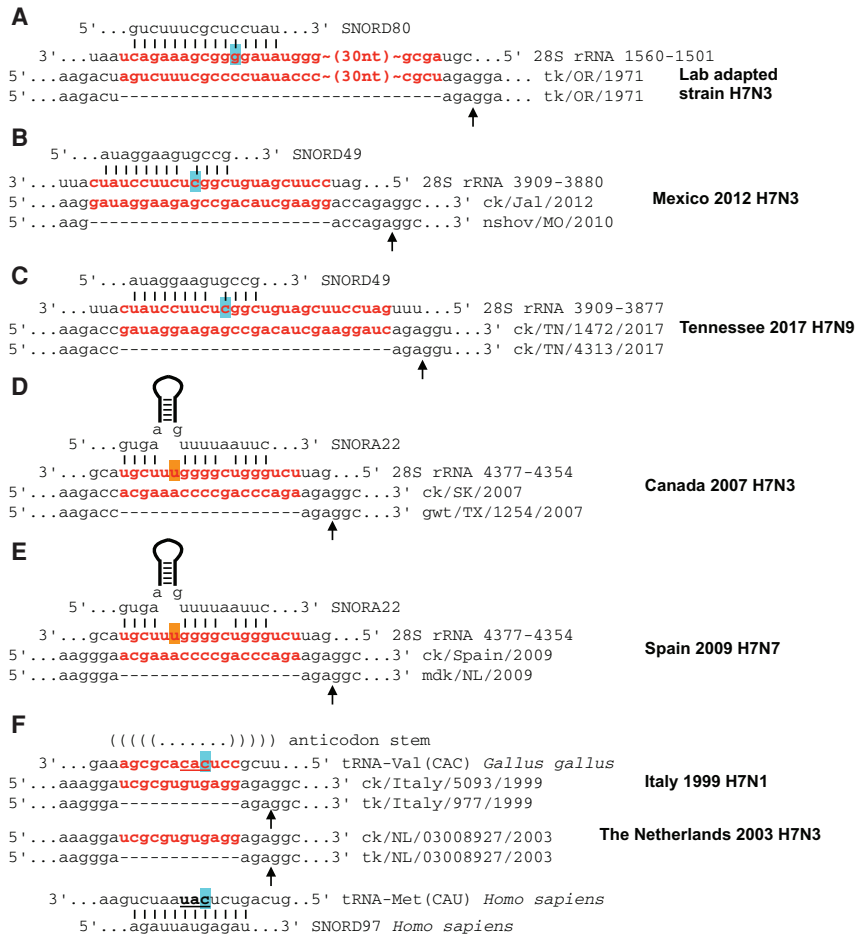
The origins of insertions of 18 nt in the HA segments of H7 HPAI strains isolated in Canada in 2007 (A/chicken/SK/HR-00011/2007 [H7N3]) and in Spain in 2009 (A/chicken/Spain/6279-2/2009 [H7N7]) have so far remained unidentified (Berhane et al. 2009; Iglesias et al. 2010). These insertions are the same in both HA sequences (accessions EU500860 and GU121458, respectively), although the viruses have distinct origins. Alignment of the inserted sequence with chicken rRNAs (accessions KT445934 and XR\_003078040) using the Smith–Waterman algorithm showed that it was reverse complementary to the chicken 28S rRNA region 4357–4374. A BLAST search in the GenBank Nucleotide collection (nr/nt) retrieved the same identity, accompanied by multiple hits to homologous rRNA fragments in other species that are 100% conserved.

The E-value of 6.5, calculated by the BLAST program for these hits, defined them as nonsignificant, as this value corresponds to the number of hits with at least a given score, expected by chance alone. However, a BLAST search with the same 18 nt against the reference chicken genome yielded the same alignment to 28S rRNA with a considerably lower E-value of 0.027, with no other hits with 100% identity. Thus, in the reasonable assumption that the insertions resulted from HA recombination with one of the host (chicken) RNAs, one can conclude that both recombination events leading to the 2007 Canadian and 2009 Spanish HPAI involved the 28S rRNA.

Together with the previously identified cases (Khatchikian et al. 1989; Maurer-Stroh et al. 2013; Lee et al. 2017), this brings the total number of recorded rRNA recombination events of HA segments with 28S rRNA to five (Table 1). As observed in the 2007 Canadian and 2009 Spanish HPAI, the insertion observed in A/chicken/Tennessee/17-007147-2/2017 (H7N9) virus involved the same recombination site as in the Mexican H7N3 HPAI viruses (Maurer-Stroh et al. 2013; Lee et al. 2017). Therefore, two regions of rRNA were involved in four independent HPAI emergences, suggesting that recombination sites in the 28S rRNA are not random.

Interestingly, all three recombination points in the 28S rRNA turned out to be located within 1 nt of the 3' ends of known snoRNA binding sites (Fig. 2A–E): either 1 nt downstream from a region paired to a C/D box snoRNA, or just adjacent to the region paired to H/ACA box snoRNA. The probability of this to be due to chance alone is low: with 105 identified snoRNA targets (see Materials and Methods and Supplemental Table S1) in chicken 28S rRNA of 4437 nucleotides, the probability of three independent sites to be at or 1 nt downstream from a snoRNA/rRNA duplex is  $(2 \times 105/4437)^3 \approx 10^{-4}$ .

The optimal Smith–Waterman alignments of the insertions of 9–12 nt in H7 HA segments of HPAI viruses from Italy, 1999–2001 (H7N1) (Banks et al. 2001; Monne et al. 2014), England, 2015 (H7N7) (Abdelwhab et al. 2016), China, 2017 (H7N9) (Chen et al. 2017), or laboratory-adapted H7N7 strain SC35 (Li et al. 1990) with chicken rRNAs and other segments of the corresponding strains contained mismatches and therefore did not allow one to draw conclusions on the origins of these insertions. Using the insertions of 9 nt for BLAST searches in the chicken genome was not possible because such short queries were rejected by the BLAST program. It is possible to estimate the E-value of such a search in chicken genome, using a general expression  $E = K \times m \times n \times \exp(-\lambda S)$  for gap-free alignments with the scores of at least  $S$ , where  $m$  and  $n$  are the lengths of aligned sequences and  $K$  and  $\lambda$  depend on scoring parameters and nucleotide composition (see, e.g., Altschul et al. 1990). Thus, the pairwise BLAST alignment of the insertion of 9 nt in the HA of A/chicken/England/26352/2015(H7N7) versus the full-length HA



**FIGURE 2.** Recombination of the influenza virus H7 HA RNA with host RNAs. (A–E) The insertions in the HA segments created by recombination with 28S rRNA and the binding sites of snoRNAs near the recombination sites. The HA RNA sequences from HPAI strains (Khatchikian et al. 1989; Banks et al. 2001; Berhane et al. 2009; Iglesias et al. 2010; Maurer-Stroh et al. 2013; Lee et al. 2017; Beerens et al. 2020) are aligned to HA sequences of most closely related LPAI viruses. The apical stem between two parts of the SNORA22 antisense element is shown schematically. (F) The 12-nt sequence inserted in the Italian H7N1 (1999) and Dutch H7N3 (2003) HPAI strains is complementary to the region encompassing anticodon of chicken tRNA-Val. Base pairs of the tRNA-Val anticodon stem are shown with bracket view. Anticodon nucleotides are underlined. Chicken tRNA-Val is aligned to human tRNA-Met in order to illustrate the location of the C/D box snoRNA binding site in the tRNA-Met (Vitali and Kiss 2019). Inserted sequences and their complements in 28S rRNA and tRNA are shown in red. Locations between the codons for conserved amino acids Arg and Gly, corresponding to the HA0 polyprotein cleavage sites, are indicated by arrows. The 2'-O-methylated nucleotides are labeled with cyan, the pseudouridylated uridines with orange. For strain abbreviations, see Materials and Methods.

sequence of 1704 nucleotides (accession EPI\_ISL\_191604) yielded an E-value of 0.015. As the HA sequence is not characterized by a biased nucleotide composition, the E-value for a search in the chicken genome assembly GRCg6a (accession GCA\_000002315.5) with ungapped length of 1055.6 Mb can be estimated as proportional to the genome length. The obtained E-value of more than 9000 confirmed that such a BLAST search with a 9 nt sequence makes no sense.

BLAST searches with the insertions of 12 nt of Chinese H7N9 and Italian H7N1 strains in the reference chicken genome yielded 100%-identity hits with a reported E-value of 105. However, the actual number of hits was significantly lower than 105: 14 (taking into account two alternative variants differing by 1-nt shift) and 12, respectively, with only 10 corresponding to transcribed gene regions in each case. Remarkably, six out of 10 hits of the Italian H7N1 insertion were located in tRNA-Val (CAC) genes, suggesting that the insertion most likely originated from one of them, with the sequence complementary to the anticodon loop and part of the anticodon stem (Fig. 2F). Furthermore, the same sequence was inserted into the HA RNA of LPAI virus precursors in turkeys during an outbreak in three Dutch farms in 2002 and selected rapidly upon experimental infections of chickens in 2003 (Fig. 2F; Velkers et al. 2006; Beerens et al. 2020).

One of the important factors explaining why rRNAs, and probably tRNAs are relatively frequently involved in recombination with influenza HA RNA is the high abundance of these RNAs in eukaryotic cells. Indeed, about 85%–90% of the total mass of transcripts is rRNA molecules, followed by tRNAs (up to 10%), the latter being the most abundant RNA type in terms of number of molecules (e.g., Waldron and Lacroute 1975; Boivin et al. 2018). Recombination with influenza virus RNA from other genes may similarly be explained by their high abundance in virus-infected cells. Furthermore, 18S and 28S rRNAs have been shown to be packaged into influenza A virions (Noda et al. 2018) and thus have a high probability to be in close proximity to the vRNA. SnoRNAs are also a highly abundant RNA species, and are frequently used by influenza A virus polymerase in cap-snatching during transcription (Gu et al. 2015; Koppstein et al. 2015; Sikora et al. 2017), although it should be noted that all three snoRNAs binding to the 28S rRNA recombination sites (Fig. 2A–E) are encoded in introns of other genes and therefore uncapped.

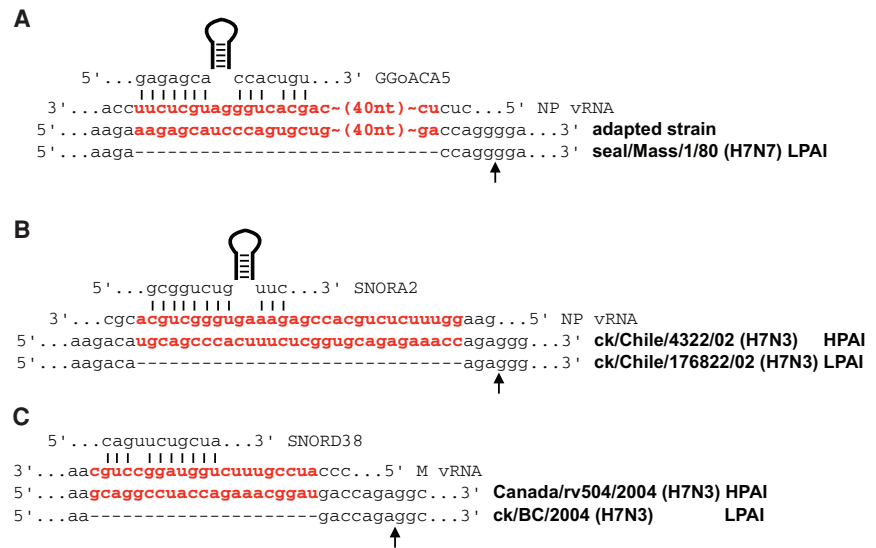


strongest with  $-28.1$  kcal/mol, which was the fourth strongest duplex out of 53 chicken C/D box snoRNA duplexes formed with 28S rRNA, ranging between  $-28.8$  and  $-12.0$  kcal/mol. However, the rank of SNORD49 pairing to rRNA was only 29 with  $-22.4$  kcal/mol, and SNORA22 formed the weakest binding duplexes out of 52 identified binding sites of H/ACA box snoRNAs to 28S rRNA:  $-3.8$  and  $-3.0$  kcal/mol.

The presence of mismatches was a feature of snoRNA/rRNA duplexes that seemed to correlate with repeated involvement of the 28S rRNA sites near SNORD49 and SNORA22 binding in the recombination with HA RNA (Fig. 2). Only five other chicken snoRNAs, namely, SNORD12/SNORD106, SNORD24, SNORD50, SNORA17 (GGN58), and SNORA63, formed duplexes with mismatches upon binding to 28S rRNA. Moreover, out of these five duplexes only the SNORA17 pairing and subsequent recombination with HA RNA would result in a furin cleavage signal in the HA polyprotein. A mismatch in the snoRNA/rRNA pairing interaction might play a role in rRNA recombination at the binding site due to a faster dissociation of the snoRNA, which would be required to use the binding region as a template by the polymerase. Kinetics of snoRNA release from the rRNA recombination site can be important to allow for template switching during a time window of influenza polymerase slowdown, which is likely to occur in the cleavage site region due to its structured character (Gulyaev et al. 2019).

Structural constraints in this HA RNA domain (Gulyaev et al. 2016, 2019) may also determine the fitness of the viruses containing the insertions. The insertions occur in the proposed hairpin loop of the conserved H7 HA RNA structure, stabilizing it by base pairs formed by inserted sequences (Gulyaev et al. 2016). However, predictions of possible structures of this domain with inserted sequences from other rRNA regions showed similar stabilization (not shown), so that we could not derive any specific feature characteristic for SNORD49 and SNORA22 targets, which would explain their repeated involvement in the recombination with HA RNA.

Interestingly, we identified chicken snoRNAs that could bind to the known sites (Table 1) of recombination of NP and M segments with HA vRNA (Fig. 4). As a number of snoRNA functions are determined by their targeting various RNAs (Bachelier et al. 2002; Bratkovič et al. 2019; Bergeron et al. 2020), it is possible to suggest that emergence of HPAI strains by HA recombination with RNA mol-



**FIGURE 4.** Putative snoRNA binding at the recombination sites of HA RNAs with other influenza A virus segments. The sequences of NP and M segments involved in the recombination are from the LPAI progenitors: (A) Orlich et al. (1994); (B) Suarez et al. (2004); (C) Hirst et al. (2004). Chicken orphan H/ACA snoRNA GGoACA5 has been identified by Shao et al. (2009). Notations are similar to those in Figure 2.

ecules other than rRNA can be also triggered by snoRNA binding.

## DISCUSSION

Here, we identified that for all five described cases of recombination of H7 HA RNA with chicken rRNA, the recombination point was located near snoRNA binding sites. Interestingly, two snoRNAs (SNORD22 and SNORD49) were involved in two independent cases of recombination. We therefore hypothesize that recombination of HA segments with rRNA occurs at the hotspots determined by snoRNA/rRNA duplexes. Recombination may be facilitated by polymerase slowdown induced by the template secondary structure around the recombination site (Gulyaev et al. 2019). Upon release of snoRNA, the rRNA is exploited as a template by a copy-choice mechanism, starting from the 3' end of the snoRNA binding site. After the second template switching back to the viral template RNA, a nascent viral cRNA with an rRNA-derived insertion is produced. If the insertion encodes a MBCS at the HA1/HA2 boundary in the encoded polypeptide and does not considerably interfere with virus fitness, a highly pathogenic strain can emerge. Presumably, similar recombination events may occur in other regions of HA RNA, but they remain undetected because distorted HA protein structures make the resulting virions not viable.

Mechanistic details of such a recombination remain unclear. During influenza virus RNA replication the template and nascent strand regions near the elongation point are

located in the channels formed by the polymerase subunits (te Velthuis and Fodor 2016; Pflug et al. 2017; Fodor and te Velthuis 2020). Recombination would probably require a temporary disruption of some interactions between and/or within the subunits to accommodate another template in a copy-choice mechanism. Alternatively, an involvement of another polymerase molecule cannot be excluded. It should be noted that noncanonical copying of RNA templates by DNA-dependent RNA polymerases, including rRNA-transcribing RNA polymerase I, has been observed (Macnaughton et al. 2002; Jain et al. 2020). It is also possible that the recombination hotspots at the snoRNA binding sites in rRNA and vRNA are determined by a duplex-specific cleavage that could yield the free 3' ends of the targeted regions. We cannot exclude a role of snoRNA-guided modifications in the structure and/or accessibility of targeted regions, such as structure stabilization and additional hydrogen bond acceptor upon pseudouridylation or stabilization of transient alternative conformations upon 2'-O-methylation (Charette and Gray 2000; Abou Assi et al. 2020).

Similar to HA recombination with 28S rRNA, the suggested recombinations with tRNA in the Italian H7N1 and Dutch H7N3 HPAI strains would occur during cRNA synthesis, as in all these cases the fragments inserted into the coding sequences are complementary to host noncoding RNAs (Fig. 2). Although tRNA modifications in eukaryotes are mostly snoRNA-independent, potential snoRNA targets have been identified in nematode tRNAs (Zemann et al. 2006) and a precedent of a C/D box snoRNA-tRNA interaction guiding 2'-O-methylation of wobble nucleotide in the anticodon of human tRNA has been recently set (Vitali and Kiss 2019). It is tempting to speculate that the insertions of the 12-nt tRNA fragment in HA RNAs could be triggered by an interaction of the similar region in avian tRNA-Val with yet unknown C/D box snoRNA. A putative homologous binding site in chicken tRNA-Val would be located 1 nt upstream of the terminal nucleotide of the sequence inserted in HPAI strains, exactly like the SNORD80 and SNORD49 binding sites in 28S rRNA (Fig. 2A,B,C,F). Although we did not identify any chicken snoRNA with a guide sequence that could target the tRNA-Val anticodon, such a snoRNA may exist as many snoRNAs remain unannotated (Gardner et al. 2010; de Araujo Oliveira et al. 2016; Bratkovič et al. 2019; Bergeron et al. 2020). Also, we did not identify any snoRNA that could target the recombination points of the insertions in HA of H5 subtype, detected for the first time by ultra-deep sequencing upon serial H5N2 virus passaging (Laleye and Abolnik 2020).

If true, the preference for snoRNA-bound locations in template switching by the influenza A virus polymerase is an important part of the mechanism of the emergence of HPAI H7 strains by recombination of virus HA segments with other RNAs. The hypothesis provides a rationale be-

hind the existence of hotspots in recombination with rRNA. Future experimental studies could elucidate the details of this mechanism.

## MATERIALS AND METHODS

### Sequences and alignments

The strains of HPAI viruses yielded by HA recombination and their likely LPAI progenitors were identified from literature (Khatchikian et al. 1989; Orlich et al. 1994; Banks et al. 2001; Hirst et al. 2004; Suarez et al. 2004; Berhane et al. 2009; Iglesias et al. 2010; Maurer-Stroh et al. 2013; Monne et al. 2014; Lee et al. 2017; Beerens et al. 2020) and/or BLAST (Camacho et al. 2009) searches. Accessions: A/turkey/Oregon/1971(H7N3) (abbreviated here as tk/OR/1971), M31689; A/chicken/Jalisco/CPA1/2012(H7N3) (ck/Jal/2012), JX397993; A/northern/shoveler/Missouri/100S4750/2010(H7N3) (nshov/MO2010), CY133416; A/chicken/Tennessee/17-007147-2/2017(H7N9) (ck/TN/1472/2017), KY818811; A/chicken/Tennessee/17-007431-3/2017(H7N9) (ck/TN/4313/2017), KY818819; A/chicken/SK/HR-00011/2007(H7N3) (ck/SK/2007), EU500860; A/green-winged teal/TX/1254/2007(H7N3) (gwt/TX/1254/2007), KF573723; A/chicken/Spain/6279-2/2009(H7N7) (ck/Spain/2009), GU121458; A/mallard duck/Netherlands/2/2009(H7N7) (mdk/NL/2009), KX978114; A/chicken/Italy/5093/99(H7N1) (ck/Italy/5093/1999), DQ991328; A/turkey/Italy/977/1999(H7N1) (tk/Italy/977/1999), GU052999; A/chicken/Netherlands/03008927/2003(H7N3) (ck/NL/03008927), EPI\_ISL\_391798; A/turkey/Netherlands/03008927/2003(H7N3) (tk/NL/03008927), EPI\_ISL\_391799; A/chicken/Chile/4322/02(H7N3) (ck/Chile/4322/02), AY303631; A/chicken/Chile/176822/02(H7N3) (ck/Chile/176822/02), AY303630; A/Canada/rv504/2004(H7N3) (Canada/rv504/2004), CY015006; A/GSC\_chicken/British Columbia/04(H7N3) (ck/BC/2004), AY650270. Chicken 28S rRNA sequence: accession XR\_003078040. Sequences of tRNAs were retrieved from the tRNAdb database (Jühling et al. 2009). Alignments of insertions versus rRNA or influenza virus RNAs were carried out by the Smith-Waterman algorithm of the program Water provided by the EMBL-EBI tools (Madeira et al. 2019).

Available chicken snoRNAs that target 28S rRNA were initially retrieved from the snoRNA databases: snoRNABase (Lestrade and Weber 2006) and snOPY (Yoshihama et al. 2013). This list was complemented by published records of chicken snoRNA sequences (Shao et al. 2009; Zhang et al. 2009; Deryusheva et al. 2020). The final nonredundant data set of chicken snoRNA sequences and structures was verified by identification of GenBank database entries, genomic sequences (using BLAST searches in the chicken genome) and structural models from the database of RNA families Rfam (Kalvari et al. 2018). In case of alternative names for the same snoRNA, the symbols approved by the HUGO Gene Nomenclature Committee (HGNC) (Braschi et al. 2019) were preferably used, that is, SNORD... and SNORA... for C/D and H/ACA box snoRNAs, respectively. Sequence accessions of the snoRNAs that bind to 28S rRNA near the points of recombination with HA RNA: SNORD80 (or snR60), XR\_003076593; SNORD49 (or GGgCD65), XR\_003071336; SNORA22 (or GGN50), XR\_003071339. SnoRNAs that may bind to NP and M segments near the points of recombination: SNORD38,

XR\_003076587; SNORA2, NC\_008465 (chromosome 33), positions 3350951-3351091 (complement); GGoACA5, NC\_006112 (chromosome 25), 3660678-3660803 (complement).

### Analysis of snoRNA–rRNA duplexes

The binding sites of chicken snoRNAs in 28S rRNA were identified by analogy to duplexes identified for human snoRNAs, given in the snoRNABase (Lestrade and Weber 2006) and using the identification of guide sequences in snoRNAs that are complementary to chicken 28S rRNA. Free energies of RNA duplexes were calculated using thermodynamic parameters from the Nearest Neighbor Database (Turner and Mathews 2010).

### SnoRNA expression levels

The expression levels of chicken snoRNAs were obtained from the data given by Expression Atlas incorporated in the Ensembl genomic browser (Papatheodorou et al. 2018; Yates et al. 2020).

### SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

### ACKNOWLEDGMENTS

This work was supported by National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIAID/NIH) contract HHSN272201400008C. This work has also received funding from the European Union's Horizon 2020 research and innovation program under DELTA-FLU, grant agreement no. 727922. We gratefully acknowledge the originating and submitting laboratories for sequences in the GISAID's EpiFlu Database, used in this study.

Received July 27, 2020; accepted November 6, 2020.

### REFERENCES

- Abdelwhab EM, Veits J, Mettenleiter TC. 2013. Genetic changes that accompanied shifts of low pathogenic avian influenza viruses toward higher pathogenicity in poultry. *Virulence* **4**: 441–452. doi:10.4161/viru.25710
- Abdelwhab EM, Veits J, Ulrich R, Kasbohm E, Teifke JP, Mettenleiter TC. 2016. Composition of the hemagglutinin polybasic proteolytic cleavage motif mediates variable virulence of H7N7 avian influenza viruses. *Sci Rep* **6**: 39505. doi:10.1038/srep39505
- Abou Assi H, Rangadurai AK, Shi H, Liu B, Clay MC, Erharter K, Kreutz C, Holley CL, Al-Hashimi HM. 2020. 2'-O-methylation can increase the abundance and lifetime of alternative RNA conformational states. *Nucleic Acids Res* **48**: 12365–12379. doi:10.1093/nar/gkaa928
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**: 403–410. doi:10.1016/S0022-2836(05)80360-2
- Bachelier JP, Cavaille J, Huttenhofer A. 2002. The expanding snoRNA world. *Biochimie* **84**: 775–790. doi:10.1016/S0300-9084(02)01402-5
- Banks J, Speidel ES, Moore E, Plowright L, Piccirillo A, Capua I, Cordioli P, Fioretti A, Alexander DJ. 2001. Changes in the haemagglutinin and the neuraminidase genes prior to the emergence of highly pathogenic H7N1 avian influenza viruses in Italy. *Arch Virol* **146**: 963–973. doi:10.1007/s007050170128
- Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, et al. 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**: 1587–1593. doi:10.1126/science.1230612
- Berens N, Heutink R, Harders F, Bossers A, Koch G, Peeters B. 2020. Emergence and selection of a highly pathogenic avian influenza H7N3 virus. *J Virol* **94**: e01818-19. doi:10.1128/JVI.01818-19
- Bergeron D, Fafard-Couture E, Scott MS. 2020. Small nucleolar RNAs: continuing identification of novel members and increasing diversity of their molecular mechanisms of action. *Biochem Soc Trans* **48**: 645–656. doi:10.1042/BST20191046
- Berhane Y, Hisanaga T, Kehler H, Neufeld J, Manning L, Argue C, Handel K, Hooper-McGrevy K, Jonas M, Robinson J, et al. 2009. Highly pathogenic avian influenza virus A (H7N3) in domestic poultry, Saskatchewan, Canada, 2007. *Emerg Infect Dis* **15**: 1492–1495. doi:10.3201/eid1509.080231
- Boivin V, Deschamps-Francoeur G, Couture S, Nottingham RM, Bouchard-Bourelle P, Lambowitz AM, Scott MS, Abou-Elela S. 2018. Simultaneous sequencing of coding and noncoding RNA reveals a human transcriptome dominated by a small number of highly expressed noncoding genes. *RNA* **24**: 950–965. doi:10.1261/ma.064493.117
- Braschi B, Denny P, Gray K, Jones T, Seal R, Tweedie S, Yates B, Bruford E. 2019. Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Res* **47**: D786–D792. doi:10.1093/nar/gky930
- Bratkovič T, Bozic J, Rogelj B. 2019. Functional diversity of small nucleolar RNAs. *Nucleic Acids Res* **48**: 1627–1651. doi:10.1093/nar/gkz1140
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421. doi:10.1186/1471-2105-10-421
- Charette M, Gray MW. 2000. Pseudouridine in RNA: what, where, how, and why. *IUBMB Life* **49**: 341–351. doi:10.1080/152165400410182
- Chen J, Lee KH, Steinhauer DA, Stevens DJ, Skehel JJ. 1998. Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. *Cell* **95**: 409–417. doi:10.1016/S0092-8674(00)81771-7
- Chen J, Zhang J, Zhu W, Zhang Y, Tan H, Liu M, Cai M, Shen J, Ly H, Chen J. 2017. First genome report and analysis of chicken H7N9 influenza viruses with poly-basic amino acids insertion in the hemagglutinin cleavage site. *Sci Rep* **7**: 9972. doi:10.1038/s41598-017-10605-6
- de Araujo Oliveira JV, Costa F, Backofen R, Stadler PF, Walter MEMT, Hertel J. 2016. SnoReport 2.0: new features and a refined support vector machine to improve snoRNA identification. *BMC Bioinformatics* **17**: 464. doi:10.1186/s12859-016-1345-6
- Deryusheva S, Talhouarne GJS, Gall JG. 2020. “Lost and Found”: snoRNA Annotation in the *Xenopus* genome and implications for evolutionary studies. *Mol Biol Evol* **37**: 149–166. doi:10.1093/molbev/msz209
- Duckert P, Brunak S, Blom N. 2004. Prediction of proprotein convertase cleavage sites. *Prot Eng Des Sel* **17**: 107–112. doi:10.1093/protein/gzh013
- Fodor E, te Velthuis AJW. 2020. Structure and function of the influenza virus transcription and replication machinery. *Cold Spring Harb Perspect Med* **10**: a038398. doi:10.1101/cshperspect.a038398
- Fouchier RAM, Schneeberger PM, Rozendaal FW, Broekman JM, Kemink SAG, Munster V, Kuiken T, Rimmelzwaan GF, Schutten M, van Doornum GJJ, et al. 2004. Avian influenza A virus



- (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. *Proc Natl Acad Sci* **101**: 1356–1361. doi:10.1073/pnas.0308352100
- Gardner PP, Bateman A, Poole AM. 2010. SnoPatrol: how many snoRNA genes are there? *J Biol* **9**: 4. doi:10.1186/jbiol211
- Garten W, Braden C, Arendt A, Peitsch C, Baron J, Lu Y, Pawletko K, Hards K, Steinmetzer T, Böttcher-Friebertshäuser E. 2015. Influenza virus activating host proteases: identification, localization and inhibitors as potential therapeutics. *Eur J Cell Biol* **94**: 375–383. doi:10.1016/j.ejcb.2015.05.013
- Gu W, Gallagher GR, Dai W, Liu P, Li R, Trombly MI, Gammon DB, Mello CC, Wang JP, Finberg RW. 2015. Influenza A virus preferentially snatches noncoding RNA caps. *RNA* **21**: 2067–2075. doi:10.1261/rna.054221.115
- Gultyaev AP, Spronken MI, Richard M, Schrauwen EJA, Olsthoorn RCL, Fouchier RAM. 2016. Subtype-specific structural constraints in the evolution of influenza A virus hemagglutinin genes. *Sci Rep* **6**: 38892. doi:10.1038/srep38892
- Gultyaev AP, Richard M, Spronken MI, Olsthoorn RCL, Fouchier RAM. 2019. Conserved structural RNA domains in regions coding for cleavage site motifs in hemagglutinin genes of influenza viruses. *Virus Evol* **5**: vez034. doi:10.1093/ve/vez034
- Hirst M, Astell CR, Griffith M, Coughlin SM, Moksa M, Zeng T, Smailus DE, Holt RA, Jones S, Marra MA, et al. 2004. Novel avian influenza H7N3 strain outbreak, British Columbia. *Emerg Infect Dis* **10**: 2192–2195. doi:10.3201/eid1012.040743
- Iglesias I, Martínez M, Muñoz MJ, de la Torre A, Sánchez-Vizcaino JM. 2010. First case of highly pathogenic avian influenza in poultry in Spain. *Transbound Emerg Dis* **57**: 282–285. doi:10.1111/j.1865-1682.2010.01145.x
- Jain N, Blauch LR, Szymansky MR, Das R, Tang SKY, Yin YW, Fire AZ. 2020. Transcription polymerase-catalyzed emergence of novel RNA replicons. *Science* **368**: eaay0688. doi:10.1126/science.aay0688
- Jühling F, Mörl M, Hartmann RK, Sprinzl M, Stadler PF, Pütz J. 2009. tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* **37**: D159–D162. doi:10.1093/nar/gkn772
- Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD, Petrov AI. 2018. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res* **46**: D335–D342. doi:10.1093/nar/gkx1038
- Khatchikian D, Orlich M, Rott R. 1989. Increased viral pathogenicity after insertion of a 28S ribosomal RNA sequence into the haemagglutinin gene of an influenza virus. *Nature* **340**: 156–157. doi:10.1038/340156a0
- Koppstein D, Ashour J, Bartel DP. 2015. Sequencing the cap-snatching repertoire of H1N1 influenza provides insight into the mechanism of viral transcription initiation. *Nucleic Acids Res* **43**: 5052–5064. doi:10.1093/nar/gkv333
- Laleye AT, Abolnik C. 2020. Emergence of highly pathogenic H5N2 and H7N1 influenza A viruses from low pathogenic precursors by serial passage *in ovo*. *PLoS One* **15**: e0240290. doi:10.1371/journal.pone.0240290
- Lee CW, Swayne DE, Linares JA, Senne DA, Suarez DL. 2005. H5N2 avian influenza outbreak in Texas in 2004: the first highly pathogenic strain in the United States in 20 years? *J Virol* **79**: 11421–11421. doi:10.1128/JVI.79.17.11421-11421.2005
- Lee DH, Torchetti MK, Killian ML, Berhane Y, Swayne DE. 2017. Highly pathogenic avian influenza A(H7N9) virus, Tennessee, USA, March 2017. *Emerg Infect Dis* **23**: 1860–1863. doi:10.3201/eid2311.171013
- Lee DH, Criado MF, Swayne DE. 2020. Pathobiological origins and evolutionary history of highly pathogenic avian influenza viruses. *Cold Spring Harb Perspect Med* doi:10.1101/cshperspect.a038679
- Lestrade L, Weber MJ. 2006. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res* **34**: D158–D162. doi:10.1093/nar/gkj002
- Li S, Orlich M, Rott R. 1990. Generation of seal influenza virus variants pathogenic for chickens, because of hemagglutinin cleavage site changes. *J Virol* **64**: 3297–3303. doi:10.1128/JVI.64.7.3297-3303.1990
- Macnaughton TB, Shi ST, Modahl LE, Lai MMC. 2002. Rolling circle replication of hepatitis delta virus RNA is carried out by two different cellular RNA polymerases. *J Virol* **76**: 3920–3927. doi:10.1128/JVI.76.8.3920-3927.2002
- Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, et al. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**: W636–W641. doi:10.1093/nar/gkz268
- Maurer-Stroh S, Lee RT, Gunalan V, Eisenhaber F. 2013. The highly pathogenic H7N3 avian influenza strain from July 2012 in Mexico acquired an extended cleavage site through recombination with host 28S rRNA. *Virology* **453**: 139. doi:10.1016/j.virus.2013.10.139
- Merkin J, Russell C, Chen P, Burge CB. 2012. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* **338**: 1593–1599. doi:10.1126/science.1228186
- Monne I, Fusaro A, Nelson MI, Bonfanti L, Mulatti P, Hughes J, Murcia PR, Schivo A, Valastro V, Moreno A, et al. 2014. Emergence of a highly pathogenic avian influenza virus from a low-pathogenic progenitor. *J Virol* **88**: 4375–4388. doi:10.1128/JVI.03181-13
- Naguib MM, Verhagen JH, Mostafa A, Wille M, Li R, Graaf A, Järhult JD, Ellström P, Zohari S, Lundkvist Å, et al. 2019. Global patterns of avian influenza A(H7): virus evolution and zoonotic threats. *FEMS Microbiol Rev* **43**: 608–621. doi:10.1093/femsre/fuz019
- Noda T, Murakami S, Nakatsu S, Imai H, Muramoto Y, Shindo K, Sagara H, Kawaoka Y. 2018. Importance of the 1+7 configuration of ribonucleoprotein complexes for influenza A virus genome packaging. *Nat Commun* **9**: 54. doi:10.1038/s41467-017-02517-w
- Orlich M, Gorrwald H, Rott R. 1994. Nonhomologous recombination between the hemagglutinin gene and the nucleoprotein gene of an influenza virus. *Virology* **204**: 462–465. doi:10.1006/viro.1994.1555
- Papatheodorou I, Fonseca NA, Keays M, Tang YA, Barrera E, Bazant W, Burke M, Füllgrabe A, Fuentes AM, George N, et al. 2018. Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res* **46**: D246–D251. doi:10.1093/nar/gkx1158
- Peng X, Gralinski L, Ferris MT, Frieman MB, Thomas MJ, Proll S, Korth MJ, Tisoncik JR, Heise M, Luo S, et al. 2011. Integrative deep sequencing of the mouse lung transcriptome reveals differential expression of diverse classes of small RNAs in response to respiratory virus infection. *MBio* **2**: e00198-11. doi:10.1128/mBio.00198-11
- Pflug A, Lukarska M, Resa-Infante P, Reich S, Cusack S. 2017. Structural insights into RNA synthesis by the influenza virus transcription-replication machine. *Virus Res* **234**: 103–117. doi:10.1016/j.virusres.2017.01.013
- Samir M, Vidal RO, Abdallah F, Capece V, Seehusen F, Geffers R, Ali AAH, Bonn S, Pessler F. 2020. Organ-specific small non-coding RNA responses in domestic (Sudani) ducks experimentally infected with highly pathogenic avian influenza virus (H5N1). *RNA Biol* **17**: 112–124. doi:10.1080/15476286.2019.1669879
- Shao P, Yang JH, Zhou H, Guan DG, Qu LH. 2009. Genome-wide analysis of chicken snoRNAs provides unique implications for the evolution of vertebrate snoRNAs. *BMC Genomics* **10**: 86. doi:10.1186/1471-2164-10-86

- Sikora D, Rocheleau L, Brown EG, Pelchat M. 2017. Influenza A virus cap-snatches host RNAs based on their abundance early after infection. *Virology* **509**: 167–177. doi:10.1016/j.virol.2017.06.020
- Smith CM, Steitz JA. 1997. Sno storm in the nucleolus: new roles for myriad small RNPs. *Cell* **89**: 669–672. doi:10.1016/S0092-8674(00)80247-0
- Stamm S, Lodmell JS. 2019. C/D box snoRNAs in viral infections: RNA viruses use old dogs for new tricks. *Non-coding RNA Res* **4**: 46–53. doi:10.1016/j.ncrna.2019.02.001
- Suarez DL, Senne DA, Banks J, Brown IH, Essen SC, Lee CW, Manvell RJ, Mathieu-Benson C, Moreno V, Pedersen JC, et al. 2004. Recombination resulting in virulence shift in avian influenza outbreak, Chile. *Emerg Infect Dis* **10**: 693–699. doi:10.3201/eid1004.030396
- te Velthuis AJW, Fodor E. 2016. Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. *Nat Rev Microbiol* **14**: 479–483. doi:10.1038/nrmicro.2016.87
- Turner DH, Mathews DH. 2010. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res* **38**: D280–D282. doi:10.1093/nar/gkp892
- Velkers FC, Bouma A, Matthijs MGR, Koch G, Westendorp T, Stegeman JA. 2006. Outbreak of avian influenza H7N3 on a turkey farm in the Netherlands. *Vet Rec* **159**: 403–405. doi:10.1136/vr.159.13.403
- Vitali P, Kiss T. 2019. Cooperative 2'-O-methylation of the wobble cytidine of human elongator tRNA<sup>Met</sup>(CAT) by a nucleolar and a Cajal body: specific box c/D RNP. *Genes Dev* **33**: 741–746. doi:10.1101/gad.326363.119
- Waldron C, Lacroute F. 1975. Effect of growth rate on the amounts of ribosomal and transfer ribonucleic acids in yeast. *J Bacteriol* **122**: 855–865. doi:10.1128/JB.122.3.855-865.1975
- Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, et al. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**: W296–W303. doi:10.1093/nar/gky427
- Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, et al. 2020. Ensembl 2020. *Nucleic Acids Res* **48**: D682–D688. doi:10.1093/nar/gkz1138
- Yoshihama M, Nakao A, Kenmochi N. 2013. snOPY: a small nucleolar RNA orthological gene database. *BMC Res Notes* **6**: 426. doi:10.1186/1756-0500-6-426
- Zemann A, op de Bekke A, Kiefmann M, Brosius J, Schmitz J. 2006. Evolution of small nucleolar RNAs in nematodes. *Nucleic Acids Res* **34**: 2676–2685. doi:10.1093/nar/gkl359
- Zhang Y, Wang J, Huang S, Zhu X, Liu J, Yang N, Song D, Wu R, Deng W, Skogerbo G, et al. 2009. Systematic identification and characterization of chicken (*Gallus gallus*) ncRNAs. *Nucleic Acids Res* **37**: 6562–6574. doi:10.1093/nar/gkp704