Translational Decoding in

Viral Infection and Cancer Development

Xumin Ou

欧旭敏　著

Cover design: the author of this thesis
Layout design: the author of this thesis

Translational Decoding in Viral Infection and Cancer Development

Uitlezen van translatiecodons en
Virale infectie en kankerontwikkeling

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof. Dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on

Tuesday 16th November 2021 at 13:00 pm

by

Xumin Ou

born in Qingshen, Sichuan Province, China.

Doctoral Committee:

Promotor:	Prof. Dr. M. P. Peppelenbosch
Inner committee:	Prof. Dr. L. J. W. van der Laan
	Prof. Dr. B. Berkhout
	Dr. T. Mahmoudi
Copromotor:	Dr. Q. Pan

TABLE OF CONTENTS

PART I: Role of tRNA biology in viral infection and cancer development

PART II: The implications of translational decoding in viral cross-species infection

Chapter 1

General Introduction and Outline of this Thesis

Viruses are common causes of various diseases. These pathogens are nanometer-size infectious particles that replicate inside the viable cells of an organism and depend on host resources for their replication. According to an estimation of the International Committee on Taxonomy of Viruses, currently, there are more than 9,000 archived viral species [1]. These viruses can infect nearly all domains of life, such as mammals, avians, fishes, reptiles, insects, plants, and even bacteria (i.e., by bacteriophages). Understanding virus-host interaction is important, as this defines how viruses replicate themselves within host organisms at different levels, such as at a molecular, cellular, and organismal level. At the molecular level, the central dogma of molecular biology defines the flow of genetic information. The dogma states that the genetic information is stored in the cellular DNA, and subsequently transcribed into RNA. RNA is finally translated into proteins that are effectors of the genetic information in cellular physiology. The central dogma is at the basis of modern life sciences, including virology. Viruses, however, are among the simplest microorganism in existence and do not have the transcription and/or translation machinery necessary to perform the actions contained in the central dogma and thus depend on the host cell for executing the flow of their genetic information. The viral particle (also known as virion) is an assembly of genetic material (in the form of DNA or RNA) and the viral genetic material-encapsulating capsids, whereas in certain viruses there is also a phospholipid bilayer envelope. Of note, the only heritable materials in virion are the viral DNA or RNA, which in essence is an assembly of genetic codes [2]. The combination of genetic codes in a viral gene determines the amino acid sequence of corresponding proteins, such as the capsid or viral polymerases. Faithful and efficient decoding of the viral genetic information is crucial for viral reproduction and hence understanding of viral biology requires insights into the mechanisms that mediate this decoding.

The course of viral infection generally follows six stages: attachment, penetration, uncoating, gene expression, replication, assembly, and release. Viral gene expression is crucial for viral replication and the following stages. Failure of this step in the viral life cycle will lead to the absence of production of viral proteins and consequently failure of viral propagation. In essence, viral gene expression is the translational decoding of the genetic codes presenting in the viral messenger RNA (mRNA). The translation of viral mRNAs into proteins takes place in the host ribosome and this process resembles the translation of the cellular mRNA into host proteins. In this context, it is important to consider zoonotic viruses, which jump from animals to humans. Translation decoding strategies show host specificity, especially concerning which codons are more efficiently translated and which codons are less efficiently translated. A virus that performs a species jump has to deal with the different translational environments of the new host species, even if the genetic codes are universal for all domains of life. Viruses jumping from one host to another may correspondingly be subject to evolutionary pressure to change their specific codon content. A virus with high translation efficiency will finally become dominant in its new host and cultivate into a new (quasi-)species, in turn provoking disease (e.g., zoonotic Hepatitis E virus infection) [3] or even provoke an infectious pandemic

(e.g., bat-to-human SARS-CoV-2) [4]. Translational decoding is an important driving force in viral evolution and understanding its principles may open novel avenues for the rational treatment of this group of infectious diseases.

Similar considerations apply for oncological disease (cancer), which is a type of host disease that involves uncontrolled cell growth. Almost all tissue-derived cells can transform into cancerous cells via the acquisition of mutations that allow escape from growth control mechanisms. The resulting mutations result in the formation of alternative codons that are potentially well-adapted to the translational context of the involved tumor cells and as a consequence, the translation of oncogenes may be very efficient. Thus, in practice, cancer transformation is not only dependent on the acquisition of oncogenic mutations but also the subsequent gene expression of the resulting cancer-promoting mRNAs. Hence translational decoding of oncogenes is a potential Achilles heel of cancer. It has to be keep in mind that concerning translational decoding in oncology, this disease progression requires a high expression of cancer-promoting genes but also reduces the expression of cancer-suppressive genes. Genes relevant in the context of cancer may include immune suppressors, immune co-stimulatory molecules, cell growth factors, and apoptosis inhibiting factors of cancerous cells. Hence there is a clear need to understand translational decoding in the context of cancerous diseases.

Understanding the role of host translation machinery in controlling viral infection and cancer development is a significant endeavor. Once we know how genes of a virus and oncogene(s) of cancer are translated [5], we might be able to develop corresponding antiviral drugs and novel anticancer therapies by targeting the key factors of the host translation machinery. In particular, the transfer RNA (tRNA) molecules that directly decode the mRNA of the viruses and oncogenes into corresponding proteins might constitute appealing targets in this respect [6]. tRNA is the fundamental unit in the decoding of mRNA codons for protein synthesis of any specific gene. However, the key factors of translational decoding that regulate viral infection and cancer development are not being well-studied, and diseases associated with hepatitis viruses and also liver cancer are important examples in this respect [7,8].

Translational decoding of a gene

Translational decoding is executed by ribosome, which constitutes the translation machinery that produces the cellular proteome by decoding mRNAs derived from either viruses or cancers [9]. There are a host of factors that synchronize the efficiency and fidelity of translation, such as ribosome subunits, eukaryotic initiation /elongation/ termination factors, GTP, RNA-binding proteins, a pool of amino acid-charging tRNAs, as well as the secondary structure, codon contexts, and base modifications of being translated mRNAs inside the ribosome [10,11]. The direct interaction between mRNA codon and tRNA is essential as these two factors mainly determine the decoding efficiency and/ or fidelity [12,13] (**Figure 1**).

mRNA codon

Following the discovery of the double helix DNA in 1953, the genetic codes hidden inside the DNA molecules were (partially) revealed by 1961 [14]. Soon after, the information represented by all 64-triplet genetic codes was established and the direct link between the RNA sequence and the corresponding protein sequence had been identified. Because twenty amino acids are specified by 61 codons (3 out of 64 codons are stop codons without coupling to any amino acid). While some genetic codes are redundant in which more than one codon (also known as synonymous codons) specifies a single amino acid; this feature is named codon degeneration. Codon degeneration frequently occurs at position 3 of a triplet codon. This implies that the first two positions of the triplet codon are the main determinants of the type of amino acid, the so-called "two-out-of-three" theory [15]. However, not all 64 codons are equally used for coding amino acids, some appear more often in the exome as compared to others (a phenomenon named codon usage bias) and for some codons, there has been clear evolutionary pressure to avoid their use (also known as rare codons). Generally speaking, for the efficiency of mRNA translation into protein, the degree of codon usage bias present in an mRNA is a key rate-limiting factor. Therefore, strategies to increase the yield of recombinant protein in *E.coli*, an organism that is frequently employed for bulk production of recombinant proteins, involve deleting rare codons from the being expressed gene and replacing under-represented synonymous codons with the over-represented ones (also known as codon optimization) [16]. Conversely, viral codon deoptimization-based vaccines have been conceived that attenuates viral replication through replacing over-represented synonymous codons or codon pairs with the under-represented ones. Many attenuated vaccine candidates have now been successfully developed by this approach, including poliovirus, avian influenza virus, human respiratory syncytial virus, and many others [17-20]. Collectively, the basic principle of these applications is to regulate the translation efficiency of involving genes of bacterial or virus [16]. As a result, following inoculation viruses will reproduce slowly, giving the immune system the time to develop protective immune responses. If protein sequence-synonymous but "codon-optimized" pathogenic viruses infect the individual later, the immune system is already capable of recognizing the virus and can easily control infection. The success of this approach demonstrates the importance of translational decoding for understanding viral biology.
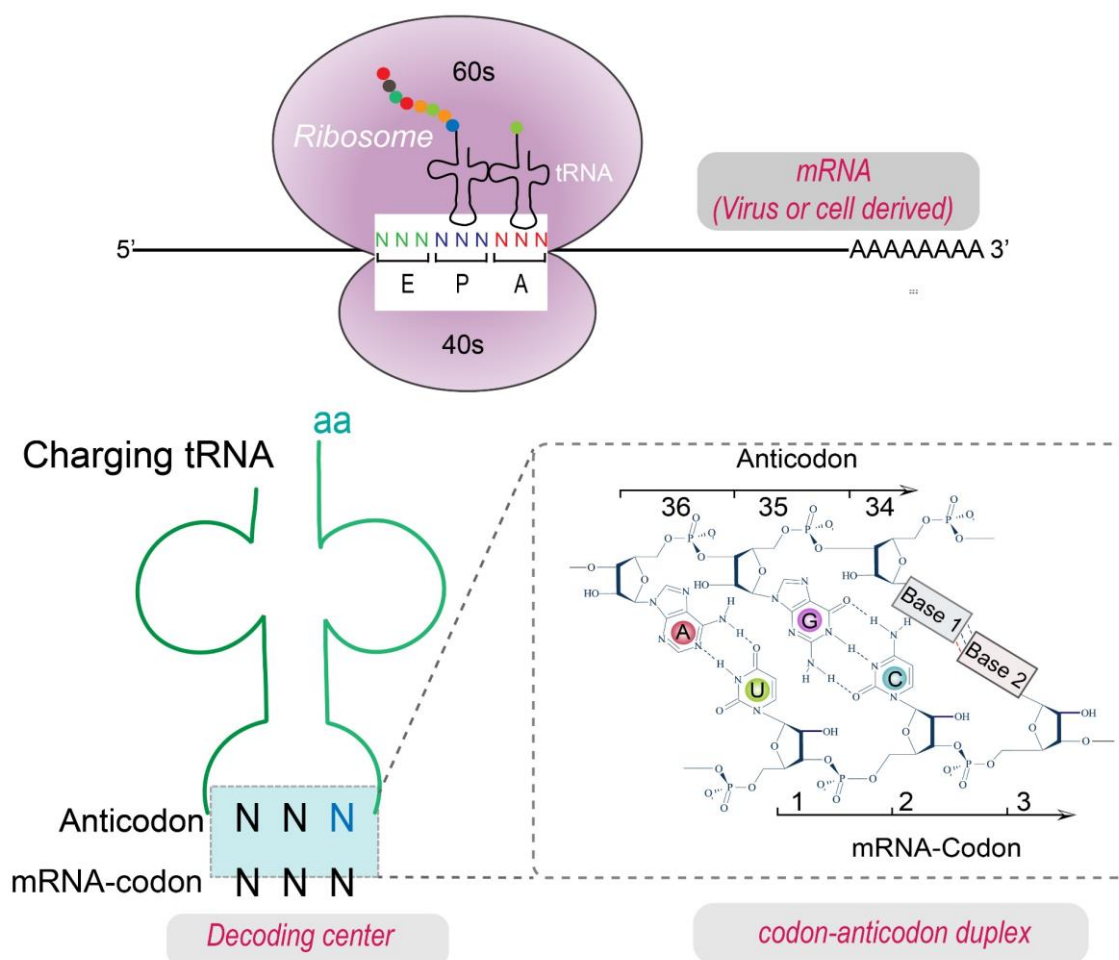
Figure 1. Illustration of translational decoding. Both viruses and cell-derived mRNAs are translated by the cellular ribosome. The triplet codons of the to-be-translated mRNA are decoded by a pool of amino acid-charging tRNAs. tRNAs can recognize all the 64 triplet codons after properly base pairing by its anticodon. tRNA, transfer RNA; mRNA, messenger RNA; aa, amino acid.

Transfer RNA

mRNA codons are decoded by mature tRNAs (equivalent to amino acid charging tRNA) that transfer specific amino acids to synthesize proteins [21]. The decoding process mainly involves Watson-Crick base pairing (i.e., A=T & G=C) at the interface of mRNA codon and tRNA anticodon, as well as non-Watson-Crick base-pairing like wobbling or superwobblling [22]. In the canonical interpretation, 61 aminoacyl-tRNAs and 3 suppressor tRNAs (that terminate translation) decode 64 triplet codons that specify the twenty amino acids. The resulting redundancies in the genetic codes give rise to the synonymous codons, and their decoding involves wobbling at position 3 of the codon. For each amino acid, the number of codons that can give rise to the incorporation of a specific amino acid in the nascent peptide chain varies from two to six. In parallel, the number of different tRNAs that can be charged with the specific amino acid (based on the identity of their anticodons) can also vary from two to six. Translational decoding of the mRNA codons is constrained by factors during codon-anticodon

recognition and often constitutes the rate-limiting step during protein synthesis. Besides the abundance of tRNA species, mRNA translational decoding is also regulated by nearly a hundred epigenetic tRNA modifications, especially at the wobble site of the tRNA anticodon [23]. Accordingly, optimizing tRNA wobble and codon usage in mRNA can substantially enhance translation efficiency and accuracy [24]. The implications of these fundamental mechanisms to the pathology remain often obscure at best, and improving this situation drove me to perform the works described in the present thesis.

Hepatitis viruses

Hepatitis viruses are common in both humans and animals, and important examples are the five human hepatitis viruses, categorized from hepatitis A virus (HAV) to hepatitis E virus (HEV), as well as animal hepatitis viruses like duck Hepatitis A virus (DHAV), duck Hepatitis B virus (DHBV), and probably DHCV, DHDV, DHEV [25]. Different from human hepatitis viruses, the duck hepatitis viruses are not being well-studied as they pose no potential threat to public health. Among these (likely) five types of duck hepatitis viruses, only DHAV infection is highly fatal for birds when they are ducklings of two weeks or less, but mature ducks are not markedly affected by DHAV infection. DHBV infection is associated with duck liver cancer and especially hepatocellular carcinoma (HCC), and some cases of cholangiocellular carcinoma. As a similar etiology with human hepatitis B, the duck DHBV infection model has been proposed and was used for screening of anti-HBV drugs. However, the clinical presentation of hepatitis-related DHBV infection is not obvious, thus limiting its potential for hepatitis research in general. This has triggered researchers to identify new animal hepatotropic viruses (e.g., DHAV) that can induce liver disease, including acute and chronic hepatitis. This type of viral hepatitis animal models may allow monitoring the disease progression in real-time and help elucidate the underlying mechanisms of pathogenesis. Different from animal viral hepatitis, the impact of human viral hepatitis on public health is enormous. As WHO reported, there are 325 million people globally living with a hepatitis virus infection. All five human hepatitis viruses cause liver disease but have different modes of transmission, disease severity, and geographical distribution. In particular, human HBV and HCV lead to chronic hepatitis and are the most common cause of cirrhosis and liver cancer. While human HAV and HEV infections frequently cause acute viral hepatitis or fulminant hepatitis and in most cases is self-limiting [25,26].

Duck hepatitis A virus

DHAV-1, a member of the family *Picornaviridae*, was recently classified as the only member of the genus of *Avihepatovirus* [27]. It is highly pathogenic to ducklings less than one-week old (morbidity and mortality, 100% and 95%, respectively) [28]. This virus has a single-stranded positive RNA genome with the similar genomic organization of Picornavirus, including 5' untranslated regions (UTR) covalently linked by a VPg, a large open reading frame (ORF), 3'UTR polyadenylated by the PolyA tail (**Figure 2**). Once the viral RNA is released into the host cells, it serves as a template for both translation and replication followed by assembly of a large number of progeny DHAV virions. Different from the majority of host mRNA translation, a

single viral RNA will be translated into the entire viral proteome (that consists of ten different proteins) via proteolytic splitting of a resulting polyprotein, but the host mRNA is only translated by one specific protein. Vaccination with a live attenuated vaccine that was developed by serial passaging of the duck isolates in chicken embryos, can prevent infection of susceptible ducklings by DHAV. These protective effects not only attribute to its high immunogenicity, which is dependent on effectively translational decoding of the viral gene but also the reduced virulence after passaging of the virus in the chicken host, which may be caused by mutations, in form of synonymous or non-synonymous mutations. It would be interesting to study if this notion can be objectified by investigating codon changes in this virus changes before and after passaging it in the chicken host. One hypothesis driving the research described in this thesis is that understanding the role of viral codon variations will shed new insights into many aspects of viral translation that may further add a mechanistic explanation of the artificially viral-cross species infection.
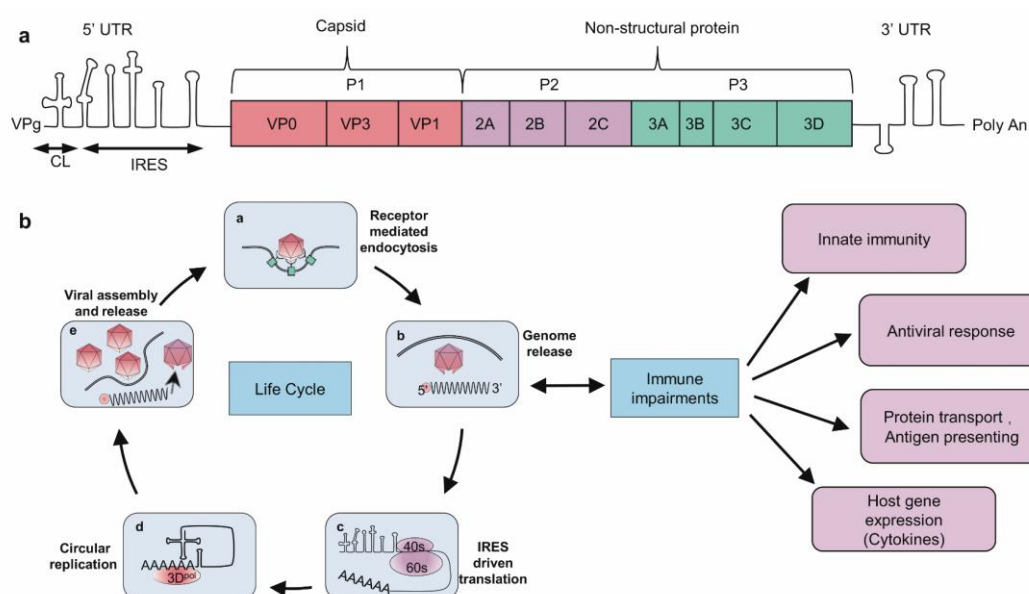
Figure 2. Schematic of the picornavirus genome, life cycle, and immune evasion. (a) The 5'UTR is covalently bound to a short peptide, VPg, while the 3'UTR contains a polyA tail. The unique ORF is embedded between the two untranslated regions. All genes encoding structural and non-structural proteins are listed at the corresponding positions. The types of cis-acting elements are described at the 5'UTR and 3'UTR. (b) The viral life cycle is briefly displayed at the left bottom and categorized according to its basic biological processes, such as endocytosis, genome release, translation, replication, and viral assembly and release. Simultaneously, several strategies are depicted by picornavirus to evade host immunity and antiviral responses.

Human hepatitis E virus

Hepatitis E virus (HEV) infection represents an emerging global health issue, whereas the clinical outcomes of such infection vary dramatically between different populations [26]. The severity of HEV infection is associated with the pregnancy status and the immune status of a patient. Concerning the latter, especially the use of immunosuppressants, as is for instance required following organ transplantation, is an important consideration. About the zoonotic

nature of HEV, the continuous transfer of animal HEV strains to humans remains a significant challenge to public health [29]. Globally, it is estimated that 2.3 billion people have even exposed to this virus, and HEV is considered to be the prime cause of acute viral hepatitis. HEV, a single-strand positive RNA virus, is currently classified in the genus *Orthohepevirus* of the family *Hepeviridae* [30]. It contains three open reading frames (ORF1-3). ORF1 encodes non-structural proteins. ORF2 encodes capsid proteins. ORF3 overlaps ORF2 but not ORF1 (**Figure 3**). According to the types of host reservoirs, HEV is classified into four species, Orthohepevirus (A-D). In Orthohepevirus A, at least eight host species have been identified, *in casu* humans, pigs, wild boars, rabbits, camels, deer, moose, and mongooses [29]. The virus is currently classified into eight distinct genotypes, denominated genotype 1 (gt1) to gt8. Little is known about the zoonotic factors that determine the risk for HEV cross-species infection. Previous studies have suggested that ORF1 was involved in HEV cross-species infection. Intergenotypic chimeric virus using ORF1 of human gt 1 HEV and ORF2 of the zoonotic gt 3 or 4 HEV fails to infect pigs [31]. Substituting ORF1 of gt4 HEV with its counterpart from gt1 HEV induces the capacity of this virus to infect pig cell lines [32]. Among all genotypes in Orthohepevirus A, gt3 HEV has the broadest host spectrum, as it can infect humans, domestic pigs, wild boars, deer, rabbits, and mongooses [28]. The variety of host species of HEV raises an intriguing question that how the virus manages itself to adapt to different translational contexts among multiple hosts [29].
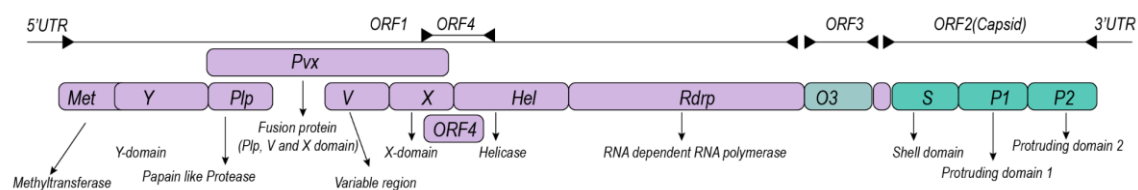
Figure 3. Genome organization of human Hepatitis E virus.

Coronaviruses

The ongoing coronavirus disease 2019 (COVID-19) pandemic is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Seven coronaviruses are known to infect humans, including four low-pathogenic (229E, NL63, OC43, and HKU1) and three high-pathogenic coronaviruses (SARS-CoV-2, SARS-CoV, and MERS-CoV). For SARS-CoV-2, based on the sequence similarity, it was believed that the bats probably carried this virus and transmitted it to humans. As of Agu 31. 2021, over 211.73 million human cases have been confirmed globally. To understand cross-species dynamics better, in this thesis I concentrated on North America for which a lot of data is available. Over 81.82 million cases were from America at the time of the study, particularly hailed from North America [33]. Better understanding the cross-species transmission of SARS-CoV-2 and its human-to-human transmission history may help to guide public health measures [33]. It is well known that viral genetic codon variations are associated with many aspects of virology, most notably viral infectivity, and zoonotic transmission. Because the codon usage of a virus essentially regulates

the translation efficiency of a virus in a given host. It is interesting to see if codon usage of this zoonotic virus before and after jumping to another host is different. Understanding how the viral codons change and mode of mechanism will help public health surveillance by knowing the risk factors of the potentially viral zoonotic jump, that is, translation enhancement by codon recombination.

Liver cancer

Oncological diseases remain a major challenge to the health care system. Normally, human cells grow and multiply by cell division and eventually undergo programmed cell death when the cell is damaged or no longer useful, e.g. because of aging or if the cell has reached the top of the villus in the intestine. Cells precursor to cancer can escape this growth controlling mechanism and follow initial compartment expansion after acquiring mutations, consequently transforming into full-blown cancer. Many types of cancer have been recognized and collectively they provoke untold human misery. According to the Global Cancer Statistics 2020, liver cancer is one of the top three leading causes of cancer-related death, below lung cancer and colorectal cancer. Generally, liver cancer refers to at least two main types of cancer, hepatocellular carcinoma (HCC), an aggressive malignancy derived from hepatocytes, and cholangiocarcinoma (CCA), a hepatobiliary aggressive malignancy derived from biliary tract epithelium with a cholangiocyte aspect [34]. Infection with hepatitis B, C, and D viruses is the leading cause of HCC on a global scale, in particular, the chronic infection following non-resolved HBV infection (the Eastern World) and HCV (the Western World) are precursors to HCC. The mechanistic explanation as to why HBV and HCV predispose to HCC has not been fully resolved but involves the development of persistent and severe liver inflammation. If the immune response is not sufficient to eradicate the virus, chronic infection and persistent inflammation develop, in turn predisposing to cancer development [35]. HCC and CCA are life-threatening diseases and potentially curative treatment is only possible following early detection of these cancers and involves surgical resection, liver transplantation, radiofrequency ablation, chemotherapy, and immunotherapy. However, the prognosis is poor with a 5-year survival rate of 10 % to 18.4%, and a high risk for recurrence, in particular for CCA patients [36]. Understanding the key interplay between the liver cells/inflammatory cells and the hepatitis viruses will pave new clues for novel therapies, and improved knowledge on translational efficiency should prove exceedingly useful handles in this respect.

Implications of tRNA-codon decoding in viral infection and cancer development

Dysregulation of gene expression is well-recognized in both viral infectious diseases and oncological diseases. For certain viral infections, the dysregulation of host translation machinery is mechanically related to an overall inhibition of host cap-dependent mRNA translation via the phosphorylation of eukaryotic initiation factor 2 alpha (eIF2a), or the direct degradation of the host mRNA by viral genome-encoded RNases, as well as partially through

the remodeling of the host tRNA pool to facilitate the decoding efficiency of viral RNA [37]. For cancers, dysregulation of tRNA decoders has been documented for certain types of cancer [38,39]. Understanding the interaction between tRNA and protein-encoding mRNA derived from viruses and/or cancerous oncogenes is important because it determines the likelihood of propagation of the viruses and/ or the proliferation of the cancer cells. Unfortunately, however, knowledge of tRNA decoding biology is limited, and especially related to hepatitis virus infection, HCC, and CCA, and more information would advance the research field.

Viral adaption is shaped by a variety of effectors, including the need to highjack the cellular protein-synthesizing machinery for the efficient translation of viral genes. Rate-limiting steps in the production of viral proteins are critically dependent on the availability of tRNAs that decode both cellular and viral codons. Accordingly, many viruses show adaptation concerning codon usage towards their host [40]. Thus, studying the interplay between viral codons and cellular tRNA is essential for understanding infection biology.

For cancers, the dysregulation of tRNAs has been implicated at least in breast cancer, lung cancer, and melanoma [39,41]. By shaping the tRNA pool to match pro-tumorigenic mRNAs, the translation of oncogenes facilitated prime cancer development, such as highly up-regulated tRNA-Glu-UUC and tRNA-Arg-CCG in breast cancer [39]. Besides regulation at the tRNA transcription level, modification of tRNA anticodon (U34) further supports tumorigenesis by up-regulating U34 enzymes and enhancing translation efficiency of tumor-promoting genes via wobble decoding [8]. For example, a high level of the U34 enzyme promotes alternative translation and has been linked to the resistance of anti-BRAF therapy through wobble decoding of hypoxia-inducible factor 1A (HIF1A) mRNA in a codon-dependent manner [8]. Furthermore, the amino acid metabolism and corresponding tRNA charging process are also important as they support the basic amino acid transfer function of tRNA per se and also require further investigation.

Aim of this thesis

This research is to investigate the role of translation decoding in viral infection and cancer development. The efficiency and productivity of translational decoding of the virus is a precondition for efficient viral infection and accordingly associates with the severity of the resulting disease. Also for cancer, protein synthesis of oncogenes and other cancer-relevant genes determines disease progression. Thus, it is important to gain a deep understanding of the translational decoding, in particular the tRNA- mRNA codon interactions, which are the main aims of this thesis.

Outline of this thesis

All can be studied from the known to the unknown, and from the surface to the deep. From the sequenced individual genomes of humans and two important avian species (in casu the chicken and the duck), in **Chapter 2**, I performed a comparative analysis of the tRNA-encoding genome, identifying disparate and conserved elements that shed light on our thinking as to how evolutionary forces have shaped the tRNA landscape [42]. To confirm the insights, experimentation is necessary but I was confronted by the lack of a robust and technology that allows easy determination of cellular tRNA levels. This prompted me to pursue a research direction, in **Chapter 3**, that aims to develop a simple tRNA detection approach by quantitative PCR [43]. Using this novel methodology, I was able to observe that the hepatic tRNAome is remodeled by HEV infection. Further exploring this research line, I was also interested to learn about the mechanisms employed by cells to avoid erroneous translation. I discovered, however, that mRNA translation is not always perfect and errors in the amino acid composition may occur. In **Chapter 4**, I discuss the key origins of mistranslation by probing the tRNA decoding process [9]. I found that such errors mainly derive from tRNA misdecoding and misacylation, especially when certain codon-pairing tRNA species are missing (see also **Chapter 2**). I was also able to develop my thoughts and discussion on this subject by identifying potential implications of erroneous decoding on viral evolution and cancer development, which show that mRNA translation is often the nexus of viral replication and cancer progression, sometimes even more than that provided by DNA and/ or RNA replication.

Encouraged by the above, I decided to exploit the simple tRNA methodology I developed in **Chapter 3**, in **Chapter 5** to map the tRNAome landscape in liver cancer patients and I explored potential therapeutic targets at the interface of charging amino acids with certain tRNA species [38]. The result is a new class of drugs potentially useful in this devastating disease.

Viral hepatitis is a common disease in both humans and animals. Currently, there is a shortage of small animal models that allow experimental and translational studies to further human hepatology. In **Chapter 6**, I investigated the potential of the mature duck DHAV infection model for human acute and chronic hepatitis research [44]. After inoculation with DHAV, the liver histopathology, serum biomarkers, fibrosis, viremia, and T cell responses were continually assessed during a period that ended 280-day post-infection. The clinical manifestations following infection of mature ducks with this virus were not obvious but were potentially consistent with chronic infection, while the virus-infected ducklings are suffering from the fatal disease with acute hepatitis as the primary trigger of this lethality. This deadly course of disease in ducklings does not occur following vaccination with living DHAV attenuated vaccine that was developed by serial passaging of the duck isolates in chicken embryos. To understand the role of codon variations in this artificially viral cross-species adaption (that is an approach generally used for the development of a live attenuated vaccine), In **Chapter 7**, I characterized the codon variations of DHAV duck isolates after passaging them in the chicken embryo and performed phylogenetic analysis [45]. The research further supports the idea that codon

adaptation is an important element for attenuation of viral strains for vaccination purposes.

During the global COVID pandemic, in **Chapter 8**, my research focus moved to SARS-CoV-2 and two historical coronaviruses, in casu SARS-CoV, and MERS-CoV [46]. These three human coronaviruses are thought to have originated from bats. By investigating these naturally cross-species infections, in this chapter, I have identified several genetic signatures including codon variations of these three pairs of bat-to-human coronavirus. To understand the early SARS-CoV-2 transmission history in the North American human population, I further conducted a geographically stratified genome-wide association study (North American isolates and the remaining isolates) and a retrospective study.

Thus overall, my thesis significantly enhances our understanding of translational decoding in virology and cancer biology. This knowledge may in turn allow better development of novel therapies against viruses and cancer by targeting specific molecules of the translation machinery.

References

1. Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Research* 2017; **46**(D1): D708-D17.

2. Crick FJN. The central dogma of molecular biology. 1970; **227**(5258): 561-3.

3. Zhou JH, Li XR, Lan X, et al. The genetic divergences of codon usage shed new lights on transmission of hepatitis E virus from swine to human. *Infection Genetics and Evolution* 2019; **68**: 23-9.

4. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nature medicine* 2020; **26**(4): 450-2.

5. Sulima SO, Hofman IJF, De Keersmaecker K, Dinman JD. How Ribosomes Translate Cancer. *Cancer Discov* 2017; **7**(10): 1069-87.

6. Schuller AP, Green R. Roadblocks and resolutions in eukaryotic translation. *Nature reviews Molecular cell biology* 2018; **19**(8): 526-41.

7. Truitt ML, Ruggero D. New frontiers in translational control of the cancer genome. *Nat Rev Cancer* 2016; **16**(5): 288-304.

8. Rapino F, Delaunay S, Rambow F, et al. Codon-specific translation reprogramming promotes resistance to targeted therapy. *Nature* 2018; **558**(7711): 605-9.

9. Ou X, Cao J, Peppelenbosch MP, Pan Q. Errors in translational decoding: tRNA wobbling or misincorporation? *PLoS genetics* 2019; **15**(3): e1008017.

10. Demeshkina N, Jenner L, Westhof E, Yusupov M, Yusupova G. A new understanding of the decoding principle on the ribosome. *Nature* 2012; **484**(7393): 256-9.

11. Powell JR, Dion K. Effects of Codon Usage on Gene Expression: Empirical Studies on Drosophila. *Journal of molecular evolution* 2015; **80**(3-4): 219-26.

12. Dana A, Tuller T. The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res* 2014; **42**(14): 9171-81.

13. Novoa EM, de Pouplana LR. Speeding with control: codon usage, tRNAs, and ribosomes. *Trends in Genetics* 2012; **28**(11): 574-81.

14. Crick FH. Codon--anticodon pairing: the wobble hypothesis. *Journal of molecular biology* 1966; **19**(2): 548-55.

15. Parker J, Friesen JD. "Two out of three" codon reading leading to mistranslation in vivo. *Molecular and General Genetics MGG* 1980; **177**(3): 439-45.

16. Quax TE, Claassens NJ, Soll D, van der Oost J. Codon Bias as a Means to Fine-Tune Gene Expression. *Molecular cell* 2015; **59**(2): 149-61.

17. Fan RL, Valkenburg SA, Wong CK, et al. Generation of Live Attenuated Influenza Virus by Using Codon Usage Bias. *J Virol* 2015; **89**(21): 10762-73.

18. Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E, Mueller S. Virus attenuation by genome-scale changes in codon pair bias. *Science* 2008; **320**(5884): 1784-7.

19. Manokaran G, Sujatmoko, McPherson KG, Simmons CP. Attenuation of a dengue virus replicon by codon deoptimization of nonstructural genes. *Vaccine* 2019; **37**(21): 2857-63.

20. Le Nouen C, Luongo CL, Yang L, et al. Optimization of the Codon Pair Usage of Human Respiratory Syncytial Virus Paradoxically Resulted in Reduced Viral Replication In Vivo and Reduced Immunogenicity. *J Virol* 2020; **94**(2): e01296-19.

21. Chan PP, Lowe TM. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in

complete and draft genomes. *Nucleic Acids Res* 2016; **44**(D1): D184-9.

22. Alkatib S, Scharff LB, Rogalski M, et al. The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS genetics* 2012; **8**(11): e1003076.

23. Ranjan N, Rodnina MV. tRNA wobble modifications and protein homeostasis. *Translation (Austin)* 2016; **4**(1): e1143076.

24. Endres L, Dedon PC, Begley TJ. Codon-biased translation can be regulated by wobble-base tRNA modification systems during cellular stress responses. *RNA biology* 2015; **12**(6): 603-14.

25. Lanford RE, Feng Z, Chavez D, et al. Acute hepatitis A virus infection is associated with a limited type I interferon response and persistence of intrahepatic viral RNA. *Proceedings of the National Academy of Sciences of the United States of America* 2011; **108**(27): 11223-8.

26. Nimgaonkar I, Ding Q, Schwartz RE, Ploss A. Hepatitis E virus: advances and challenges. *Nature Reviews Gastroenterology & Hepatology* 2017; **15**: 96.

27. Ding C, Zhang D. Molecular analysis of duck hepatitis virus type 1. *Virology* 2007; **361**(1): 9-17.

28. Song C, Wan H, Yu S, et al. Rapid detection of duck hepatitis virus type-1 by reverse transcription loop-mediated isothermal amplification. Journal of virological methods; 2012. p. 76-81.

29. Meng XJ. Expanding Host Range and Cross-Species Infection of Hepatitis E Virus. *PLoS pathogens* 2016; **12**(8): e1005695.

30. Shukla P, Nguyen HT, Faulk K, et al. Adaptation of a genotype 3 hepatitis E virus to efficient growth in cell culture depends on an inserted human gene segment acquired by recombination. *J Virol* 2012; **86**(10): 5697-707.

31. Feagins AR, Córdoba L, Sanford BJ, et al. Intergenotypic chimeric hepatitis E viruses (HEVs) with the genotype 4 human HEV capsid gene in the backbone of genotype 3 swine HEV are infectious in pigs. *Virus research* 2011; **156**(1-2): 141-6.

32. Chatterjee SN, Devhare PB, Pingle SY, Paingankar MS, Arankalle VA, Lole KS. Hepatitis E virus (HEV)-1 harbouring HEV-4 non-structural protein (ORF1) replicates in transfected porcine kidney cells. *Journal of General Virology* 2016; **97**(8): 1829-40.

33. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020; **579**(7798): 270-3.

34. Sia D, Villanueva A, Friedman SL, Llovet JM. Liver cancer cell of origin, molecular class, and effects on patient prognosis. *Gastroenterology* 2017; **152**(4): 745-61.

35. Gehring AJ, Protzer U. Targeting innate and adaptive immune responses to cure chronic HBV infection. *Gastroenterology* 2019; **156**(2): 325-37.

36. Razumilava N, Gores GJ. Cholangiocarcinoma. *The Lancet* 2014; **383**(9935): 2168-79.

37. Nunes A, Ribeiro DR, Marques M, Santos MAS, Ribeiro D, Soares AR. Emerging Roles of tRNAs in RNA Virus Infections. *Trends in biochemical sciences* 2020; **45**(9): 794-805.

38. Zhang R, Noordam L, Ou X, et al. The biological process of lysine-tRNA charging is therapeutically targetable in liver cancer. *Liver Int* 2020.

39. Goodarzi H, Nguyen HCB, Zhang S, Dill BD, Molina H, Tavazoie SF. Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. *Cell* 2016; **165**(6): 1416-27.

40. Zhao F, Yu CH, Liu Y. Codon usage regulates protein structure and function by affecting translation elongation speed in Drosophila cells. *Nucleic Acids Res* 2017; **45**(14): 8484-92.

41. Rapino F, Close P. Wobble uridine tRNA modification: a new vulnerability of refractory melanoma. *Mol Cell Oncol* 2018; **5**(6): e1513725.

42. Ou X, Peng W, Yang Z, et al. Evolutionarily missing and conserved tRNA genes in human and avian. *Infect Genet Evol* 2020: 104460.

43. Ou X, Ma B, Zhang R, et al. A simplified qPCR method revealing tRNAome remodeling upon infection by genotype 3 hepatitis E virus. *FEBS Lett* 2020; **594**(12): 2005-15.

44. Ou X, Mao S, Cao J, et al. The neglected avian hepatotropic virus induces acute and chronic hepatitis in ducks: an alternative model for hepatology. *Oncotarget* 2017; **8**(47): 81838-51.

45. Ou X, Wang M, Mao S, et al. Incompatible Translation Drives a Convergent Evolution and Viral Attenuation During the Development of Live Attenuated Vaccine. *Frontiers in Cellular and Infection Microbiology* 2018; **8**.

46. Ou X, Yang Z, Zhu D, et al. Tracing genetic signatures of bat-to-human coronaviruses and early transmission of North American SARS-CoV-2. *Transboundary and Emerging Diseases*, 2021,00:1-13

PART I

Role of tRNA biology in viral infection and cancer development

Chapter 2

Evolutionarily missing and conserved tRNA genes

in human and avain

Xumin Ou, Wenjing Peng, Zhishuang Yang, Jingyu Cao, Mingshu Wang, Maikel P Peppelenbosch, Qiuwei Pan, Anchun Cheng

Abstract

Viral infection heavily relies on host transfer RNA (tRNA) for viral RNA decoding. Counterintuitively, not all tRNA species based on anticodon are matched to all 64-triplet codons during evolution. Life solves this problem by cognate tRNA species via wobbling decoding. We found that 14 out of 64 tRNA genes in humans and the main avian species (chicken and duck) were parallelly missing, including 8 tRNA-A_{34}NN and 6 tRNA-G_{34}NN species. By analyzing the conservation of key motifs in tRNA genes, we found that box A and B served as intragenic tRNA promoters were evolutionally conserved among human, chicken, and duck. Thus, decoding viral RNA by similar wobbling strategies and tRNA transcripts may be parallelly used by human, chicken, and duck. We envisioned that many basic mechanisms regarding viral RNA decoding were possibly conserved in these hosts and may consequently promote cross-species infection.

Keywords: tRNA, Viral decoding, Infection and evolution, Human, Avian

Highlights

- The 8 tRNA - $A_{34}NN$ and 6 tRNA - $G_{34}NN$ are concurrently missing in humans (Homo sapiens), chicken (Gallus gallus), and duck (Anas platyrhynchos) during evolution.
- Box A and B, the intragenic tRNA promoters, are evolutionally conserved between human, chicken, and duck.
- Box A and B in human, chicken, and duck are TGGNNNAG(A)TGG and GGGTTCGANNCC, respectively.

Introduction

Transfer RNAs (tRNAs) are essentially required for gene decoding. Despite the universal nature of genetic codon, not all tRNA genes are common to all organisms. Here, we would like to discuss fundamental problems and possible effects arising from the evolutionarily missing and conserved tRNA genes in human, chicken, and duck (Alkatib et al., 2012; Ou et al., 2019; Rogalski et al., 2008). Among these three organisms, viruses especially the avian influenza virus can crossly infect (Pepin et al., 2010). For multi-host viruses, similar viral RNA decoding strategies may be parallelly used by different hosts. Because viral cross-species infection heavily relies on host tRNAs of different species for viral RNA decoding (Ou et al., 2020; van Weringh et al., 2011). We envisioned that many basic mechanisms regarding viral RNA decoding were possibly conserved in these three hosts and may consequently promote cross-species infection.

Evolutionarily missing tRNA genes in human and avian

In the ribosome, 61 amino acid charging tRNAs and 3 suppress tRNAs are theoretically needed to decode 64 genetic codons that specify 20 amino acids (Ou et al., 2019). However, we found that certain types of tRNA in humans and the main species of avian are missing during evolution. This triggers an interesting question that how the 64 genetic codons of cellular mRNA or viral RNA can be decoded by partially missing tRNAs. Recently, it has been verified that the number of obligatory tRNA species (based on anticodons) for decoding is substantially less than 64. Life solves this problem by wobbling that the unpaired codons are decoded by cognate tRNA species otherwise misdecoding (Ou et al., 2019; Rogalski et al., 2008). It has been proven in humans that 15 out of theoretically necessary 64 tRNA species, including 8 tRNA-A_{34}NN and 7 tRNA-G_{34}NN, are missing confirmed by tRNAscan-SE (Lowe and Eddy, 1997) and tRNAome sequencing (Shigematsu et al., 2017; Zheng et al., 2015). Intriguingly, we also found that these missing tRNA species in main poultry, such as chicken (*Gallus gallus*) and duck (*Anas platyrhynchos*), matched 93.3% (14/15) to the missing types in human (*Homo sapiens*), including 8 tRNA - A_{34}NN and 6 tRNA - G_{34}NN species (**Figure 1A**). The wobbling decoding has also been extensively investigated in other organisms like plant and E coli, in which 25 tRNA species suffice protein biosynthesis (Alkatib et al., 2012). This evidence indicates that an evolutionarily conserved wobbling decoding mechanism is likely universal in human, chicken, and duck as the identically missing tRNA species (Ou et al., 2019).

Conservation of tRNA promoter among human and avian

Though tRNA genes are partially missing (based on anticodon), they are redundant by other criteria as multiple iso-decoders and iso-receptors exist, for instance, the 9 iso-decoders of tRNA-His-GUG in the human genome (Schimmel, 2018) (**Table S1**). As verified, the human tRNAs are encoded by more than 600 tRNA gene locus, and that is nearby 1-fold more than that in chicken (No.: 272) and duck (No.: 361) (Chan and Lowe, 2016) (**Table S1**). As for

transcription of the tRNA gene locus, box A and B serve as intragenic tRNA promoter that is distinct from the extragenic promoter of protein-encoding genes. These two boxes are constructively involved in the D loop and TΨC loop of tRNA. By comparing the conservation of key motifs in tRNA, we found that box A and B are evolutionarily conserved among human and avian (Vassetzky and Kramerov, 2013) (**Figure 1B**). Specifically, box A and B of tRNA genes in human and avian are TGG_NNN_AG(A)TGG and GGGTTCGA_NN_CC, respectively (Vassetzky and Kramerov, 2013). Strikingly, these two boxes are identical in chicken and duck, in which box A is TGG*GCT*AATGG and GGGTTCGA*TC*CC for box B (**Figure 1B**).

A

Two box tRNA sets

Isotype				
Phe	AAA — None None None	GAA — 10 9 12		
Asn	ATT — None None None	GTT — 20 11 13		
Lys			CTT — 15 6 17	TTT — 12 5 4
Asp	ATC — None None None	GTC — 13 9 8		
Glu			CTC — 8 8 8	TTC — 7 7 7
His	ATG — None None None	GTG — 10 7 10		
Gln			CTG — 13 6 9	TTG — 6 2 5

Two box tRNA and other tRNA sets

Isotype				
Ile	AAT — 14 7 8	GAT — 3 None None		TAT — 5 3 2
Met/Me			CAT — 9/10 10/4 18	
Tyr	ATA — None None None	GTA — 13 15 17		
Suppres		TCA	CTA	TTA
Cys	ACA — None None None	GCA — 29 12 20		
Trp			CCA — 7 6 6	
SelCys				TCA — 2 1 None

Four box tRNA sets

Isotype				
Ala	AGC — 22 16 19	GGC — None None None	CGC — 4 3 9	TGC — 8 7 11
Gly	ACC — None None None	GCC — 14 7 13	CCC — 5 5 5	TCC — 9 8 12
Pro	AGG — 9 8 10	GGG — None None None	CGG — 4 3 9	TGG — 7 5 9
Thr	AGT — 9 4 7	GGT — None None None	CGT — 5 2 2	TGT — 6 6 3
Val	AAC — 9 8 7	GAC — None 1 1	CAC — 11 7 7	TAC — 5 3 2

NNA	NNG	NNC	NNU	NNC/A	NNU/G	←Anticodon
H/C/D	H/C/D	H/C/D	H/C/D	H/C/D	H/C/D	← tRNA No.

Two-box Two-box   **H/C/D**=Human/Chicken/Duck
Four-box
Six-box

Six box tRNA sets

Isotype							
Ser	AGA — 9 8 11	GGA — None None None	CGA — 4 3 3	TGA — 4 4 4	ATC — None None None	GCT — 8 7 9	
Arg	ACG — 7 6 11	GCG — None None None	CCG — 4 2 1	TCG — 6 3 3			CCT — 5 3 4 / TCT — 6 4 3
Leu	AAG — 9 5 8	GAG — None None None	CAG — 9 8 9	TAG — 3 2 1			CAA — 6 4 3 / TAA — 4 2 1

B

Duck tRNA genes

Box A — Anticodon — Box B

Chicken tRNA genes
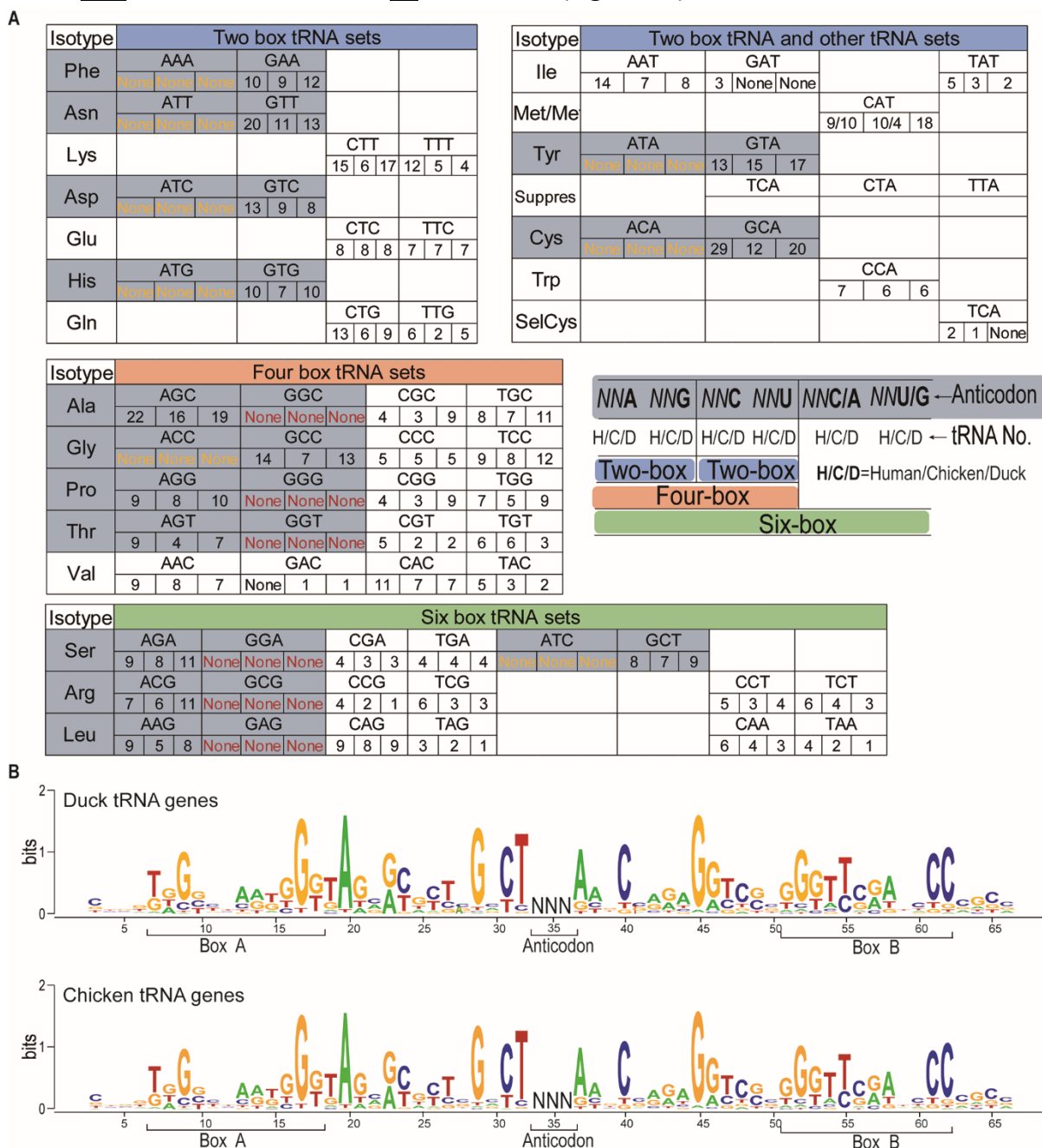
Box A — Anticodon — Box B

Figure 1 Evolutionarily missing and conserved tRNA genes in human and avian. (A) Sixty-four tRNA sets are summarized according to the order of human and avian (chicken and duck) and specified in parallel with numbers of tRNA gene locus. The identically missing types in human

and avian, including 8 tRNAA$_{34}$NN (yellow text) and 6 tRNAG$_{34}$NN (red text), are indicated. (**B**) Comparative analysis indicates that the boxes A and B of tRNA genes in human and avian are TGGNNNAG(A)TGG and GGGTTCGANNCC, respectively. These two boxes are identical in chicken and duck, in which box A is TGGGCTAATGG and GGGTTCGATCCC for box B. Conservation of tRNA genes (chicken & duck: 272: 361) were generated by multiple alignments and visualized by WebLogo 3.1

The implication of evolutionarily missing and conserved tRNA genes for viral decoding

Viral decoding is regulated by the landscape of tRNA – viral codons, probably through adapting to host codon usage or modulating the host tRNAome (Ou et al., 2020). It has been reported that codon usage bias is generally distinct between virus and host due to translational selection (Chen et al., 2020). For viral decoding in human, chicken, and duck, because the same tRNA species (based on the anticodon) are parallelly missing, the unpaired codons of viruses are obligately decoded by the cognate tRNA species through similar wobbling decoding strategies (Ou et al., 2019). Due to the difference in codon usage bias between virus and host, we speculate that efficient viral translation may require rewiring the host tRNAome. Supportively, we recently found that the expression landscape of hepatic tRNAome was remodeled by infection of the zoonotic hepatitis E virus (HEV) (Ou et al., 2020). Because of the conservation of the D loop and TΨC loop (Box A and B) of tRNA species, regulation of tRNAome at the transcriptional level is well possible though not all tRNA species are equally transcribed. Using similar wobbling decoding strategies or host tRNAome remodeling, viruses may adapt two or more hosts like human, chicken, and duck thus may contribute to viral cross-species infection. For the avian influenza virus, the evolutionally missing and conserved tRNA genes between human, chicken, and duck can possibly explain why this virus can crossly infect among these hosts. Because the similar tRNA transcripts and wobbling decoding strategies are used for RNA decoding of the avian influenza virus. Besides, the viral cross-species infection also links to the development of live attenuated vaccines, such as the Sabin vaccine, dengue vaccine, influenza vaccine, and mumps vaccine (Bhamarapravati and Sutee, 2000; Koprowski, 1960). Because they were developed through a series of passages from one host to another. As tested in chicken and duck, we developed a live attenuated vaccine by serial passage of the virulent strain of duck hepatitis A virus in the chicken embryo (Ou et al., 2017; Ou et al., 2018).

Collectively, we found those tRNA genes in humans and the main avian species were parallelly missing and evolutionarily conserved. We envisioned that many basic mechanisms regarding viral RNA decoding were possibly conserved in these hosts. For viral cross-species infection in humans and avian, the investigation from the standpoint of tRNA-codon interaction will define novel mechanisms in this respect and decipher viral translation in many aspects (Schimmel, 2018).

References

Alkatib, S., Scharff, L.B., Rogalski, M., Fleischmann, T.T., Matthes, A., Seeger, S., Schottler, M.A., Ruf, S., Bock, R., 2012. The contributions of wobbling and superwobbling to the reading of the genetic code. PLoS Genet 8, e1003076.

Bhamarapravati, N., Sutee, Y., 2000. Live attenuated tetravalent dengue vaccine. Vaccine 18 Suppl 2, 44-47.

Chan, P.P., Lowe, T.M., 2016. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res 44, D184-189.

Chen, F., Wu, P., Deng, S., Zhang, H., Hou, Y., Hu, Z., Zhang, J., Chen, X., Yang, J.-R., 2020. Dissimilation of synonymous codon usage bias in virus–host coevolution due to translational selection. Nature Ecology & Evolution 4, 589-600.

Koprowski, H., 1960. Historical aspects of the development of live virus vaccine in poliomyelitis. Br Med J 2, 85-91.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic acids research 25, 955-964.

Ou, X., Cao, J., Peppelenbosch, M.P., Pan, Q., 2019. Errors in translational decoding: tRNA wobbling or misincorporation? Plos Genetics 15, e1008017.

Ou, X., Ma, B., Zhang, R., Miao, Z., Cheng, A., Peppelenbosch, M.P., Pan, Q., 2020. A simplified qPCR method revealing tRNAome remodeling upon infection by genotype 3 hepatitis E virus. FEBS Lett 594, 2005-2015.

Ou, X., Mao, S., Cao, J., Cheng, A., Wang, M., Zhu, D., Chen, S., Jia, R., Liu, M., Sun, K., Yang, Q., Wu, Y., Chen, X., 2017. Comparative analysis of virus-host interactions caused by a virulent and an attenuated duck hepatitis A virus genotype 1. PLOS ONE 12, e0178993.

Ou, X., Wang, M., Mao, S., Cao, J., Cheng, A., Zhu, D., Chen, S., Jia, R., Liu, M., Yang, Q., Wu, Y., Zhao, X., Zhang, S., Liu, Y., Yu, Y., Zhang, L., Chen, X., Peppelenbosch, M.P., Pan, Q., 2018. Incompatible Translation Drives a Convergent Evolution and Viral Attenuation During the Development of Live Attenuated Vaccine. Frontiers in Cellular and Infection Microbiology 8, 249.

Pepin, K.M., Lass, S., Pulliam, J.R.C., Read, A.F., Lloyd-Smith, J.O., 2010. Identifying genetic markers of adaptation for surveillance of viral host jumps. Nature Reviews Microbiology 8, 802-813.

Rogalski, M., Karcher, D., Bock, R., 2008. Superwobbling facilitates translation with reduced tRNA sets. Nat Struct Mol Biol 15, 192-198.

Schimmel, P., 2018. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. Nat Rev Mol Cell Biol 19, 45-58.

Shigematsu, M., Honda, S., Loher, P., Telonis, A.G., Rigoutsos, I., Kirino, Y., 2017. YAMAT-seq: an efficient method for high-throughput sequencing of mature transfer RNAs. Nucleic Acids Res 45, e70.

van Weringh, A., Ragonnet-Cronin, M., Pranckeviciene, E., Pavon-Eternod, M., Kleiman, L., Xia, X., 2011. HIV-1 modulates the tRNA pool to improve translation efficiency. Mol Biol Evol 28, 1827-1834.

Vassetzky, N.S., Kramerov, D.A., 2013. SINEBase: a database and tool for SINE analysis. Nucleic Acids Res 41, D83-89.

Zheng, G., Qin, Y., Clark, W.C., Dai, Q., Yi, C., He, C., Lambowitz, A.M., Pan, T., 2015. Efficient and quantitative high-throughput tRNA sequencing. Nat Methods 12, 835-837.

Table S1 tRNA copies and location among human, chicken, and duck.

#	gene	Human No.	Human Chromosome	Chicken No.[①]	Chicken Chromosome[①]	Chicken No.[②]	Chicken Chromosome[②]	Duck No.	Duck Chromosome
1	tRNA-Phe-AAA	None		None		None		None	
2	tRNA-Phe-GAA	10	6,11,12,13,19	9	1,2,15,28	9		12	1,2,3,16,20
3	tRNA-Asn-ATT	None		None		None		None	
4	tRNA-Asn-GTT	20	1,10,13,17,19	11	1,2,9,16,27,28	11		13	1,2,4,7,9,17,20,28,Z
5	tRNA-Lys-CTT	15	1,5,6,14,15,16,19	6	5,10,16	6		17	1,2,3,4,5,9,11,17
6	tRNA-Lys-TTT	12	1,6,11,12,16,17	4	11,26	5	11,26,Un	4	11,27
7	tRNA-Asp-ATC	None		None		None		None	
8	tRNA-Asp-GTC	13	1,6,12,17	9	1,13,15	9		8	1,16
9	tRNA-Glu-CTC	8	1,6	8	2,3,16,20	8		8	2,3,17,19,29
10	tRNA-Glu-TTC	7	1,2,13,15	6	2,6,16,20	7	2,4[③],6,16,20	7	3,17,21,27
11	tRNA-His-ATG	None		None		None		None	
12	tRNA-His-GTG	10	1,6,9,15	7	1,5,6,10,11,Z	7		10	1,2,11,Z
13	tRNA-Gln-CTG	13	1,6,15,17	5	10,18	6	10,18,Un	9	11,19
14	tRNA-Gln-TTG	6	6,17	2	18,27	2		5	3,5,15,19,28
15	tRNA-Ile-AAT	14	6,14,17	7	1,5,10	7		8	1,5,9,11
16	tRNA-Ile-GAT	3	X	None		None		None	
17	tRNA-Ile-TAT	5	2,6,19	3	3,27,Z	3		2	3,28
18	tRNA-Met-CAT	19	6,8,16	14	1,2,3,4,7,10,11,18,25	14		18	2,3,7,11,19
19	tRNA-Tyr-ATA	None		None		None		None	
20	tRNA-Tyr-GTA	13	2,6,8,14	15	2	15		17	1,2,3
21	tRNA-Cys-ACA	None		None		None		None	
22	tRNA-Cys-GCA	29	1,3,4,7,14,15,17	12	27	12		20	28
23	tRNA-Trp-CCA	7	6,7,12,17	6	1,3,26	6		6	1,3,27
24	tRNA-SelCys-TCA	2		1		1		None	
25	tRNA-Ala-AGC	22	2,6,8,14	10	2,3,5	16	2,3,5	19	1,2,3,5
26	tRNA-Ala-GGC	None		None		None		None	
27	tRNA-Ala-CGC	4	2,6	3	1,2,7	3		9	1,5,7,16,Z
28	tRNA-Ala-TGC	8	5,6,12	6	7,15	7	5,7,15	11	3,7,16
29	tRNA-Gly--ACC	None		None		None		None	
30	tRNA-Gly-GCC	14	1,2,6,16,17,21	7	3,7,11,19	7		13	1,3,4,5,11,20,Z
31	tRNA-Gly-CCC	5	1,2,16,17	5	1,5,8,14,22	5		5	1,2,5,8,15
32	tRNA-Gly-TCC	9	1,17,19	8	3,5,10,18,28	8		12	1,2,5,11,20,21
33	tRNA-Pro-AGG	9	1,6,7,11,14,16	8	1,2,7	8		10	1,7,25
34	tRNA-Pro-GGG	None		None		None		None	
35	tRNA-Pro-CGG	4	1,6,16,17	3	1,2,7	3		9	1,7
36	tRNA-Pro-TGG	7	5,11,14,16	5	1,7,16	5		9	1,7[④],17
37	tRNA-Thr-AGT	9	6,17,19	4	1,7,11,17	4		7	1,7,12,17,18
38	tRNA-Thr-GGT	None		None		None		None	
39	tRNA-Thr-CGT	5	6,16,17	2	14,18	2		2	15,19
40	tRNA-Thr-TGT	6	1,5,6,14	6	3,5,16	6		3	3,5,17
41	tRNA-Val-AAC	9	3,5,6	8	1,2,9,16	8		7	1,9,16,17
42	tRNA-Val-GAC	None		1	7	1[⑤]		1	7
43	tRNA-Val-CAC	11	1,5,6,19	7	1,2,16,28	7		7	1,17,20
44	tRNA-Val-TAC	5	6,10,11,X	3	1,2	3		2	1,2
45	tRNA-Ser-AGA	9	6,8,17	8	1,2,3	8		11	1,2,3,11,Z
46	tRNA-Ser-GGA	None		None		None		None	
47	tRNA-Ser-CGA	4	6,12,17	3	7,33	3		3	7[⑥]
48	tRNA-Ser-TGA	4	6,10	4	6,7,19	4		4	1,7,8,29
49	tRNA-Ser-ATC	None		None		None		None	
50	tRNA-Ser-GCT	8	6,11,15,17	7	2,5,7,27	7		9	1,2,3,5,7,28

51	tRNA-Arg-ACG	7	3,6,14	6	2,3	6	11	2,3
52	tRNA-Arg-GCG	None		None		None	None	
53	tRNA-Arg-CCG	4	6,16,17	2	18	2	1	19
54	tRNA-Arg-TCG	6	6,9,15,17	3	10,18	3	3	11,19
55	tRNA-Arg-CCT	5	7,16,17	3	1,18	3	4	1,19
56	tRNA-Arg-TCT	6	1,6,9,11,17	4	8,17,18,25	4	3	8,18,19⑦
57	tRNA-Leu-AAG	9	5,6,14,16	5	14,16	5	8	1,2,17
58	tRNA-Leu-GAG	None		None		None	None	
59	tRNA-Leu-CAG	9	1,6,16	8	1,2,3,4,7,11	8	9	1,3,4,5,7,11,Z
60	tRNA-Leu-TAG	3	14,16,17	2	1,14	2	1	15
61	tRNA-Leu-CAA	6	1,6	4	4,16	4	3	10,17
62	tRNA-Leu-TAA	4	6,11	2	3,12	2	1	13
	Total	417		272①		282②	361	

Note：

①, NCBI genome database.

②, GtRNA database.

③, tRNA-Glu-TTC：The original "tRNA-Glu-TCC" in NCBI database was re-anotated as "tRNA-Glu-TTC" because the anticodon was "TTC"
(TCCCATATGGTCTAGCGGTTAGGATTCCTGGTTTTCACCCAGGCGGCCCGGGTTCGACTCCCGGTATGGGAA).

④, tRNA-Pro-AGG：The original "tRNA-Pro-TGG" in NCBI database was re-anotated as "tRNA-Pro-AGG" because the anticodon was "AGG"
(GGCTCGTTGGTCTAGGGGTATGATTCTCGCTTAGGGTGCGAGAGGTCCCGGGTTCAACTCCCGGACGAGCCC).

⑤, Low confidence.

⑥, tRNA-Ser-TGA：The original "tRNA-Ser-CGA" in NCBI database was re-anotated as "tRNA-Ser-TGA" because the anticodon was"TGA"
(GCTGTGATGGCCGAGGGGTGAAGGCGTTGGACTTGAAATCCAATAGGGTTTCCCTGCGCAGGTTCGAATCCTGCTC ACAGCG).

⑦, tRNA-Arg-TCT：The original "tRNA-Arg-TAT" in NCBI database was re-anotated as "tRNA-Arg-TCT" because the anticodon was "TCT"
(GCCCCAGTGGCCTAATGGATAAGGCACTGGCCTTCTAAGTCAGGGATTGTGGGTTCGAGTCCCACCTGGGGTG).

Chapter 3

A simplified qPCR method revealing tRNAome remodeling upon infection by genotype 3 hepatitis E virus

Xumin Ou, Buyun Ma, Ruyi Zhang, Zhijiang Miao, Anchun Cheng, Maikel P Peppelenbosch, Qiuwei Pan

Abstract

The landscape of tRNA-viral codons regulates viral adaption at translational level, presumably through adapting host codon usage or modulating host tRNAome. We found that the major zoonotic genotype of hepatitis E virus (HEV) has not adapted to host codon usage, prompting exploration of HEV infection on host tRNAome. However, tRNAome quantification is largely impeded by their extremely short sequence and redundant tRNA genes. Hereby, we present a length-extension and stepwise simplified qPCR method by utilizing a universally DNA/RNA hybrid tRNA adaptor and degenerate primers. Using this novel methodology, we observe that HEV infection dramatically reprograms the hepatic tRNAome that is likely to facilitate translation of viral RNAs. This tRNAome quantification method bears broad implications for future tRNA research and possibly tRNA-based diagnostics.

Keywords: hepatitis E virus; length-extension and simplified qPCR method; tRNAome

Introduction

Viral adaption is shaped by a variety of effectors, including the need to highjack the cellular protein synthesizing machinery for efficient translation of viral genes [1]. Rate-limiting steps in the production of viral proteins differ depending on host species and the virus involved. But in many cell types, efficiency of the translational machinery is critically dependent on the tRNAs availability that decode both cellular and viral codons [2,3]. Accordingly, many viruses show adaptation with respect to codon usage towards their host [4-7]. Conversely, it has been reported that interferons, the potent antiviral cytokines [8], can alter the availability of tRNAs to facilitate the production of antiviral proteins and to attenuate decoding of viral codons [9,10]. Thus, studying the interplay between viral codons and cellular tRNAome is essential for understanding the infection biology.

tRNAs are the recognition modules that decode the mRNA in the ribosome and are covalently charged by amino acids [11]. Following recognition, the corresponding intramolecular bond is hydrolysed and the released amino acid is covalently linked to nascent peptide chain. The tRNAs encoded in genomic DNA are typically 70 to 90 base pairs in length often with multiple genomic copies of the same tRNA [12]. The human tRNAome is encoded by more than 600 tRNA gene locus [12]. Following transcription, tRNAs are subjected to post-transcriptional modification, and a common CCA ribonucleotide sequence is added to the 3' end of tRNA [13]. Accurate detection of the mature tRNAome is technically challenging and especially hampered by the short length and redundant tRNA genes [14,15]. As the number of different tRNA sequences is relatively limited, we believe that optimized qPCR-based techniques hold possibility for quantifying tRNAs.

Hepatitis E virus (HEV) is a ssRNA(+) virus and is the leading cause of acute viral hepatitis [16]. Globally, HEV causes annually 20 million infections with over 56,000 lethal cases and especially in pregnant women [17]. Among the major genotypes, genotype 1 and 2 HEV only infects human, whereas genotype 3 and 4 are zoonotic. Genotype 3 is highly prevalent in western countries with a broad host spectrum [18,19]. In this study, we aim to first develop a simplified qPCR method for quantifying mature tRNAs, and then to investigate how codon usage of the zoonotic HEV genotype relates to the human tRNAome composition.

Materials and methods

Experimental models

For HEV infectious model, a full-length genotype 3 HEV genome (Kernow-C1 p6 clone, GenBank Accession Number: JQ679013) was used [20]. The human hepatoma Huh7 cell line was used to harbor the HEV genome and produce infectious virus particles for secondary infection of naïve Huh7 cells [21]. The HEV subgenomic model was based on Huh7 cells containing the subgenomic genotype 3 HEV sequence (Kernow-C1 p6/luc) coupled to a

Gaussia luciferase reporter gene [21,22]. The cell line was regularly checked for identity using STR verification as provided by the pathology department of the Erasmus MC. Key regents, viral strains and software used in this study were listed in **Table S1**.

Bioinformatic analysis

Codon usage bias and codon adaption index (CAI) of HEV ORF (1-3) and ISGs were analyzed by Codon W software (http://www.molbiol.ox.ac.uk/cu, version 1.4.2) using standard genetic codes. Correlation analysis were performed by Python Matplot package. Codon usage and tRNAome data were visualized by Hemi software with hierarchical clustering analysis (http://hemi.biocuckoo.org/).

The qPCR protocol for quantifying mature tRNAs

Preparation of samples and reagents ●TIMING 1d

1| Prepare primary cells, cell lines or tissues to yield 1-4 μg total RNA for each sample. This amounts to roughly 105-106 mammalian cells. These values are rough guidelines for experimental design, but yields will vary according to the cell types and amounts.

(**Optional**) Protocol can begin with isolated RNA samples. In this case, we recommend to adjust RNA concentration to 214 ng/μL. Then, mix 4 μL RNA sample with 1 μL 5×deacylation buffer (step 9) and incubate at 37 ℃ for 40 min, after that add 10 μL TE buffer for PH adjustment (step 10). Finally, 7 μL of the RNA containing deacylated tRNAs (400 ng ≈ 214 ng/μL × 4/15 ×7 μL) are ready for following annealing with U-adaptor (step 11).

Total RNA isolation and tRNA deacylation ●TIMING 1.5−2h

2| Wash cells. Remove cell medium, and wash cells with warm PBS buffer.

3| Lyse cells. Add 350 μL of RA lysis buffer to each well, then suck up whole cell lysate into 1.5 EP tube. All the reagents for RNA extraction except deacylation buffer are provided by NucleoSpin@ RNA kit.

4| Adjust RNA binding condition. Add 350 μL 70% ethanol to the homogenized lysate and mix by pipetting up and down. (minimal 5 times)

5| RNA binding. For each sample take one Nucleospin® RNA Column place in a collection tube. Pipette lysate up and down 2-3 times and load the lysate to the column. Centrifuge for 30 s at 11,000×g. Place the column on a new collection tube (2 ml).

6| Desalt silica membrane. Add 350 μL MDB (Membrane Desalting Buffer) and centrifuge at 11,000×g for 1 min to dry the membranes.

7| Digest DNA. Prepare DNase reaction mixture in a sterile 1.5 ml tube. For each isolation, add 10 µL DNase to 90 µL reaction buffer for DNase. Mix by flicking the tube. Add 95 µL DNase reaction mixture directly onto the center of the silica membrane of the column. Incubate at room temperature for 15 min.

8| Wash and dry the silica membrane.

  1st Wash: Add 200 µL RAW2 buffer to the Nucleospin® RNA Column. Centrifuge for 30 s at 11,000×g. Place the column on a new collection tube (2 ml).

  2nd Wash: Add 600 µL RA3 buffer to the Nucleospin® RNA Column. Centrifuge for 30 s at 11,000×g. Place the column on a new collection tube (2 ml).

  3rd Wash: Add 250 µL RA3 buffer to the Nucleospin® RNA Column. Centrifuge for 2 min at 11,000×g to dry the membrane completely. Place the column on a new collection tube (2 ml).

9| Deacylation. Prepare deacylation mixture. Add 2 µL of 5×Tris-HCl (100 mM, PH 9.0) into 8 µL ddH$_2$O; mix it by pipetting up and down (minimal 5 times). Add 10 µL Deacylation mixture to the Nucleospin® RNA Column. Incubate at 37 ℃ for 40 min.

10| Elute RNA. Place the column on a new 1.5 EP tube. Elute the RNA by adding an additional 20 µL TE buffer on the Nucleospin® RNA Column directly. Centrifuge for 1 min at 11,000×g.

Adaptor ligation ●TIMING 1.5−2h

11| Annealing. Measuring RNA concentration by Nanodrop. Prepare mixture of 400 ng RNA and 20 pm adaptor (2 µL); then the final volume is adjusted to 9 µL. Incubate at 90 ℃ for 3 min.

12| Add 1 µL 10×Tris-HCl (50 mM, PH8.0) annealing buffer at 37 ℃ for 20 min.

13| Ligation reaction. 2 µL ligase buffer, 7.9 µL H$_2$O and 0.1 µL T4 RNA ligase was added into the above mixture, then incubate at 37 ℃ for 1 h.

Reverse Transcription ●TIMING 0.5h

14| Annealing with mixture of specific tRNA reverse primers. 20 ul of above mixture is added with 3.42 µL (1.14 µL× 3 = 3.42 µL) of mixture of tRNA reverse primers (**Table S2**). Incubate the mixture at 65 ℃ for 5 min. Place on ice immediately.

15| Co-reverse transcript with random and oligo dT Primers. Add 6 µL 5× Prime Script RT Master Mix and 0.58 µL H$_2$O into the above tube. Mix by flicking the tube. Incubate the mixture at 37 ℃ for 15 min followed by 85 ℃ for 5 s.

qPCR ●TIMING 1−2h

16| qPCR mix preparation (**Table S3**). Prepare qPCR mixture according to the following recipe (10 µL). To cover all 57 tRNA types, we recommend adding 220 µL ddH$_2$O into above 30 µL of reverse transcription solution.

Notice: Because of the large number of tRNA sets, it is recommended to prepare tRNA-X-R and tRNA-X-F primer mixture to minimize loading variation. For example, to prepare 80 times of tRNA-R/F mix, we recommend that 20 µL (10 pm) of tRNA-X-R and tRNA-X-F was diluted by 200 µL ddH$_2$O. Therefore, the finally used volume of tRNA primers mix is 3 µL.

17| Perform qPCR with the designed cycling condition according to **Table S4**.

Quantification

The 2-Δ ΔCt was used to calculate the relative expression of each tRNA species, and the ΔCt values were determined by subtracting the average Ct values of the endogenous control gene GAPDH from the average Ct values of the each tRNA type.

Statistical Analysis

Linear correlations between the relative synonymous codon usage (RSCU) of human and animal HEV strains were estimated by Pearson coefficients. Statistical comparisons of tRNAome data were performed with Mann-Whitney U test for non-paired independent samples. *P < = 0.05 was indicated as significant. The statistical analysis was performed using the SPSS19.0 software.

Results

The codon usage of genotype 3 HEV has not adapted to its human host

As in many biological systems tRNA availability [23] and codon context [24] constitute the rate limiting step for protein synthesis. Viruses are expected to adapt codon usage as to fit host codon usage patterns. With respect to the zoonotic genotype 3 HEV, however, this has not yet been analyzed. Hence, we compared the RSCU of three open reading frames (ORFs) of genotype 3 HEV isolated from human to the counterpart of the human host. Similar analysis was also performed for HEV strains isolated from animals, including swine, deer, mongoose and wild boar (Table 1). Strikingly, we did not find evidence that the codon usage of genotype 3 HEV isolated from human and the animal host has adapted to the human codon usage pattern for any of the three ORFs investigated (**Fig. 1**). Because linear correlations between the RSCU of human and animal HEV strains are very poor (Pearson coefficients less than 0.40). Moreover, the RSCU analysis of HEV between the non-human primate (e.g., monkey) and human also indicated that the usage of major codons is similar except serine codon (UCU) (**Fig. S1**). Besides RSCU, the codon adaption index (CAI) has also been suggested to predict the efficiency of translation elongation [25]. Our results indicated that the CAI of HEV was similar

among different host species-derived strains, and much lower than that of human housekeeping gene (e.g., GAPDH) (**Fig. 1**). We conclude that there has been very little evolutionary pressure on genotype 3 HEV to adapt the codon usage of its human host. Thus, we decided to investigate how HEV infection regulates host tRNAome.

Table 1. list of HEV genotype 3 strains used in this study.

Host	Source material	Country of origin	Collection data	Accession #
Homo sapiens	Hepatocytes	United Kingdom	Jul-2010	JQ679013
Sus domesticus	Serum	United States	1997	AF082843
Sus scrofa	Liver	Germany	2006	FJ705359
Cervus nippon	Liver or serum	Japan	22 Feb 2003	AB189071
Herpestes javanicus	Serum	Japan	2002	AB236320
Cynomolgus	Serum	Japan	Nov-2009	JQ026407.1

A simplified qPCR method for characterizing the human mature tRNAome

To develop a novel qPCR method for easy quantification of mature tRNAs, we first designed a DNA/RNA hybrid adaptor (Universal adaptor, U-adaptor). It is compatible with all types of mature tRNAs with a common 3' CCA acceptor (**Fig. 2**). Following binding of the adaptor to the tRNA molecule, the adaptor can be ligated to both the 5' end and 3' end of the mature tRNAs. Of note, the mature tRNAs are frequently charged by amino acids that hamper the above ligation reaction. Thus, a deaminoacylation reaction for the charged tRNA is needed during RNA isolation (see Materials and methods). Importantly, the adaptor is designed to form a loop and hybridize on tRNA itself. This hybridization is compatible for any types of base-pairs adjacent to 3' CCA tail of tRNAs, as the pairing base of the U-adaptor involved is degenerately designed. The ligation reaction specifically links the 5' end and 3' end of the adaptor to the corresponding ends of tRNA molecule that is catalyzed by T4 RNA ligase (**Fig. 2**). Finally, a hybrid pseudocircular molecule, containing the deaminoacylated original mature tRNA molecule and the U-adaptor, is formed. This strategy serves two important ends: it extends length of tRNA for subsequent qPCR while simultaneously eliminating unmature tRNAs. Afterwards, tRNAs can be determined by conventional qPCR (see the experimental strategy **Fig. 3**).

To this end, a mixture of tRNA specific reverse primers was used, which were designed to bind to the region between the anticodon loop and the D loop. A complication with regard to designing reverse primers for characterizing the tRNAome is that many iso-decoder tRNAs (i.e., tRNA with the same anticodon but vary in backbone) are responsible for decoding of the identical genetic codon [12]. We thus retrieved all 57-human genomic tRNA genes from the GtRNA 2.0 database and accordingly designed degenerate primers (**Table S2**). Using RNA obtained from Huh7 human liver cells, all primer pairs successfully amplified their iso-decoder

tRNA molecules in qPCR reactions (**Fig. S2** and **Fig. S3**). We conclude that our approach employing the U-adaptor and degenerate primers allows easy quantification of cellular tRNA profiles.
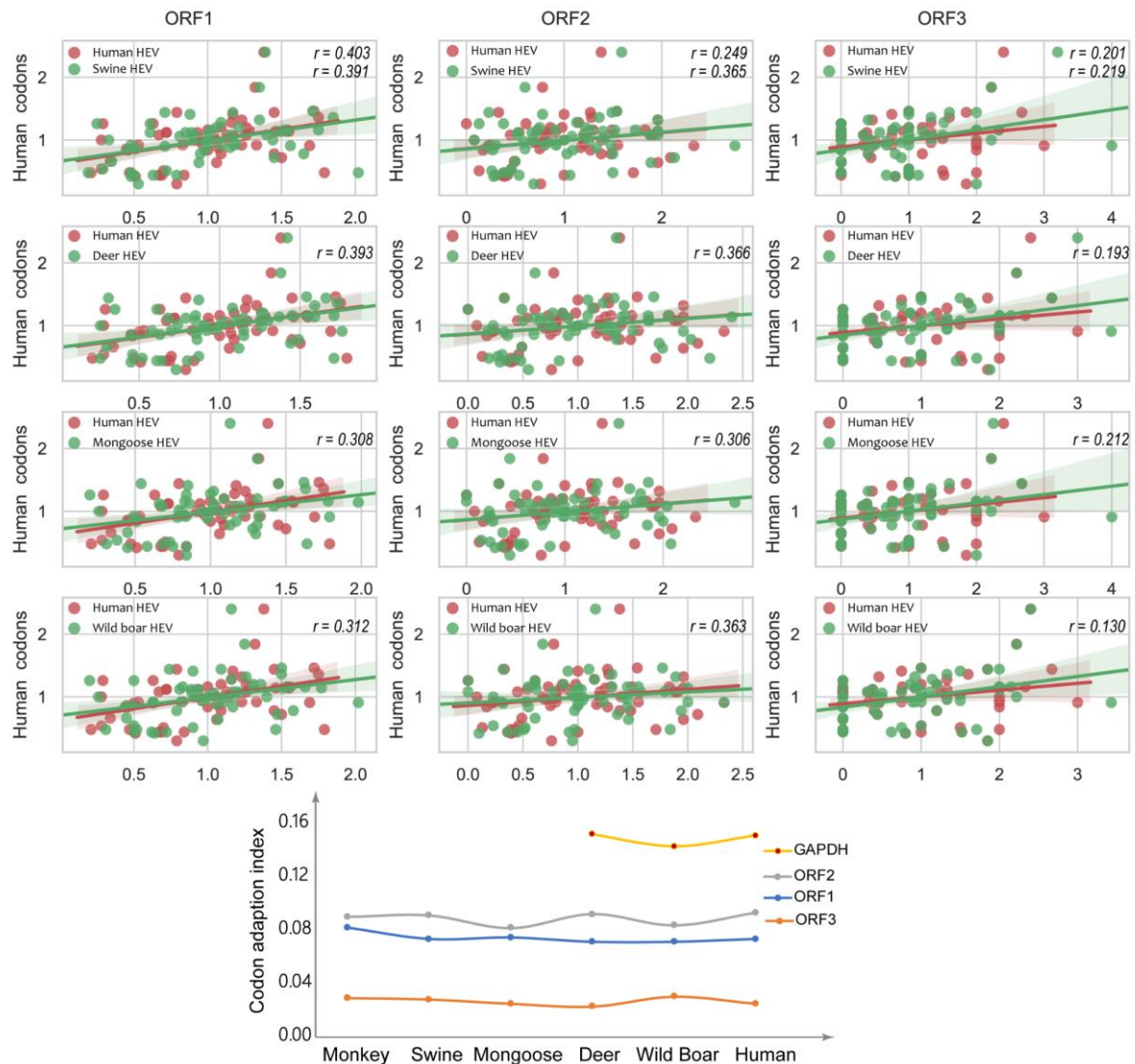

Fig. 1. Lack of adaptation of genotype 3 HEV codon usage to host codon usage. Codon usage of human HEV was correlated to overall human codon usage. The codon usage of swine, wild deer, mongoose and wild boar HEV was also related to human codon usage. For each possible codon and 64 in total, its RSCU in the human genome is plotted against the Y axis and the corresponding RSCU in the relevant viral ORF on the X axis. Results were shown for all three ORFs in the upper panels. In the lower panel, codon adaption index of genotype 3 HEV isolated from different host species were compared for ORF1-3. The CAI of GAPDH from deer, wild boar and human are also indicated. The CAI calculation was performed by Codon W software with Saccharomyces cerevisiae as reference set.
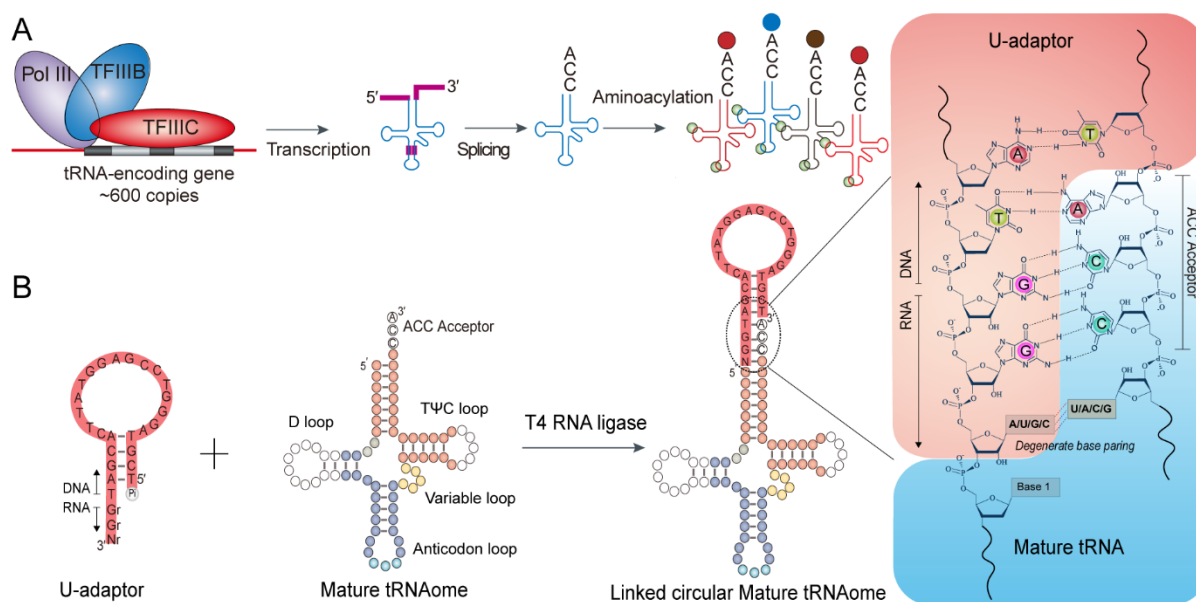
Fig. 2. A universal U-adaptor that interacts specifically with all mature tRNAs. A) More than 600 tRNA-encoding genes can be transcribed by RNA polymerase III. After post-transcriptional processing, splicing and adding of ACC acceptor to these new tRNA transcripts. The resulting mature tRNAome is available for all cellular mRNA translation and also for viral RNA decoding. The final step in tRNA maturation is adding the CCA sequence to which the acylating amino acid is coupled. The CCA tail is thus common and specific to all mature tRNAs. **B**) A universal(U) DNA/RNA hybrid adaptor was designed. The -2 to -4 of 3' terminal bases (TGG) of the U-adaptor are specifically designed to pair CCA acceptor. To further improve the compatibility of U-adaptor to all mature tRNAs, the last 3' terminal nucleotide was degenerately designed. The interface of U-adaptor and mature tRNA are detailed at the right of the panel.

Normalization of tRNAome by reference gene

In order to normalize tRNA level, tRNA specific reverse primers, random and oligo d(T) primers were concurrently used for reverse transcriptional reaction. GAPDH, a commonly accepted reference gene, is selected for tRNAome normalization [26]. tRNA-His-GUG is the only tRNA species capable of decoding Histidine codons (CAC & CAU), and thus we expect that its expression would be more stable. Indeed, we observed good correlation between levels of GAPDH mRNA and tRNA-His-GUG (**Fig. S4**). Comparing the U-adaptor and tRNA-histidine adaptor for ligation reaction, we found that the U-adaptor did not compromise the ligation efficiency as it caused only 0.11 variation calculated by ΔCt (**Table S1**). Efforts were made for further optimization of the protocol, for instance by combining steps involved annealing, but these efforts failed (**Fig. S5**). The final protocol (refer to Materials and methods) is organized in seventeen discrete steps and allows quantitative assessment of mature tRNAome.
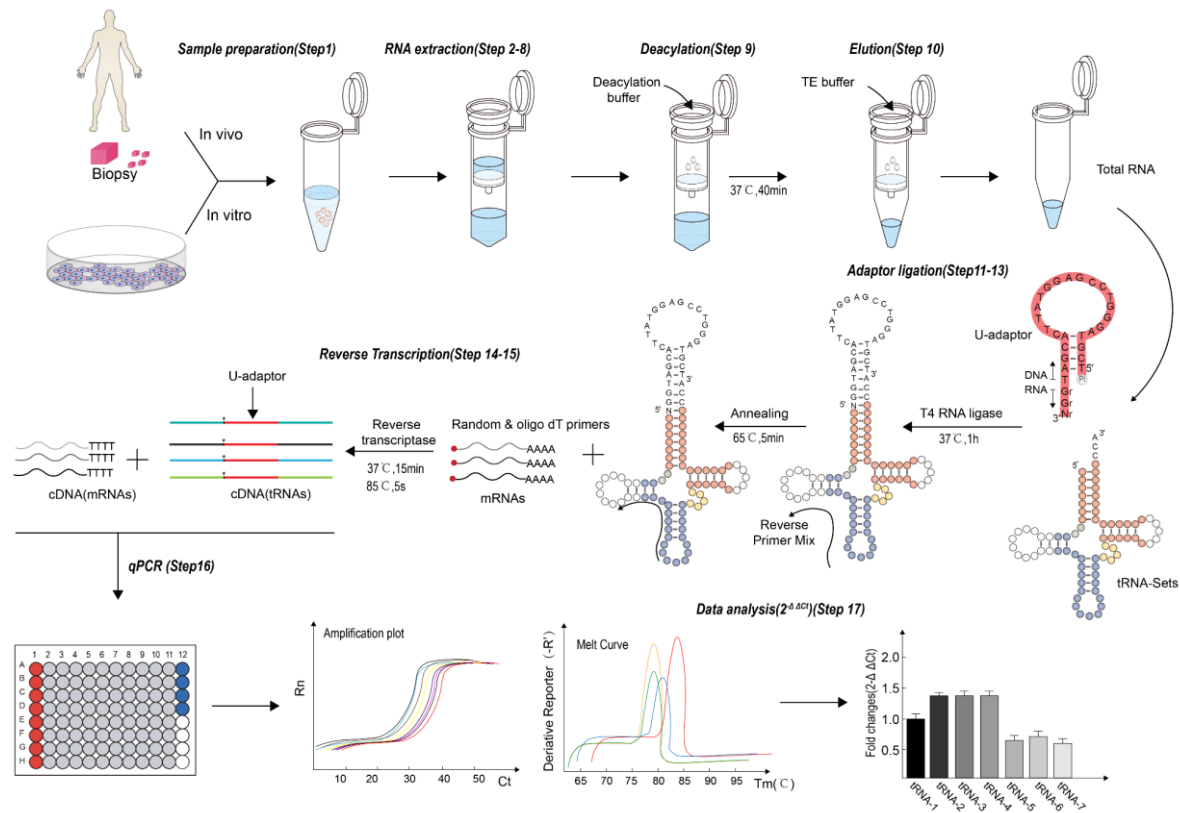
Fig. 3. Workflow of quantifying tRNAome. First total RNA was extracted from the experimental systems involved according to a routine column-based RNA isolation protocol. Before elution the column is incubated with diacylation buffer at 37 °C for 40 min to remove acylating amino acids at tRNA CCA tail followed by elution with TE buffer. Subsequently, the tRNA and U-adaptor are linked by T4 RNA ligase followed by annealing with reverse primer mix. Finally, the cDNA of total mRNA and U-adaptor linked tRNAs are concurrently synthesized for qPCR amplification. The relative expression of tRNAs is calculated by 2-ΔΔCt method that can be normalized by housing keeping genes, such as GAPDH. The stepwise procedures are detailed by seventeen discrete steps in the Materials and Methods.

HEV infection provokes remodeling of mature tRNAome in host cells

Although viruses are considered to adapt codon usage to their host [3,27], we observe that HEV is not subject to such adaptation. Hence, we decided to compare tRNAome composition in the presence and absence of HEV infection. To this end Huh7 cells were infected with full length HEV or exposed to vehicle control (**Fig. S4A**), and the levels of different tRNAs were determined (**Fig. S6A**). We observed a major effect on tRNAome remodeling by HEV infection. 29 out 57 tRNA species showed significant upregulation. The most marked example is tRNA-Pro-GGG that shows 162 times higher expression in HEV-infected cells as compared to non-infected cells. tRNA-Gly-CCC was approximately 50 times and tRNA-Pro-UGG was about 35 times upregulated in HEV infected cells. Conversely, other tRNA species were downregulated following HEV infection. tRNA-Leu-UAG is most prominently downregulated by approximately

70 %. We conclude that HEV infection provokes major remodeling of the mature tRNAome in cells capable of sustaining its replication.

HEV infection triggered tRNAome remodeling is largely dependent on ORF2

The capsid protein of HEV encoded by ORF2 features many ß-sheets in its secondary structure [28] and consequently requires incorporation of a multitude of Proline and Glycine residues during synthesis. By far the most prominent effects of HEV infection on the tRNAome are the upregulation of tRNAs supporting Proline and Glycine decoding (tRNA-Pro-GGG, tRNA-Pro-UGG, and tRNA-Gly-CCC). Thus, we investigated the role of ORF2 in HEV-induced tRNA remodeling. We quantified tRNAome in cells carrying subgenomic HEV replicon with ORF2-deletion. This subgenomic HEV replicon contains a luciferase reporter and viral replication can be quantified by luciferase activity (**Fig. 4A**). Intriguingly, although effects on tRNAome composition were still observed in the subgenomic replicon (**Fig. S6B**), they are much moderate as compared to that in the full-length HEV model (**Fig. S6C**). Effects on tRNA-Pro-GGG, tRNA-Pro-UGG, and tRNA-Gly-CCC levels were not statistically significant in the subgenomic replicon.

Further insight came from unsupervised clustering analysis of the tRNAome profiles [29]. Three biologically independent experiments were performed with naïve Huh7 cells, Huh7 cells infected by HEV with the full-length genome and cells carrying the subgenomic replicon. The experimental conditions segregated the profiles generated. As expected, the effects on tRNAome provoked by infection of HEV with full-length genome were much more pronounced as compared to those evoked by the subgenomic replicon (**Fig. 4B**). We thus conclude that ORF2 is essential for HEV-triggered tRNA remodelling.

HEV-induced tRNA remodelling may counteract decoding of interferon-stimulated genes

Innate antiviral immunity, especially that mediated by the so-called interferon-stimulated genes (ISGs), is important for controlling HEV infection [30]. Interestingly, tRNAome remodeling by HEV mainly supported better decoding of ORF2 when compared to ORF1 and ORF3 (**Fig. 4C, D**). Hence, we analyzed the effects of tRNAome remodeling triggered by full-length infectious HEV on the decoding of essential ISGs. We selected a panel of ISGs known to inhibit HEV infection. We found that HEV-provoked tRNA remodeling is not correlated to the RSCU of these ISGs (**Fig. 4C** and **Table S5**), whereas viral ORF2 translation per se is related to the same changes (**Fig. 4C, D**). We thus conclude that HEV infection reprograms cellular tRNAome that is likely to facilitate viral translation but hamper cellular antiviral immunity.
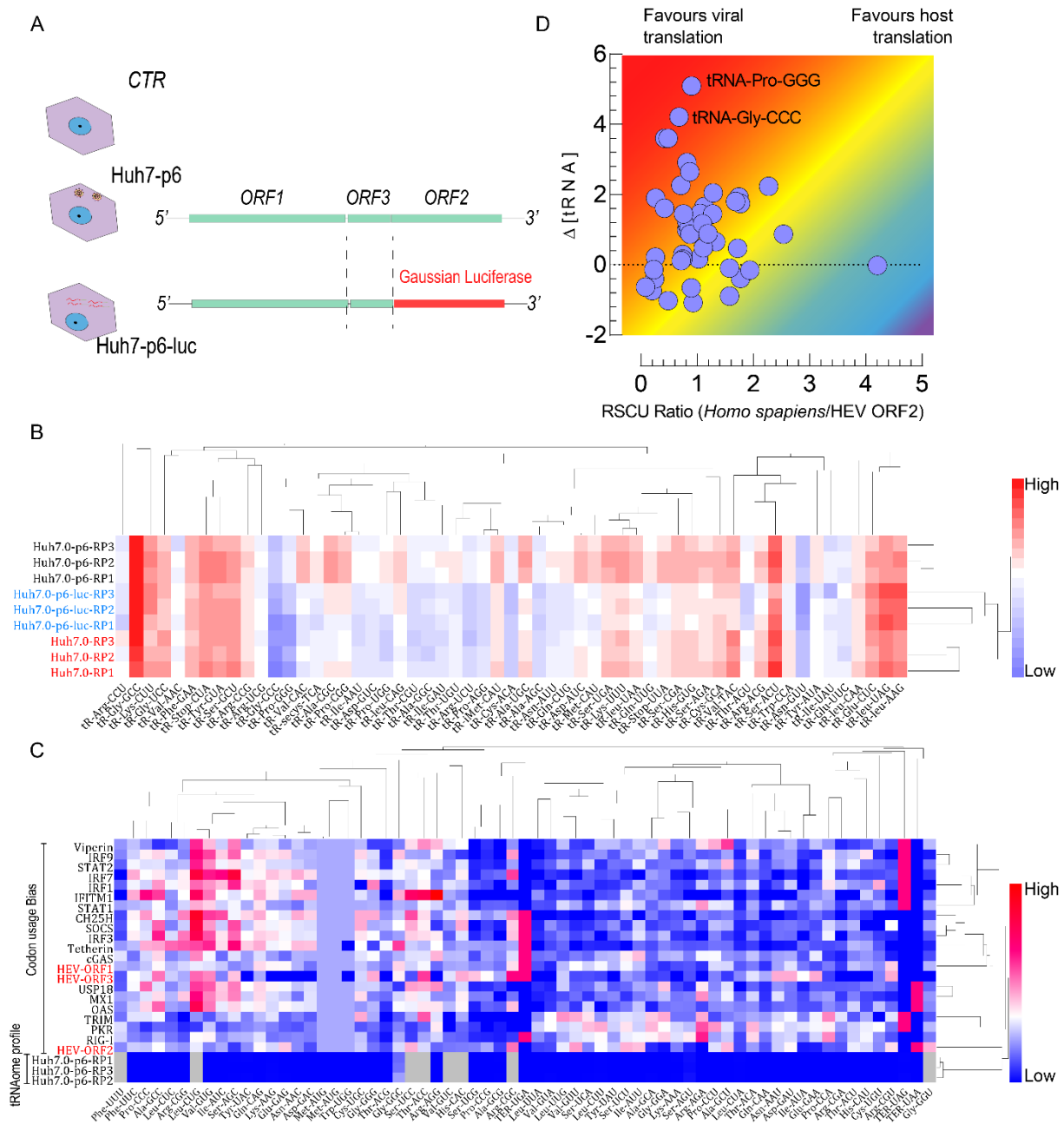
Fig. 4. tRNAome remodeling following HEV infection favors viral translation. A) For profiling tRNAome, human hepatoma Huh7 cells harboring the infectious HEV clone with full-length genome (Huh7-p6), the subgenomic replicon lacking of ORF2 (Huh7-p6-luc), and naïve Huh7 cells were used (n = 3). **B**) Cluster analysis of the tRNAome profiles in the three cell models. **C**) Comparative analysis of the effects of HEV-induced tRNAome remodeling on codon usage of viral ORF (1-3) and a panel of antiviral interferon-stimulated genes (ISGs). **D**) Visualization of the effects of HEV infection on relative tRNA abundancies in relation to viral and host codon decoding. For each tRNA, the ratio of HEV codon usage over human codon usage is plotted on the X axis. A larger value indicates that a tRNA is more often used for human decoding as compared to viral decoding. The effect of HEV infection on tRNA abundancy is plotted on the Y axis. A positive value corresponds to upregulation of a tRNA following HEV infection.

Discussion

The development of tRNA quantification methods has been greatly facilitated by the tRNAscan-SE program that allows accurate identification of genomic tRNA sequences. Earlier methods to detect tRNAs usually involve thin layer chromatography [31], liquid chromatography mass spectrometry [32], DNA arraying [33] and single tRNA-based qPCR [15]. Recently, quantification of the tRNAome at transcriptional level has become more efficient through high-throughput sequencing [14,34], but is dependent on programming and very specific reagents. In this study, we established a simplified qPCR method for rapid quantification of the mature tRNAome. This strategy is characterized by a length-extension step, a universally compatible adaptor and degenerate primers applicable for the entire mature tRNAome. The present study was only demonstrated in a human cell model, but we envision that it is applicable to other organisms by adjusting primer design.

As a proof-of-principle, this qPCR methodology was used for characterizing the effects of HEV infection on host tRNAome. We observed substantial remodeling of tRNA composition following HEV infection. For example, the level of tRNA-Pro-GGG was increased by 162 times by HEV infection. Such regulations appear in a fashion that facilitates translation of viral proteins while simultaneously hampering the decoding of antiviral ISGs. Interestingly, these observed effects relate especially to HEV ORF2, which is the dominant ORF of this virus with respect to demands on the host translational machinery [35]. Our data suggest that counteracting HEV-dependent tRNA remodeling may constitute a novel avenue for supporting host defense in combating viral infection, but this will require further understanding of the molecular pathways involved.

It has been reported that codon usage bias of the zoonotic genotypes 3 and 4 of HEV is much weaker than that of the non-zoonotic genotypes 1 [19]. This indicates that less bias of HEV codon usage may be involved in zoonotic infection. Consistently, we found that genotype 3 HEV has not adapted codon usage to its human host. This suggests that an alternative pathway may be used by HEV to facilitate cross-species infection, such as remodeling of tRNAome. It is well-known that viral infection can highjack host cell physiology for corrupting host defense and facilitating viral protein synthesis which is dependent on cellular mature tRNAome.

Recent studies have indicated that tRNAome disturbance has broad implications in many diseases, especially in cancer development and metastasis [23,36]. Further understanding the role of tRNAs in biology and pathogenesis requires easy techniques to detect and quantify these molecules. In this respect, the U-adapter strategy explored in the present study may provide new impetus and may form the basis of novel diagnostic possibilities. Hence, we envision that a convenient method, as the one presented in the current study, may prove instrumental for many aspects of contemporary biomedical research and practice with respect to tRNA biology.

References

[1]   Stern-Ginossar, N. and Ingolia, N.T. (2015). Ribosome Profiling as a Tool to Decipher Viral Complexity. Annual Review of Virology, Vol 2 2, 335-349.

[2]   Novoa, E.M., Pavon-Eternod, M., Pan, T. and de Pouplana, L.R. (2012). A Role for tRNA Modifications in Genome Structure and Codon Usage. Cell 149, 202-213.

[3]   Ou, X.M., Cao, J.Y., Cheng, A.C., Peppelenbosch, M.P. and Pan, Q.W. (2019). Errors in translational decoding: tRNA wobbling or misincorporation? Plos Genetics 15

[4]   Coleman, J.R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E. and Mueller, S. (2008). Virus attenuation by genome-scale changes in codon pair bias. Science 320, 1784-7.

[5]   Lauring, A.S., Acevedo, A., Cooper, S.B. and Andino, R. (2012). Codon Usage Determines the Mutational Robustness, Evolutionary Capacity, and Virulence of an RNA Virus. Cell Host & Microbe 12, 623-632.

[6]   Shen, S.H. et al. (2015). Large-scale recoding of an arbovirus genome to rebalance its insect versus mammalian preference. Proceedings of the National Academy of Sciences of the United States of America 112, 4749-4754.

[7]   Eschke, K., Trimpert, J., Osterrieder, N. and Kunec, D. (2018). Attenuation of a very virulent Marek's disease herpesvirus (MDV) by codon pair bias deoptimization. Plos Pathogens 14

[8]   Li, Y., Qu, C., Yu, P., Ou, X., Pan, Q. and Wang, W. (2019). The Interplay between Host Innate Immunity and Hepatitis E Virus. Viruses 11

[9]   Li, M.Q. et al. (2012). Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. Nature 491, 125-145.

[10]   Smith, B.L., Chen, G.F., Wilke, C.O. and Krug, R.M. (2018). Avian Influenza Virus PB1 Gene in H3N2 Viruses Evolved in Humans To Reduce Interferon Inhibition by Skewing Codon Usage toward Interferon-Altered tRNA Pools. Mbio 9

[11]   Crick, F.H. (1958) On protein synthesis. In Symp Soc Exp Biol ed.^eds), pp. 8

[12]   Chan, P.P. and Lowe, T.M. (2016). GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Research 44, 184-189.

[13]   Kuhn, C.D., Wilusz, J.E., Zheng, Y.X., Beal, P.A. and Joshua-Tor, L. (2015). On-Enzyme Refolding Permits Small RNA and tRNA Surveillance by the CCA-Adding Enzyme. Cell 160, 644-658.

[14]   Zheng, G.Q., Qin, Y.D., Clark, W.C., Dai, Q., Yi, C.Q., He, C., Lambowitz, A.M. and Pan, T. (2015). Efficient and quantitative high-throughput tRNA sequencing. Nature Methods 12, 835-837.

[15]   Honda, S., Shigematsu, M., Morichika, K., Telonis, A.G. and Kirino, Y. (2015). Four-leaf clover qRT-PCR: A convenient method for selective quantification of mature tRNA. Rna Biology 12, 501-508.

[16]   Kamar, N., Izopet, J., Pavio, N., Aggarwal, R., Labrique, A., Wedemeyer, H. and Dalton, H.R. (2017). Hepatitis E virus infection. Nature Reviews Disease Primers 3, 17086.

[17]   Nimgaonkar, I., Ding, Q., Schwartz, R.E. and Ploss, A. (2018). Hepatitis E virus: advances and challenges. Nature Reviews Gastroenterology & Hepatology 15, 96-110.

[18]   Meng, X.J. (2016). Expanding Host Range and Cross-Species Infection of Hepatitis E Virus. PLoS Pathog 12, e1005695.

[19]   Zhou, J.-h., Li, X.-r., Lan, X., Han, S.-Y., Wang, Y.-n., Hu, Y. and Pan, Q. (2019). The genetic divergences of codon usage shed new lights on transmission of hepatitis E virus from swine to human. Infection, Genetics and Evolution 68, 23-29.

[20]   Shukla, P. et al. (2011). Cross-species infections of cultured cells by hepatitis E virus and discovery of an infectious virus-host recombinant. Proceedings of the National Academy of Sciences of the United States of America 108, 2438-2443.

[21]   Shukla, P., Nguyen, H.T., Faulk, K., Mather, K., Torian, U., Engle, R.E. and Emerson, S.U. (2012). Adaptation of a genotype 3 hepatitis E virus to efficient growth in cell culture depends on an inserted human gene segment acquired by recombination. J Virol 86, 5697-707.

[22]   Zhou, X.Y., Wang, Y.J., Metselaar, H.J., Janssen, H.L.A., Peppelenbosch, M.P. and Pan, Q.W. (2014). Rapamycin and everolimus facilitate hepatitis E virus replication: Revealing a basal defense mechanism of PI3K-PKB-mTOR pathway. Journal of Hepatology 61, 746-754.

[23]   Goodarzi, H., Nguyen, H.C.B., Zhang, S., Dill, B.D., Molina, H. and Tavazoie, S.F. (2016). Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. Cell 165, 1416-1427.

[24]   Gamble, C.E., Brule, C.E., Dean, K.M., Fields, S. and Grayhack, E.J. (2016). Adjacent Codons Act in Concert to Modulate Translation Efficiency in Yeast. Cell 166, 679-690.

[25]   Coghlan, A. and Wolfe, K.H. (2000). Relationship of codon bias to mRNA concentration and protein length in Saccharomyces cerevisiae. 16, 1131-1145.

[26]   van de Moosdijk, A.A.A. and van Amerongen, R. (2016). Identification of reliable reference genes for qRT-PCR studies of the developing mouse mammary gland. Scientific Reports 6, 35595.

[27]   Berkhout, B. and van Hemert, F. (2015). On the biased nucleotide composition of the human coronavirus RNA genome. Virus Research 202, 41-47.

[28]   Guu, T.S.Y., Liu, Z., Ye, Q.Z., Mata, D.A., Li, K.P., Yin, C.C., Zhang, J.Q. and Tao, Y.J. (2009). Structure of the hepatitis E virus-like particle suggests mechanisms for virus assembly and receptor binding. Proceedings of the National Academy of Sciences of the United States of America 106, 12992-12997.

[29]   Johnson, S.C. (1967). Hierarchical clustering schemes. Psychometrika 32, 241-254.

[30]   Wang, W.S. et al. (2018). The RNA genome of hepatitis E virus robustly triggers an antiviral interferon response. Hepatology 67, 2096-2112.

[31]   Spears, J.L., Gaston, K.W. and Alfonzo, J.D. (2011) Analysis of tRNA editing in native and synthetic substrates. In RNA and DNA Editing ed.^eds), pp. 209-226. Springer

[32]   Gaston, K.W. and Limbach, P.A. (2014). The identification and characterization of non-coding and coding RNAs and their modified nucleosides by mass spectrometry. Rna Biology 11, 1568-1585.

[33]   Grelet, S., McShane, A., Hok, E., Tomberlin, J., Howe, P.H. and Geslain, R. (2017). SPOt: A novel and streamlined microarray platform for observing cellular tRNA levels. Plos One 12

[34]   Schwartz, M.H. et al. (2018). Microbiome characterization by high-throughput transfer RNA sequencing and modification analysis. Nature Communications 9, 5355.

[35]   Ankavay, M. et al. (2019). New insights into the ORF2 capsid protein, a key player of the hepatitis E virus lifecycle. Scientific Reports 9, 6243.

[36]   Grewal, S.S. (2015). Why should cancer biologists care about tRNAs? tRNA synthesis, mRNA translation and the control of growth. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms 1849, 898-907.
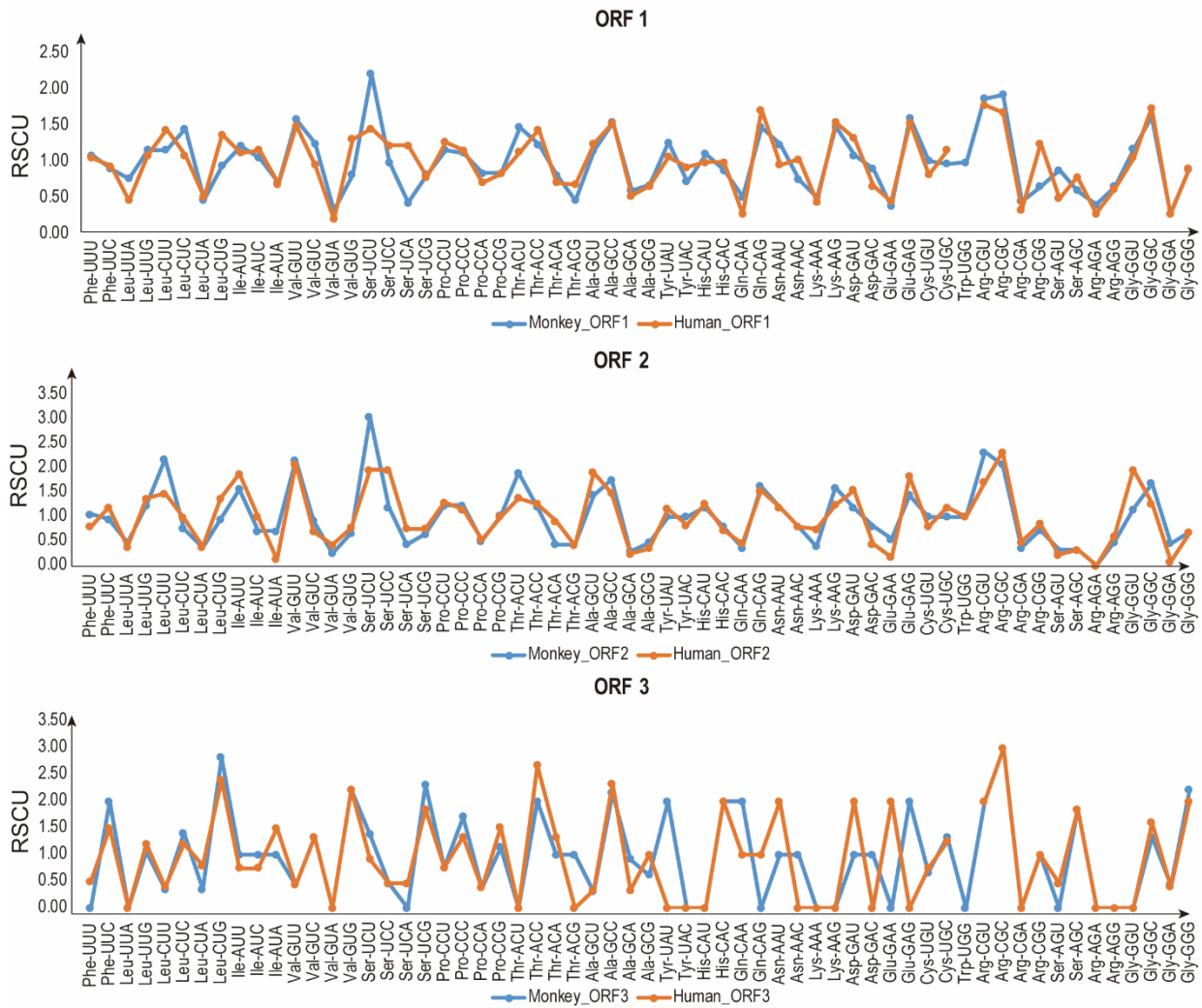
Supplementary information

Fig. S1. Comparative analysis of codon usage bias of HEV isolated from monkey and human. ORF1, ORF2 and ORF3 of genotype 3 HEV isolated from monkey and human are compared. The majority of codon usage bias are similar between monkey and human according to the index of RSCU.
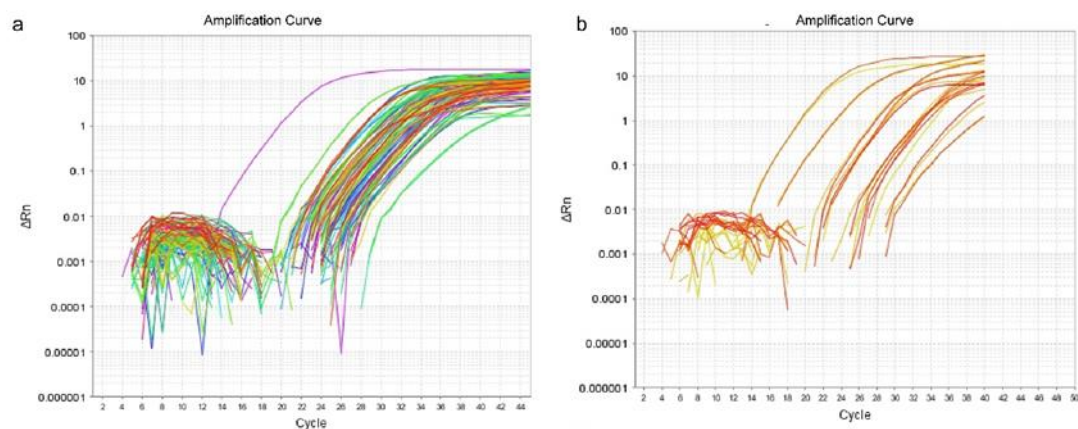
Fig. S2. Usefulness of U-adaptor to detect 57 types of tRNA sets in human genome. RNA was isolated from Huh7 cells, deaminpacylated and ligated to the U-adaptor (see **Fig. 3**) before being subjected to PCR for specific tRNA species. A) Amplification of tRNA types 1-47 (see table S1). B) Amplification of tRNA type 48-57 (see table S1). The entire complement of 57 types of genomic tRNAs displayed Ct value less as 35 (threshold, 0.1) and we were able to calculate relative abundancy of the different tRNAs present in the RNA of Huh7 cells.
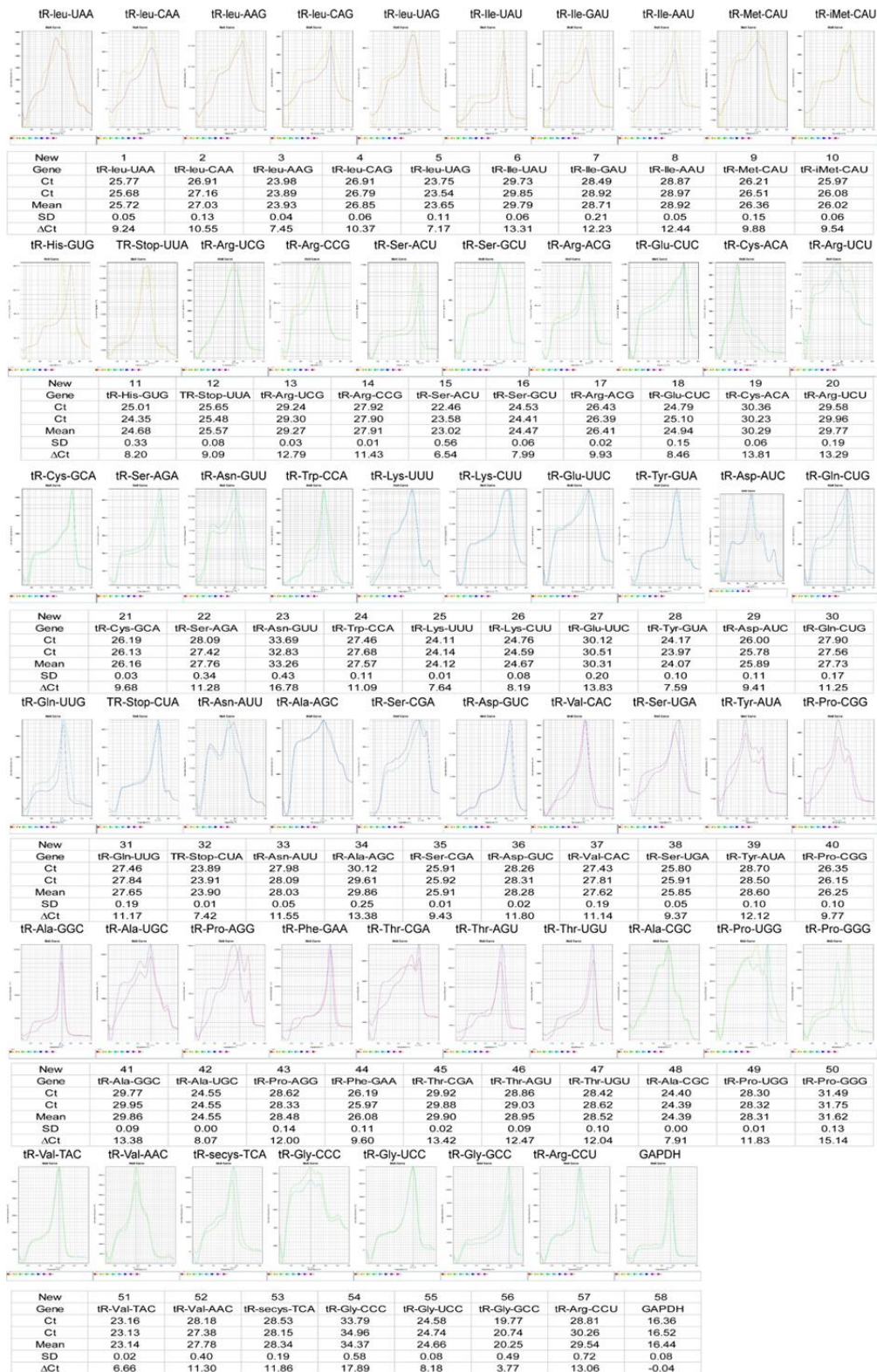
Fig. S3. Melt curves and variation analysis of 57 tRNA types. Amplicons of tRNA transcripts display single peaks and reproducible signals.
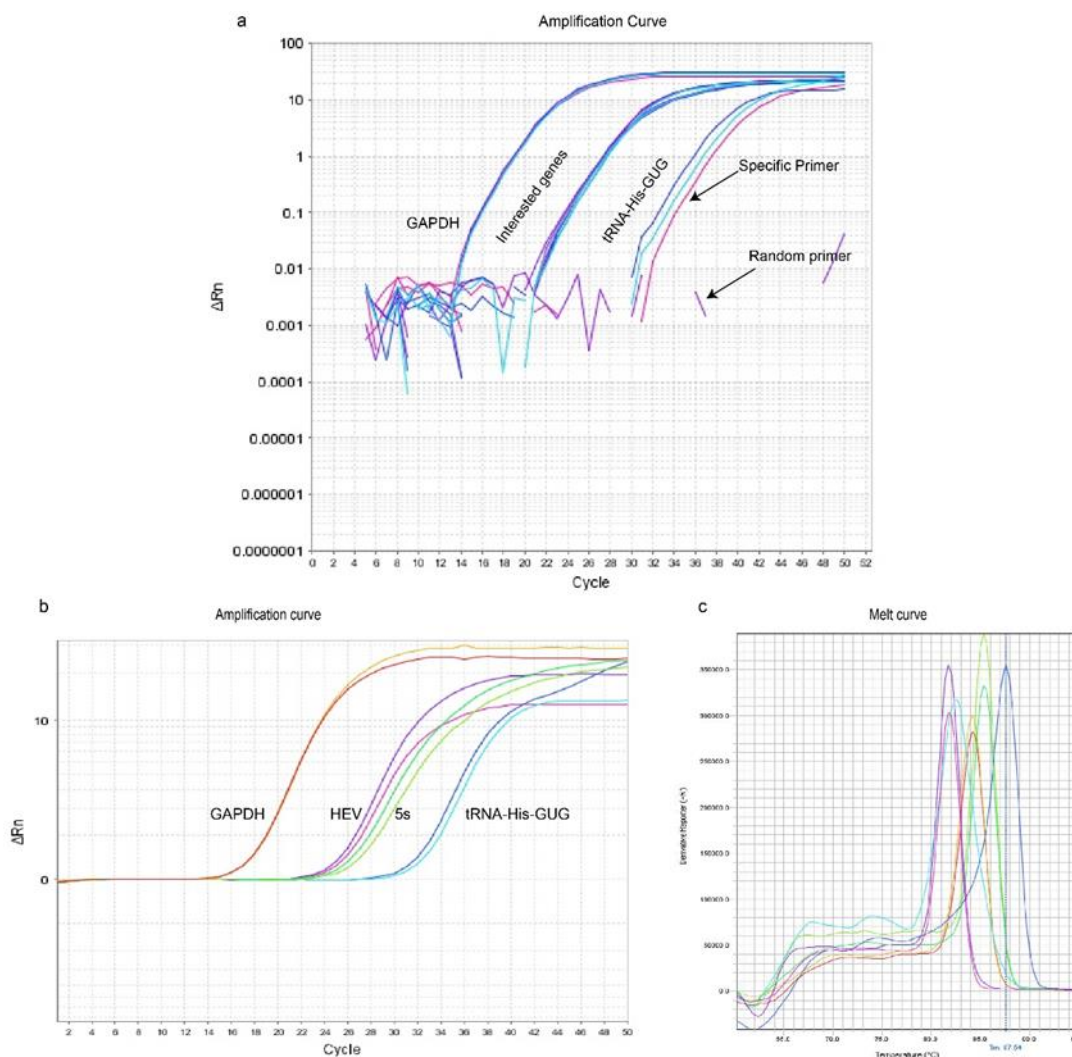
Fig. S4. Specific and quantifiable expression of tRNAs that can be related to HEV infection levels. Huh7 cells infected with HEV were subjected to the tRNAome detection protocol used in this study. A) Random primers do not allow amplification of products upon reverse transcription, suggesting that product formation is specific. Despite the deacylation, the RNA isolated allows following reverse transcription reproducible amplification of a GAPDH-specific product which allows normalization of samples based on mRNA levels. tRNA-His-GUG is the only tRNA species capable of decoding His-coding codons (CAA & CAU) and thus we considered its expression unlikely to be influenced by factors that would drive alternative expression of different tRNAs capable of decoding the same amino acid. The panel also shows reliable detection of this tRNA, which thus can serve for normalization purposes as well. B) HEV can also be detected in the same RNA preparation that allows detection of tRNA levels and GAPDH determinations and hence results can be linked to directly linked to HEV infection levels as well. C) The melt curves corresponding to the data presented in panel b shows the specific nature of the amplification involved.
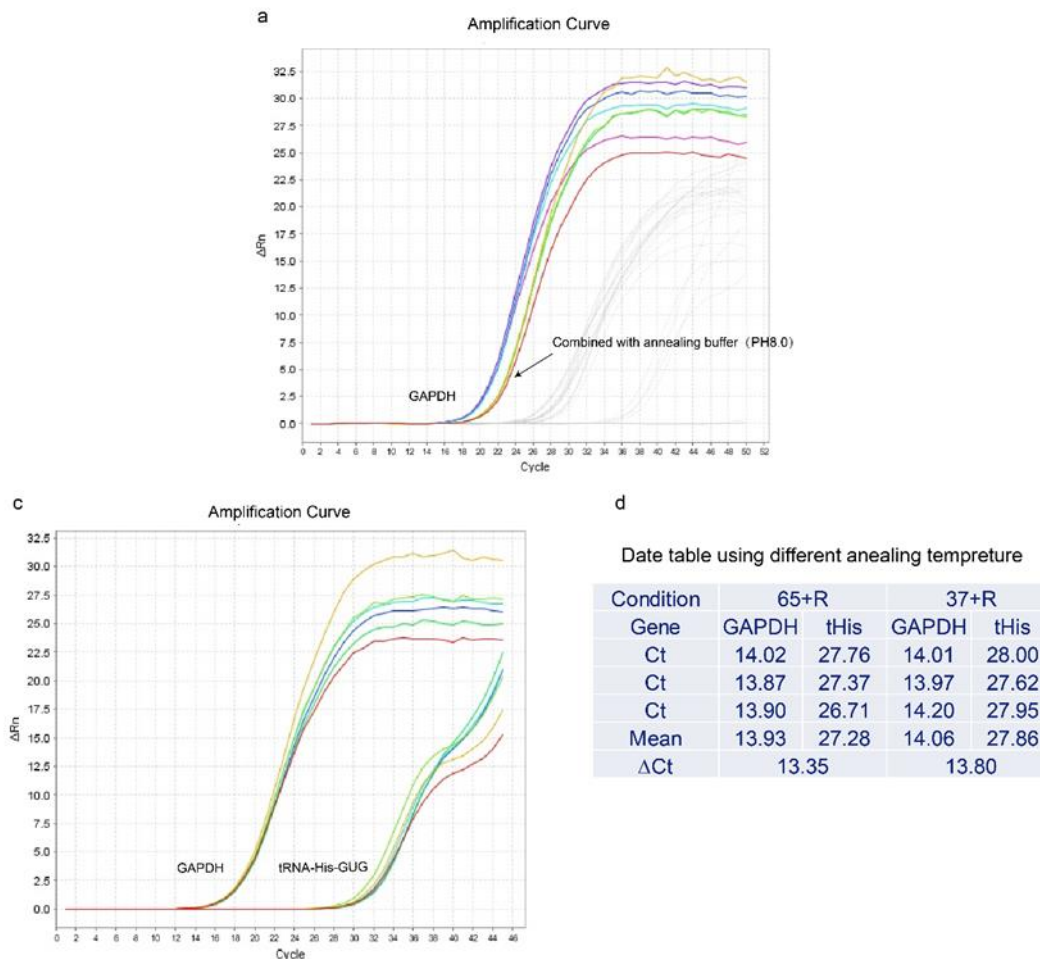
Fig. S5. Failed efforts aimed at protocol optimization. A) Huh7 cells infected with HEV were subjected to the tRNAome determination, either by the original protocol or through a protocol in which step 11 and 12 were combined. The latter, however, increased mRNA degradation. **B)** Annealing temperature for reverse transcription has a slight impact on PCR efficiency. **C)** Comparative analysis of ΔCt values by using different annealing temperatures for linked tRNA and specific reverse primers (65 ℃ and 37 ℃) had slight impact on PCR efficiency.
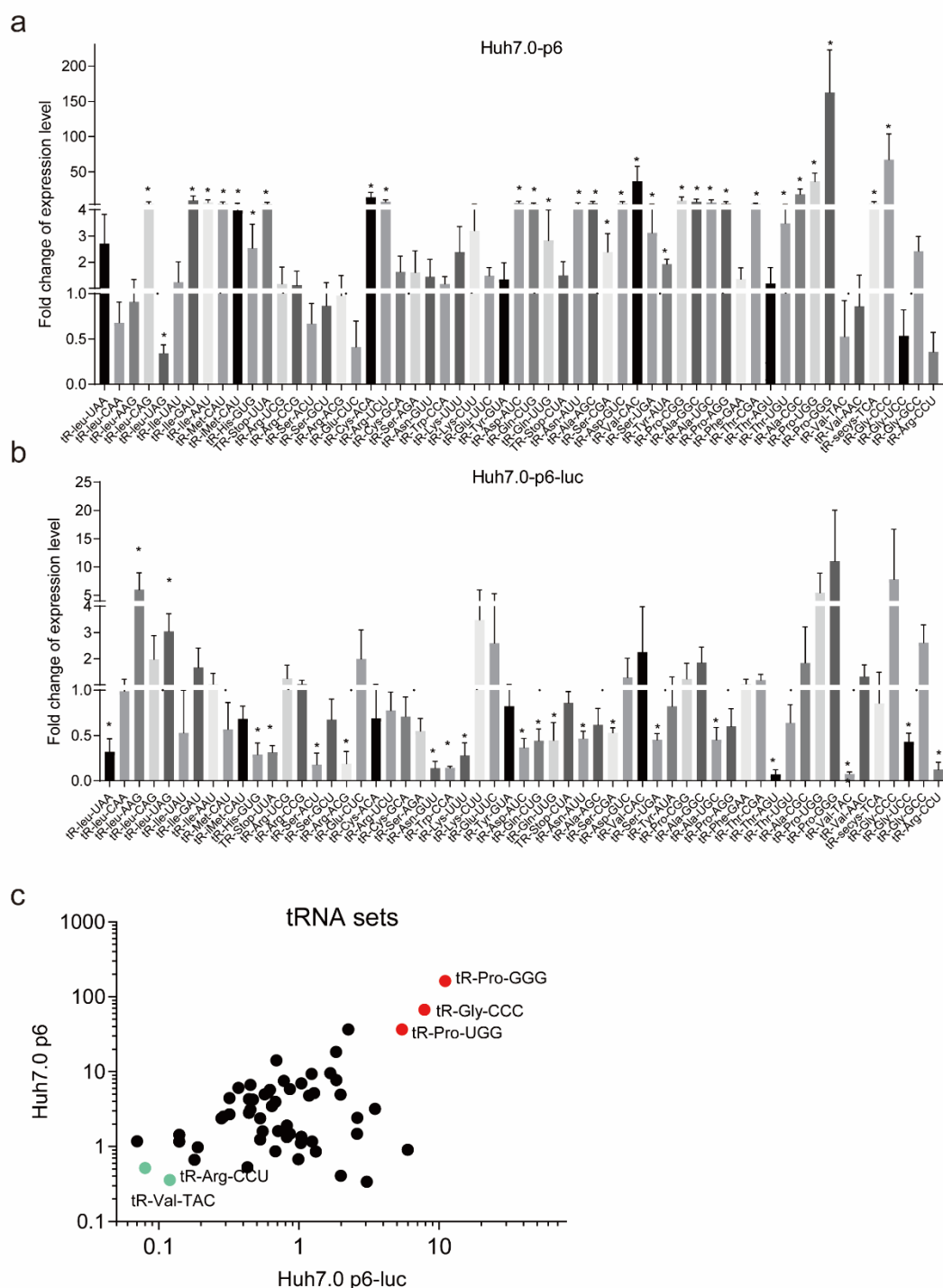
Fig. S6. **Relative transcription levels of 57 tRNA types following infection with either full length HEV or an ORF2-lacking subgenomic replicon.** In this figure a tRNA expression level of 1 indicates that levels following experimental intervention are unchanged as compared to uninfected Huh7 cells, a value between 0 and 1 indicates downregulation of tRNA levels following infection, whereas values in excess of 1 indicate upregulation of the specific tRNA species. **A)** Following infection with full length HEV (upper half of the panel) major and highly significant remodeling the tRNAome occurs (note that Y axis runs up to 200). **B)** Following infection with the HEV subgenomic replicon (Huh7-p6-Luc; lower half of the panel) remodeling is much less pronounced (Y axis runs to 20), but significant upregulation and downregulation of various tRNA types is still present. **C)** Compared to naïve Huh7 cells, tRNA-Gly-CCC, tRNA-

Pro-GGG and tRNA-Pro-UGG are concurrently elevated while tRNA-Arg-CCU and tRNA-Val-TAC are concurrently downregulated following full-length and subgenomic HEV infection although magnitude of effects is substantially different. Statistical significance of differences observed was assessed by a Mann-Whitney U test (two tailed). *P ≤ 0.05 was indicated as significant.

Table S1. Key materials used in this study.

Reagent or Resource	Source	Identifier
NEB-5-alpha Competent E. coli	NEB	Cat#: C2987I
HEV Gt3 (Kernow-C1 P6 Clone)	National Institute of Allergy and Infectious Diseases, National Institutes of Health, USA	GenBank: JQ679013
HEV p6-luc (Gaussia luciferase reporter gene)	Wang et al., 2014	N/A
Tris-HCl	Sigma-Aldrich	Cat#10812846001
EDTA	Sigma-Aldrich	Cat#1233508
MgCl2	Sigma-Aldrich	Cat# M8266
RNA isolation kit (NucleoSpin@ RNA)	MACHEREY-NAGEL	Cat#740955.250
T4 RNA ligase (dsRNA ligase)	NEB	Cat#M0239L
Prime Script RT Master Mix	Takara	Cat#RR036Q
SYBR™ Green PCR Master Mix	Applied Biosystems™	Cat#4344463
Human: HuH7 (♂) Naïve Cell Line	This paper	N/A
HEV infectious HuH-7 (♂) Cell Line	This paper	N/A
HEV replicon HuH-7 (♂) Cell Line	This paper	N/A
Primers for reverse transcription, see Table S1	This paper	N/A
Primers for qPCR, see Table S1	This paper	N/A
U-adaptor: 5phos/TCGTAGGGTCCGAGGTATTCACGATGrGrN	Life technology	N/A
tRNA-histidine adaptor: 5phos/TCGTAGGGTCCGAGGTATTCACGATGrGrU	Life technology	N/A
CodonW	John Peden et al.	http://codonw.sourceforge.net/
HEMI	Department of Biomedical Engineering, Huazhong University of Science and Technology	http://hemi.biocuckoo.org/
Python (Pandas, Seaborn, Matplot)	Guido van Rossum et al.	https://www.python.org/
2(-Delta Delta C(T)) Method	Livak et al.,2001	N/A
Deacylation buffer: 5×Tris-HCl (100 mM, PH 9.0)	This paper	N/A

Table S2. Primers used in mature tRNAome detection

	# Genes	Forward primers (5'-3')	Reverse primer (5'-3')
1.	tRNA-Leu-UAA	RTRYCCGCGTGGGTTCGA	CTTAAGTCCAACGCCTTAACCACTC
2.	tRNA-Leu-CAA	TTCTGGTCTCCRBATGGAGGC	GAGTCTGGCGCCTTAGACCA
3.	tRNA-Leu-AAG	CTCCAGTCTCTTCGGRGGC	AATCCAGCGCCTTAGACCGCTCGGCC
4.	tRNA-Leu-UAG	GCTCCAGTCWYTTCGRDGGCG	ATCCAGCGCCTTAGACCRCT
5.	tRNA-Leu-CAG	GGTCGCAGTCTCCCCTG	GAACGCAGCGCCTTAGACC
6.	tRNA-Ile-AAU	TAACGCCAAGGTCGYGGGT	TAGCACCACGCTCTRACCAACT
7.	tRNA-Ile-GAU	ATAACACCAAGGTCGCGGG	CAGCACCACGCTCTTACCAAC
8.	tRNA-Ile-UAU	YRATGCCGAGGTTGTGAGTTC	ATAAGTACCGCGCGCTAACC
9.	tRNA-iMet-CAU	ATAACCCAGAGGTCGATGGATCGAAAC	GGCCCAGCACGCTTCC
10.	tRNA-Met-CAU	ARACTGACGCGCTGCCHG	TAATCTGAARGTCSTGAGTTCRARCCTC
11.	tRNA-Stop-UUA	ATCAGAGGGTCCAGGGTTCAA	AGTCTGATGCTCTACCAACTGAACT
12.	tRNA-His-GUG	CAGCAACCTCGGTTCGAATC	CGCAGAGTACTAACCACTATACGATC
13.	tRNA-Phe-GAA	AGATCTAAAGGTCCCTGGTTCRATCC	CAGTCTAACGCTCTCCCAACTGA
14.	tRNA-Val-AAC	CACGCGAAAGGTCCCCG	TAGGCGAACGTGATAACCACTACAC
15.	tRNA-Val-UAC	CACGCAGAAGGTCCTGGG	AAAGCAGAYGTGATAACCACTACACTAT
16.	tRNA-Val-CAC	CACGCGAAAGGTCCCCG	TGAGGCGAACGTGATAACCACT
17.	tRNA-Ser-AGA	AGAAATCCATTGGGGTYTCCC	AGTCCATCGCCTTAACCACTCGGC
18.	tRNA-Ser-UGA	AATCCATTGGGGTYTCCCCG	AAGTCCAWCGCCTTAACCACTCG
19.	tRNA-Ser-CGA	AAATCCAATGGGGKYTCCCCG	GAGTCCAACRCCTTAACCACTC
20.	tRNA-Pro-AGG	GGTGCGAGAGGTCCCGG	TAAGCGAGAATCATACCCCTAGACCA
21.	tRNA-Pro-GGG	GGCTGCTGATCCCAGGC	CAGAGCGCACATTTCTAACCACTATG
22.	tRNA-Pro-UGG	GTGCGAGAGGTCCCGGG	AAAGCGAGAATCATACCCCTAGACC
23.	tRNA-Pro-CGG	GGGTGYGAGAGGTCCCGG	AAGCGAGAATCATACCCCTAGACC
24.	tRNA-Thr-AGU	TAAACAGGAGATCCTGGGTTCGAATC	CTAGACAGGCGCTTTAACCAGCTAA
25.	tRNA-Thr-UGU	TAAACCAGGGGTCGCGAGTT	CAAGACCAGYGCTCTAACCMCT
26.	tRNA-Thr-CGU	CGTAAACMGRAGATCVYGGGTTC	AGACMGRCGCYTTAACMAACTRRG
27.	tRNA-Ala-CGC	ATGTRYGAGGYCCCGGGTTCR	AAGCATGCGCTCTACCACTG
28.	tRNA-Ala-GGC	CAAACAGGAGATGCTGGATTTCAATCC	CAGACAAGTACTTTAACCCACAAAGCC
29.	tRNA-Ala-UGC	YGTRTGAGGYCYCGGGTTCR	AAGCATGCGCTCTACCACTG
30.	tRNA-Ala-AGC	CATGYAYGAGGYCCYGG	TAAGCAYGCGCTCYACCRCT
31.	tRNA-Tyr-AUA	AATCTAAAGACAGAGGTCAAGVYCT	ATAGTCTAATGCTTACTCAGCCATTTTACC
32.	tRNA-Tyr-GUA	CATCCTTAGGTCGCTGGTTCGA	GTCCTCCGCTCTACCARCTGA
33.	tRNA-Stop-CUA	CATCCTTAGGTCGCTGGTTCGA	GTCCTCCGCTCTACCARCTGA
34.	tRNA-Gln-UUG	ATCCAGCRATCCGAGTTCRAAT	AGTCCAGAGTGCTAACCATTACACC
35.	tRNA-Gln-CUG	ATCCAGCGATCCGAGTTCRART	AGTCCAGAGTGCTHACCATTACACC
36.	tRNA-Asn-AUU	AACCGAACGGTGAGTAGTTCAAGA	TAGCCGAACGCTCTGACCG
37.	tRNA-Asn-GUU	ACYGAAAGATTRGTGGTKCRAG	ACAGYCRAAYGCGCTAAC
38.	tRNA-Lys-UUU	TTAATCTGAGGGTCCRGGGTTC	AGTCTGATGCTCTACCRACTGAG
39.	tRNA-Lys-CUU	TAATCYCAGGGTCGTGGGTTCG	GAGTCYCATGCTCTACCGACTGAG
40.	tRNA-Asp-AUC	CACGAGGTCTTGGGCTGATTC	TGATAAGTACACTCTCTACCACTGAGCT
41.	tRNA-Asp-GUC	TCACGCGGGAGACCGG	AGGCGGGGATACTCACCACTA
42.	tRNA-Glu-UUC	TTCACCSMSGYGGCCCG	ARCSMSGAATCCTARCCRCTAGAC
43.	tRNA-Glu-CUC	GATTCGGCGCTCTCACCG	CTAACCACTAGACCACCAGGGA
44.	tRNA-Cys-ACA	CTTTAAAGTCATATGTAGCTGGGTTCAA	GTTACATAGCTTATAGAGTTGCTTTTGA
45.	tRNA-Cys-GCA	GATCAAGAGGTCCCYGGTTCA	AGTCAAATGCTCTACCMCTGAGC
46.	tRNA-Trp-CCA	GATCAGAAGGTTGCGTGTTCAAATC	GGAGTCAGACGCGCTACC
47.	tRNA-Arg-ACG	GATCAGAAGATTCYAGGTTCGACTCC	AGTCAGACGCGTTATCCATTGC
48.	tRNA-Arg-UCG	GATCAGAAGATTGMRGGTTCGARTC	AGTCAGACGCCTTATCCATTAGGC
49.	tRNA-Arg-CCG	GAKCWGRRGATTGWGGGTTCGAGTCC	ARKCWGAYGCCTTATCCATTAGGC

50.	tRNA-Ser-ACU	TAATGCCAGGGTCGAGGTTTCG	AGCAGCACGCTCTAACCAAC
51.	tRNA-Ser-GCU	AATCCATTGTGCTCTGCACGC	AGTCCATCGCCTTAACCACTCG
52.	tRNA-Arg-UCU	AATYCARAGGTTCYGGGTTCG	AAGTCCARYGCGCTCRTC
53.	tRNA-Arg-CCU	TAAGCCAGGGATTGTGGGTTC	AGGCCARTGCCTTATCCATTAGG
54.	tRNA-Gly-GCC	ACGCRGGAGGCCCRGGT	CAGGCRAGAATTCTACCACTGAACC
55.	tRNA-Gly-UCC	CAAGCAGTTGACCCGGGTTC	AAGGCAGCTATGCTHACCACTATACC
56.	tRNA-Gly-CCC	CKKGMGMCCCGGGTTCRA	ARKCKWGMATKMTACCACTRMACC
57.	tRNA-seCys-UCA	AAACCTGTAGCTGTCTAGYGACAGA	AAGCCTGCACCCCAGACC
U-adaptor		5phos/TCGTAGGGTCCGAGGTATTCACGATGrGrN	

Table S3. Preparation of qPCR mix.

Component	Volume	Final amount
2×SYBR^{TM} Green PCR Master Mix	5 µL	
tRNA-X-R/F	3 µL	0.25 pMol *per* primer
cDNA	2 µL	

Table S4. Specifics of the qPCR analysis.

Cycle	Denaturation	Annealing	Extension	Melt Curve
1	95 ℃, 10 min			
2-51	95 ℃, 10 s	58 ℃, 30 s	72 ℃, 40 s	
52	95 ℃, 1 min			
53				65 ℃, 1 min
≥54				+0.3 ℃/cycle, 15 s

Table S5. Interferon stimulated genes (ISGs) used in this study.

Genes	GenBank #
CH25H	NM_003956
IFITM	NM_003641.4
TRIM	NM_016388.3
OAS/RNaseL	NM_022137.5
Viperin	AF442151.1
Tetherin	NM_003021.4
STAT1	NM_007315.3
STAT2	NM_005419.4
RIG-I	NM_014314.4
cGAS	NM_138441.3
PKR	NM_001135651.3
IRF1	NM_002198.3
IRF3	NM_001571.6
IRF7	NM_001572.5
IRF9	NM_006084.4
SOCS	NM_145071.2
USP18	NM_017414.3

Chapter 4

Errors in translational decoding:

tRNA wobbling or misincorporation?

Xumin Ou, Jingyu Cao, Anchun Cheng, Maikel P Peppelenbosch, Qiuwei Pan

Abstract

As the central dogma of molecular biology, genetic information flows from DNA through transcription into RNA followed by translation of the message into protein by transfer RNAs (tRNAs). However, mRNA translation is not always perfect and errors in the amino acid composition may occur. Mistranslation is generally well-tolerated, but once reaching super-physiological levels, it can give rise to a plethora of diseases. The key causes of mistranslation are errors in translational decoding of the codons in mRNA. Such errors mainly derive from tRNA misdecoding and misacylation, especially when certain codon-paired tRNA species are missing. Substantial progress has recently been made with respect to the mechanistic basis of erroneous mRNA decoding as well as the resulting consequences for physiology and pathology. Here we aim to review this progress with emphasis on viral evolution and cancer development.

Keywords: tRNA, translational decoding, viral evolution, cancer development

Introduction

In all living organisms, DNA is transcribed into RNA and RNA is translated into protein. The latter process is executed by the ribosome, which constitutes the translation machinery that produces the cellular proteome by decoding mRNAs. Deciphering mRNA codons by tRNAs in the ribosome involves Watson-Crick base pairing (1). However, the translation machinery is not always perfect and errors in the amino acid composition may occur (2-5). The general error rates of genomic replication ($\approx 10^{-8}$) are estimated to be approximately ten thousand-fold lower than those of protein synthesis ($\approx 10^{-4}$), and thus in most instances mRNA translation is the key process contributing to inaccuracy of the cellular proteome (6). The discrepancy between error rates in DNA replication and mRNA translation may partially relate to the fact that DNA replication occurs at the level of individual nucleotides (involving $4^1=4$ possible permutations), whereas the translation machinery interprets mRNA codons in triplets (involving $4^3=64$ possible permutations) (7).

In the canonical interpretation, 61 aminoacyl-tRNAs and 3 suppress tRNAs decode 64 triplet codons that specify twenty amino acids (1). The resulting redundancies in the genetic code attribute to synonymous codons, which involve wobbling at position 3. For each amino acid, the number of codon usage varies from two to six according to codon degeneracy. In parallel, the numbers of certain amino acid specified tRNAs (based on recognition of anticodons) also vary from two to six box tRNA sets. Translational decoding of the mRNA codons is constrained by factors during codon-anticodon recognition and often constitutes the rate-limiting step during protein synthesis. Besides the abundance of tRNA species, mRNA translation is regulated by nearly a hundred of epigenetic tRNA modifications, especially at the wobble position (8, 9). The efficiency of mRNA decoding machinery is also essentially regulated by codon usages bias that is distinguished by over- or under-represented synonymous codons (10, 11). Accordingly, optimizing of tRNA wobble and codon usage in mRNA can substantially enhance translation efficiency and accuracy (10-12).

Nevertheless, mistranslation universally occurs. Pre- or post- mRNA translation may indirectly introduce errors of protein synthesis during transcription and post-translational processing (13). However, the translation machinery can directly contribute to mistranslation by tRNA misdecoding (leading to misincorporation or stop codon readthrough), tRNA misacylation (leading to wrong tRNA-amino acid coupling), codon reassignment or ribosomal translocation-provoked frameshifts (**Fig 1**) (13). It is becoming increasingly clear that such mistranslation has consequences on the pathophysiology of a variety of diseases (**Fig 1**) (**Table 1**), including multiple sclerosis, neurodegeneration, mitochondrial myopathy, encephalopathy, lactic acidosis, stroke-like episodes, Parkinson's disease and cancer (14-19). In this review, we aim to describe the key mechanisms that underlie mistranslation and illustrate potential implications using viral evolution and carcinogenesis as examples.
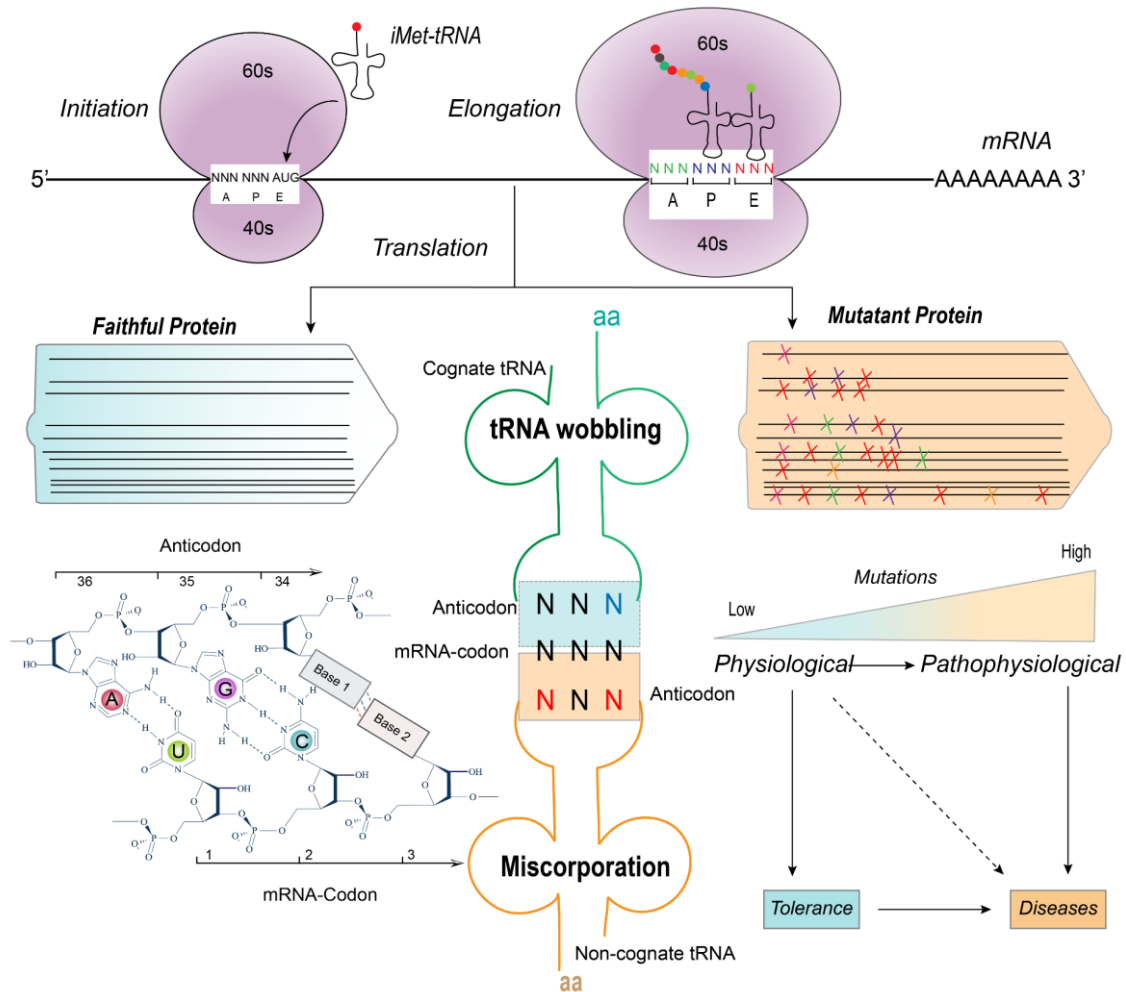
Fig 1. tRNA decoder regulates error ratio in translation decoding. In physiological conditions, errors in mRNA translation may occur but are generally well tolerated. However, the frequency of errors is dramatically increased in response to stresses. When amino acid misincorporation reaches intolerable levels, this contributes to dysfunction of cellular physiology and may cause pathogenesis. In general, the error ratio in translation decoding primarily depends on tRNA wobbling (cognate) and misdecoding (noncognate) as well as misacylation of tRNAs. aa, amino acid; iMet-tRNA, initiator tRNA Methionine; tRNA, transfer RNA.
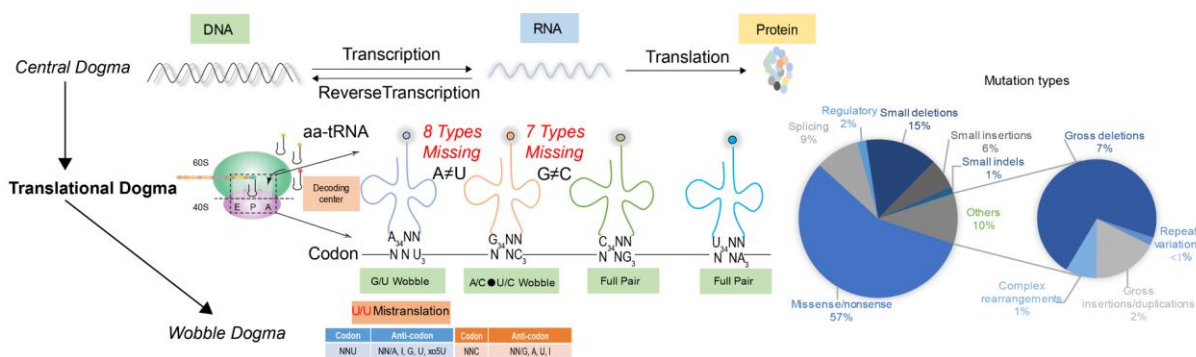
Fig 2. tRNA wobbling increases the risk of mistranslation. In the central dogma, DNA transcribes RNA and RNA translates protein. In the human mutation database, the major (57%)

mutation types are missense/nonsense (the right panel) that reflect the consequence of DNA errors at genomic level. Ribosome, as the translation machinery, essentially transduces genetic code to functional protein performed by aa-tRNAs. In human genome, 15 out of 64 tRNA types are actually missing partially because of low confidence (score < 50), including eight tRNAANN and seven tRNAGNN. Because of these missing tRNAs and the expanding wobble rules, the mRNA codon can be decoded by cognate or noncognate tRNAs, leading to modulation of translation efficiency and misincorporation (the left panel). At the bottom, the revised wobble rules and the consequent wobble types are listed according to the wobble position 3 of triplet codon. As for eight tRNAANN, NNU codons will be decoded by NNG or NNU anticodon of tRNAs. As for seven tRNAGNN, NNC codons will be decoded by NNA(I) or NNU anticodon of tRNAs. For specific missing tRNAs, the consequent wobble (tRNA wobbling or misdecoding) are detailed in **Fig 3**. aa-tRNA, aminoacyl-tRNA; tRNA, transfer RNA.

tRNA wobbling compensates for missing tRNA species

In the ribosome, tRNAs detect appropriate mRNA codons using the anticodon loop and transfer proper amino acids to polypeptides. However, the number of obligatory tRNA species (based on anticodons) for mRNA translation is substantially smaller than the theoretically required 64 species necessary for fully codon matching (1). Life solves this problem by allowing wobbling or superwobbling (also known as the "four-way wobbling") thus allowing less tRNA species to translate all mRNA codons (Table 1) (20-22). In the human genome, there are approximately tenfold excess of tRNA gene copies as compared to the number of possible codons (613 vs. 64) (23, 24). Nevertheless, the recently released GtRNAdb 2.0 database indicates that 15 out of theoretically necessary 64 tRNA species are missing, partially due to low confidence (scores < 50), including eight tRNA^{A34NN} and seven tRNA^{G34NN} (**Fig 2**) (23).

How to decode these codons without fully paired tRNAs remains an intriguing question. Because of wobbling and superwobbling, it is possible to use 32 tRNA species for decoding all 64 possible codons (1, 22). In plastid genomes, even 25 tRNA species suffice protein biosynthesis by "four-way wobbling" (21). tRNA species with an unmodified U at wobble site can decode all four triplets (NN/A, G, C and U). This relaxed wobble has been identified in Mycoplasma spp. and particular organelles, including mitochondria and, as mentioned, in plastids (21, 25, 26). Therefore, to decode those unpaired codons, cognate or non-cognate tRNAs are forced to wobble at position 3 of the codons by wobbling or superwobbling (**Fig 3**).

Table 1 The types and outcome of errors in translation machinery.

Organism	Error rates	Outcome	Cause	Mistranslation	tRNAs	Reference
E. coli	10%	Tolerance	Misacylation	Cys→ Pro	tRNA^{Pro}	(41)
				Ser→ Thr	tRNA^{Thr}	
				Glu→ Gln	tRNA^{Gln}	
				Asp→ Asn	tRNA^{Asn}	
Drosophila	~10-60%	Cell death	Misacylation	Tyr→ Phe	tRNA^{Phe}	(87)
Yeast	~6%	Stress non-sensitive	Misacylation	Pro→ Ala	tRNA^{Pro}(U/A)GG	(88)
HeLa cells	~5%	Alleviate oxidative stress	Misacylation	Glu→ Met	tRNA^{Glu}	(45)
CHO cells	~0.7%	Without changing in cellular viability	Misacylation	Tyr→ Phe	tRNA^{Tyr}	(89)
Mouse	~40-50%	Neurodegeneration	Misacylation	Gly→ Ala Ser→ Ala	tRNA^{Ala}	(15)
Ciliates	?	Genetic code evolution	Misreading	Gln→ UAA/UAG Trp→ UGA	tRNA^{glu} tRNA^{Trp}	(34)
Yeast	~45.5-54 % ① ~0.5%② ~7-86%③	Stop codon reassignments	Misreading	Gln/Tyr→ UAA(Stop)① Lys→ UAG(Stop)② Arg/Cys/Trp→ UGA(Stop)③	tRNA$^{Glu/Tyr}$ tRNA^{Lys} tRNA$^{Trp/Arg/Cys}$	(2)
Mice	~200-400%	Tumor growth	Misreading	Ser→ Ala	tRNA^{Ser}	(18)
Mycobacterial	~0.2%① ~0.8%②	Rifampicin resistance	Misincorporation	Gln→ Glu① Asn→ Asp②	tRNA^{Glu}① tRNA^{Asp}②	(43)
Human	?	Multiple sclerosis	Misincorporation	azetidine-2-carboxylic acid (Aze) → Pro	Likely tRNA^{pro}	(14)
Human	?	Mitochondrial disease	Wobble modification	Leu(UUG) reduced translation	tRNA^{leu}	(16)
Plant	?	Antibiotic sensitivity	Wobble and superwobbling	?	All codons with pyrimidines at wobble sites	(21)
Human	~20-80%	Proper translation	Modified wobble	Met→ Ile (AUA)	tRNA^{Met} $_{f5CAU}$	(90)
Yeast	~97.2%	Spontaneous	Codon reassignment	Ala→Leu (CUG)	tRNA^{Ala}-CAG	(91)
Human	?	Type 1 diabetes	Frameshift	?	?	(92)

Table — Two box tRNA sets:

Isotype	Two box tRNA sets			
Phe	AAA	GAA		
	None	10		
Asn	ATT	GTT		
	None	20		
Lys			CTT	TTT
			15	12
Asp	ATC	GTC		
	None	13		
Glu			CTC	TTC
			8	7
His	ATG	GTG		
	None	10		
Gln			CTG	TTG
			13	6

Table — Two box and other tRNA sets:

Isotype	Two box and other tRNA sets			
Ile	AAT	GAT		TAT
	14	3		5
iMet/Met			CAT	
			9/10	
Tyr	ATA	GTA		
	None	13		
Supres			TCA	CTA	TTA
Cys	ACA	GCA		
	None	29		
Trp			CCA	
			7	
SelCys				TCA
				2

Table — Four box tRNA sets:

Isotype	Four box tRNA sets			
Ala	AGC	GGC	CGC	TGC
	22	None	4	8
Gly	ACC	GCC	CCC	TCC
	None	14	5	9
Pro	AGG	GGG	CGG	TGG
	9	None	4	7
Thr	AGT	GGT	CGT	TGT
	9	None	5	6
Val	AAC	GAC	CAC	TAC
	9	None	11	5

Table — Six box tRNA Sets:

Isotype	Six box tRNA Sets					
Ser	AGA	GGA	CGA	TGA	ACT	GCT
	9	None	4	4	None	8
Arg	ACG	GCG	CCG	TCG	CCT	TCT
	7	None	4	6	5	6
Leu	AAG	GAG	CAG	TAG	CAA	TAA
	9	None	9	3	6	4

Fig 3. Errors in translation decoding are regulated by tRNA wobbling at all three codon positions. Sixty-four tRNA sets are summarized and specified in parallel with codon degeneracy (left bottom and central). Faithful or misincorporated protein can result from decoding by cognate or near-cognate tRNA at position 3. For eight missing tRNA^{ANN} (yellow text), NNU•tRNA^{UNN} wobble-dependent misdecoding by near-cognate tRNAs mainly occurs at the two box (I–VI) and six box tRNA sets (arginine-serine misincorporation) (VII). For seven missing tRNA^{GNN} (red text), NNC codons will be decoded by cognate tRNAs without amino acid misincorporation (VIII) because they happen at the four and six box tRNA sets. Since leucine and phenylalanine share UUN codon, leucine-phenylalanine misincorporation may occur across the six and two box tRNA sets. Besides wobbling at position 3, mRNA codons can be falsely decoded by "far-cognate" tRNA at position 1 and 2 (in the text). Missing tRNAs are indicated as question mark. Individual wobble and misdecoding are labeled as green and red text, respectively. tRNA, transfer RNA.

Decoding unpaired codons by excessive tRNA wobbling

Though tRNA wobbling enables translation compatibility, this also increases the probability of misdecoding by non-cognate tRNAs. Among the eight-missing species of tRNA^{ANN}, the NNU codons likely only pair with tRNA^{GNN}, tRNA^{UNN} and tRNA^{INN} as dictated by the revised wobble rules (**Table 2**) (**Fig 2**). Because the tRNA^{ANN} is missing and so does tRNA^{INN} (leading to wobbling with either adenine, cytosine or uridine) as conversion of tRNA^{ANN} to tRNA^{INN} is catalyzed by the tRNA dependent adenosine deaminases 2 (ADAT$_2$) (27). Specifically, if NNU codons pair

with tRNA^{GNN}, it will lead to a G•U wobble pair without concomitant amino acid misincorporation as the same amino acid is coded by NNU and NNC (tRNA^{GNN}) codons (**Fig 3**). If NNU codons pair with tRNA^{UNN}, however, the resultant U•U pair will cause amino acid misincorporation. NNU and NNA (tRNA^{UNN}) code for different amino acids at the two-fold degenerate codon box, resulting in Leu→Phe, Lys→Asn, Glu→Asp, Gln→His, Stop→Tyr and Selcys→Cys misdecoding (**Fig 3**). According to the original wobble hypothesis of Francis Crick, the codons decoded by the two box tRNA sets must distinguish either NNU/C or NNA/G (1). However, based on the revised wobble rules, NNU•tRNA^{UNN}-mediated decoding is at bay with Crick's assumption and might lead to misincorporation of amino acids. Leu→Phe, Lys→Asn and Gln→His misincorporations have been reported to occur in bacterial and mammalian cells when such cells suffer from Phe, Asn and His starvation, respectively (28-30). Misreading of codons by the "Two-out-of-three" hypothesis, which entails that the first two nucleotides in each codon are essential for anticodon recognition, has been suggested to pose a threat to translation fidelity (31). This type of misreading may occur in those two-fold degenerate codons as uniquely discriminated by wobble bases. It has been experimentally proven that tRNA superwobbling suffices to decode all four triplets of four-fold degenerate codons in plastids (21, 22). Interestingly, such superwobbling may allow the two-fold degenerate codons to cross-decode by NNU•tRNA^{UNN}-mediated decoding (**Fig 3**).

With respect to the seven-missing species of tRNA^{GNN}, the NNC codons are expected to pair with tRNA^{ANN}, tRNA^{UNN} and tRNA^{INN} (**Fig 2**) (**Table 2**). All missing tRNA^{GNN} can be decoded through wobble pairing without accompanying misincorporation of amino acids (**Fig 3**). Because they occur in four or six-fold degenerate tRNA boxes for which cognate tRNAs are available (**Fig 3**). However, misincorporation of amino acid may occur if NNC codons are misdecoded by non-cognate tRNA^{UNN} where the amount of tRNA^{GNN} is not limiting. Because NNC and NNA (tRNA^{UNN}) code for different amino acids in twice-degenerated codons (**Fig 3**). Furthermore, codon UA**U** (tyrosine)-UA**A** (Stop) mismatch will truncate the elongation process of nascent peptide (**Fig 3**). Conversely, if the stop codon (UGA) is mismatched by a Selcys-tRNA^{UCA} this will lead to an excessively translation elongation (**Fig 3**). Arginine and serine share an AG**N** wobble (AG**A** and AG**G** for arginine; AG**U** and AG**C** for serine) and this predisposes organisms to a potential arginine→serine misincorporation (**Fig 3**). Importantly, such arginine→serine misincorporation affects the quality of therapeutic antibody production by Chinese hamster ovary cells, illustrating the relevance of mRNA misdecoding (32). In conclusion, unpaired codons are likely to be misdecoded by non-cognate tRNAs due to excessive tRNA wobbling, raising questions as to the consequences of such misdecoding for living organisms (33).

tRNA wobbling at three codon positions compromises the fidelity of the translation decoder

Nonsense translation, so-called stop codon readthrough, can result from aberrant decoding of stop codons by non-cognate aminoacyl-tRNAs (examples are Gln(*C*AG/*C*AA), Tyr(UA*U*/UA*C*) and Lys(*A*AG/*A*AA) for the UAA and UAG stop codons respectively; Trp(UG*G*), Arg(*A*GA), and Cys(UG*U*/UG*C*) for the UGA stop codon) (2, 33). The occurrence of such readthrough highlights the possibility of position 3 and 1 wobbling in translational machinery and provides an indication as to how common translational misdecoding in living organisms is. In ciliates, ribosome profiling has demonstrated that all three stop codons can be misdecoded, while rates of such miscoding depend on the position within mRNA molecule (coding region or the end) (34). Position 1 wobbling not only occurs in stop codons, but also in sense codons, such as the misreading of Arg **C**GU/**C**GC codons as Cys **U**GU/**U**GC codons (35, 36). By using the prokaryote ortholog of elongation factor Tu (EF-Tu) for targeted mass spectrometry, it has been reported that even position 2 can be misdecoded by non-cognate tRNAs, as illustrated by the detection of the Arg C**G**U codon misdecoded by tRNA^{GAG}-Leu (5). Thus, substantial misdecoding at all three positions is possible (2, 5) (**Table 1**). This consequently compromises the fidelity of the translation decoder.

It has been reported that G•T mismatching occurs in both DNA and RNA duplex following tautomerization and ionization, and this plays important roles in replication and translation errors (37, 38). This Watson-Crick-like mismatch can evade fidelity checkpoints and appears to occur with probabilities (10^{-3} to 10^{-5}) that strongly imply a universal role of this mismatch in translation errors (38). The rG•rT mismatch at position 3 may not lead to mistranslation in decoding center, because NNU and NNC (rG•rC/rU) code for the same amino acids in such twice-degenerated codons and the same holds true for NNG and NNA (rU•rA/rG) (**Fig 3**). However, more mistranslation results if rG•rT mismatch takes place at position 1 and 2 (5, 35, 36). Hence, *in toto* a picture emerges that amino acid misincorporation in the nascent peptide chain is prone to occur mainly due to the absence of fully Watson-Crick pairing tRNAs and by excessive wobbling at all three codon positions (5).

Table 2 Revised wobble rules

Codon (XXN$_3$)	Anticodon (N$_{34}$XX)
A	U, A, I, xo^{5}U, xm^{5}s^{2}U, xm^{5}Um, Um, xm^{5}Um, k^{2}C
U	A, I, G, U, xo^{5}U
G	C, A, U, xo^{5}U, xm^{5}s^{2}U, xm^{5}Um, Um, xm^{5}Um, m^{5}C
C	G, A, U, I

Quality control of the translation machinery

Faithful translation of the mRNA codons into protein is essential for cellular physiology. The fidelity of the translation machinery firstly depends on the specific coupling of amino acids to their cognate tRNA species, which is catalyzed by aminoacyl-tRNA synthetases (aaRSs) (**Fig 4a-b**). aaRS is capable of discriminating its cognate substrates from structurally analogous tRNAs and amino acids (39). Subsequently, eukaryotic elongation factor (eEF)1A or prokaryotic EF-Tu delivers the aminoacyl-tRNA to the ribosome A site for elongation of nascent peptide chain after proper codon-anticodon recognition (40). Thus, aaRSs are cardinal in protecting protein synthesis against misacylation (39), but their specificity is not absolute. For instance, in *Escherichia coli*, four types of misacylated-tRNA, including Cys-tRNA[Pro], Ser-tRNA[Thr], Glu-tRNA[Gln], and Asp-tRNA[Asn] do not evoke a correctional reaction (41). In both mice and bacteria, serine is prone to be misacylated by alanyl-tRNA synthetases (AlaRSs) (42). In mycobacteria, an increase in the substitution of Glu→Gln and Asp→Asn by translational misincorporation has been linked to phenotypic resistance to rifampicin treatment (43). Thus, beneficial mistranslation in both prokaryotes and eukaryotes may exist and improve their survival or facilitate drug resistance (43-45). Apart from misdecoding, misacylation of amino acids to tRNA molecules is another important source of mistranslated proteins, despite the presence of mechanisms preventing such events.

How could tRNA wobbling guarantee faithful decoding by the codon-anticodon duplex? During elongation, eEF-1A or EF-Tu delivers amino acid-coupled tRNA to the ribosome A site (40). Subsequently, the ribosome re-checks the codon-anticodon duplex that involves the highly conserved G530, A1492 and A1493 of 16S RNA via stabilization of the first two Watson-Crick pairs of the duplex (31, 46). A correct confirmation of the codon-anticodon duplex will induce a conformational domain closure in the ribosome and result in the formation of the appropriate peptide bond and elongate the nascent protein (47). Analysis of X-ray structures suggests that the positions 1 and 2 of the A codon are obligatory Watson-Crick base pairs. In prokaryotes, when U●G and G●U wobbles at the first or second codon-anticodon position, the decoding center forces this pair to adopt the geometry close to that of a canonical C●G pair (40). Using nuclear magnetic resonance (NMR) relaxation dispersion, it has recently been revealed that dG●dT misincorporation during replication is likely mediated via tautomerization and ionization (37). As discussed, this Watson-Crick-like mismatches may further contribute to tRNA wobbling and consequently misdecoding (5). Although the hydrogen bond is the major force to form codon-anticodon pairs (1), the van der Waals forces, steric complementarity and shape acceptance may concurrently contribute to the codon-anticodon recognition essentially for quality control (3, 40).

mRNA mistranslation in physiology

The integrity of mRNA translation sustains essential cellular physiology in all domains of life. Low level of mistranslation, however, is well-tolerated and even contributes to stress

responses as it creates a degree of diversity in the proteome (also known as "statistical proteome") (4). Yeasts engineered to misincorporate serine at leucine CUG codon initially lose fitness, but quickly adapt by promoting the evolution of genome architecture (48). Experiments employing misacylated aa-tRNAs show that up to 10% of overall mistranslation in *Escherichia co*li does not compromise physiology of this organism and is even compatible with bacterial proliferation (41). aaRSs of mycoplasma with mutations in the editing domain provoke misacylation tRNAs with highly similar amino acids that contribute to antigen diversity as to escape host immune defenses (49). In mammalian cells, up to 10-fold methionyl-misacylation to non-Met-tRNAs will protect against reactive oxygen species (ROS)-mediated damage when cells undergo oxidative stress, such as exposure to viral infections, Toll-like receptor ligands or xenobiotics (45).

Rates of mistranslation vary dramatically between organisms and different environmental conditions (**Table 1**). An overall amino acid misincorporation rates of ≈ 3-5 ‰ during translation is regarded as compatible with normal physiology (50, 51). In contrast, exceeding 1% misincorporation is usually deleterious and may provoke pathogenesis (43). For example, the 50% tRNA^{Ala} mischarging with Ser residues by an editing-defective alanyl tRNA synthetase (AlaRS) is associated with neurodegeneration (15). In addition, defective AlaRS is also related to cardioproteinopathy (52). However, the capacity of organisms to deal with mistranslation appears diverse and sub-physiological mistranslation is tolerant and even beneficial.

Errors of translation and viral evolution

Viral genomes are dynamically mutated with frequent emergence of new quasispecies. The spectrum for the hypermutation of viral genomes, sometimes denominate as mutant clouds (53). Mutation rates at genomic level (substitutions per nucleotide per cell infection, s/n/c) range from 10^{-8} to 10^{-6} s/n/c for DNA viruses and from 10^{-6} to 10^{-4} s/n/c for RNA viruses (54). Apparently, there is an error threshold to constrain viral evolution dependent on the genome size and permutations of errors (55). Within the virome, RNA viruses in particular mutate tremendously as a consequence of RNA-dependent RNA polymerases (RdRPs) being error-prone. For instance, the mutation rate of RdRPs that mediate poliovirus and Foot-and-Mouth Disease Virus (FMDV) replication can further expand or reduce the quasispecies diversity by regulation of replication fidelity (56, 57). The consequences of the mutations highly depend on both the position and properties of the affected amino acid residues. To take FMDV as an example, a W237F mutation but not a W237I mutation in the polymerase leads to a high fidelity and thus contributes to the subsequent mutation rates (57).

Little is known of the consequences of an error-prone translation machinery on viral evolution. As discussed, erroneous protein synthesis is prone to occur especially when cells suffer cellular stresses like viral infection. In this situation, the viral RNA is possibly mistranslated during the inaccurate translation (6). Several types of errors in the translational machinery have been

linked to viral adaptability. An example is the apparent selective pressure exerted on fungal mitovirus to exclude UGA (Trp) codons from its coding sequence because of the lack of fidelity of decoding this codon by the host mitochondrion (58). Interestingly, when organisms are recoded to obtain non-assigned codons, compensatory mechanisms emerge including frameshifts and stop codon readthrough (59). In yeast, mistranslation has been demonstrated to provoke evolution of genomic architecture (48). Hence, genomic mutagenesis is substantially associated with mistranslation that promotes the likelihood of evolution (4). It is well possible that viruses also utilize mistranslated genome-copying machinery (e.g. RdRP) for viral evolution. Besides the classical mutations inherited from error-prone replication at genomic level, we propose that mistranslation may generate additional RdRP mutants at the protein level that are not inheritable. Except for negative-strand RNA viruses, there is no RdRP incorporated in the virion. Therefore, upon infection, translation of the viral genome is invariably ahead of its replication. Thus, when an RNA virus releases its genome into host cell after uncoating, errors in RdRP may be accidently introduced by mistranslation (58, 59). The resulting mixture of wild-type and mutated RdRP enzymes initiate replication associated with a spectrum of viral quasispecies (**Fig 4c**). Those species that possess the best viral fitness finally survive and become dominant.

As for DNA viruses, the mechanisms driving viral mutation are more diverse and less well-understood. Degradation of HIV-1 proviral DNA with G→A hypermutation has an important role in host responses to infection (60). This sublethal mutagenesis catalyzed by cytidine deaminases in family of apolipoprotein B RNA-editing catalytic polypeptide-like 3 (APOBEC3) can induce drug-resistant and generate immune-escape viruses (60, 61). In hepatitis B virus infected patients, however, such mutations may have undesired consequences with respect to the viral reverse transcriptase (*e.g.* the A181T and M204I mutations) and mediate adefovir resistance (62). Analogously, it has been reported that mutations in palm, finger and 3'-5' exonuclease domains of herpesviruses DNA polymerase are introduced as a consequence of nucleoside analogue-based therapy (63). Conceivably, mistranslation in DNA viruses can also generate viral proteins that are more prone to provoke mutations in the viral genome, but hard data for this notion is currently still lacking.
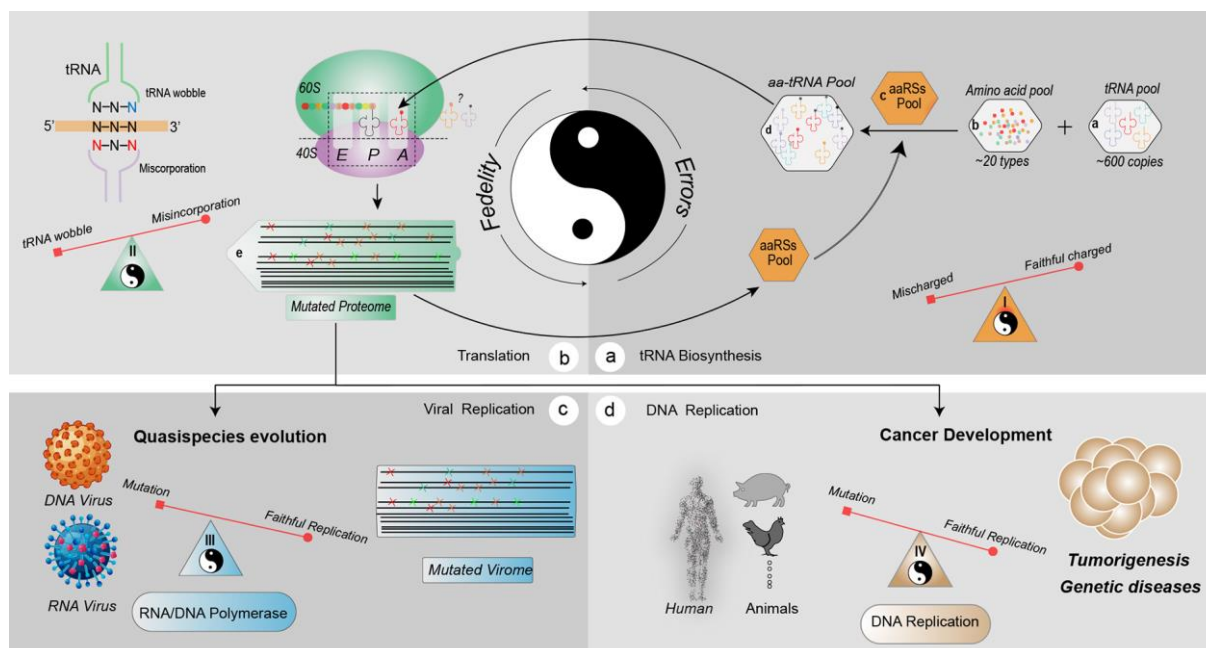
Fig 4. Fidelity and errors of translation decoding and the implications in viral evolution and cancer development. (a) aa-tRNAs are synthesized by sampling from the amino acid pool and tRNA pool and require catalysis by aaRSs. This process may accidently introduce misacylated aa-tRNAs, because the types of tRNAs and amino acids are difficult to be distinguished by involved aminoacyl synthetase because of analogous structures. (b) During elongation, tRNA wobbling will increase translation efficiency. Misincorporation can also be introduced because of tRNA misdecoding (amino acid misincorporation caused by excessive wobble decoding), especially when certain codon-paired tRNA species are missing. Finally, the fidelity of translation machinery will be impaired and produce mutated proteome, including RNA and DNA polymerases, aaRSs, and accessories. (c) Mistranslation of RdRP in RNA viruses will augment generation of a mutated virome (quasispecies) and facilitate viral evolution and adaption. (d) Similarly, mistranslation of cellular DNA replication-related enzymes and relative proteins amplifies mutagenesis in the genome and contributes to cancer development. aaRS, aminoacyl-tRNA synthetase; aa-tRNA, aminoacyl-tRNA; RdRP, RNA-dependent RNA polymerase; tRNA, transfer RNA.

Errors of translation in cancer development

Malignant transformation is usually associated with accumulation of large numbers of DNA mutations. Once occurring in essential oncogenes and tumor suppressors, these are also intimately associated with cancer development and progression (64, 65). The importance of DNA mutation-dependent alteration in protein composition is illustrated by the recent identification of ≈ 3400 driver mutations in tumor exomes (66). In the human mutation database, 57% of the mutations are missense/nonsense (**Fig 2**). This reflects the major consequences of DNA errors that are driven by either DNA replication errors or environmental factors. Apart from genomic alterations, mistranslation may also be important in cancer cells. It has been reported that DNA replication errors are responsible for two-third of the mutations observed in seventeen cancer types (67). Hence reduced fidelity of DNA replicating enzymes appears more important than environmental factors for generating cancer-associated

mutations. The implication of this notion is that if mistranslation of DNA replicating enzymes reduces replication fidelity, this would be expected to further advance cancer development (68). Of note, translation machinery is largely rewired during tumorigenesis (69). By shaping tRNA pool to match pro-tumorigenic mRNAs, the translation of oncogenes is facilitated to prime oncogenesis, such as highly upregulated tRNA[Glu]-UUC and tRNA[Arg]-CCG in breast cancer (70, 71). Moreover, mutated components of ribosome are involved in carcinogenesis as well and may foster disease by compromising the ribosome (translation fidelity) to "translate" cancer (69). For example, missense mutations of the ribosomal protein RPS15, a component of the 40S ribosomal subunit, is involved in chronic lymphocytic leukemia (72). How the compromised translation machinery contributes to the nature of hypermutated tumor transformation at genomic level is an intriguing question. In analogy to RNA viruses, mistranslation of DNA polymerases and APOBEC3H in cancer may occur before genomic replication (68, 73-75). The cellular proteome in G1 phase of cell cycle must duplicate before S phase and the demand on the translational machinery may provoke errors with respect to mRNA decoding (76). The ribosomal fidelity in (pre-) malignant cells may become compromised, resulting in mistranslated DNA polymerase molecules which in turn drive further genomic instability (69, 75). This further contributes to hypermutation and consequently tumorigenesis (Fig 4d) (70, 77). In apparent support of this notion, mistranslation caused by Ser-to-Ala misreading tRNA has been shown to promote the development of epithelial cancer in mouse models (18). Moreover, mutated DNA polymerase ε (P286R) in mice model provokes ultra-mutagenesis that can rapidly develop into lethal cancers of diverse lineages (75).

It is important to note that N→T missense mutations are widespread in cancer (66). This type of mutation increases translation efficiency through facilitating tRNA wobbling and superwobbling that provides the cancer cells with advantage to compete clones but will concomitantly provoke amino acids misincorporation, especially when the two box tRNA sets are involved (**Fig 3**). As described, epigenetic modification of tRNA (U34) further supports tumorigenesis by upregulating U34 enzymes and enhancing codon wobble of especially tumor promoting genes, an effect which prominently involves SOX9 and Elp3 (78, 79). High level of the U34 enzyme promotes alternative translation and has been linked to resistance to anti-BRAF therapy through wobble decoding of HIF1A mRNA in a codon-specific manner (80). Thus, the error-prone translation machinery appears to contribute to mutagenesis during cancer development.

Though mistranslation promotes carcinogenesis, it also offers possible target for anti-cancer therapeutics. Targeting enzymes catalyzing U34 tRNA modification has been demonstrated the potential for treating melanoma (81). Depletion of the U34-enzymes Elp3 or CTU1/2 provokes cell death in patient-derived BRAF[V600E] melanoma cultures (80). Genetic incorporation of noncanonical amino acids by decoding specific codon is another approach (82). Misincorporations of p-acetylphenylalanine at target codons have been explored to develop bi-specific antibody-based therapy for breast cancer and acute myeloid leukemia (83,

84). Moreover, certain mutant peptides of human tumors can serve as T-cell epitopes for immunotherapy (85). These tumor-specific immunogens as potentially personalized vaccines have been shown to boost immune rejection to the tumors in mouse model (85, 86).

Conclusion and perspective

mRNA mistranslation universally occurs across all living organisms. It is generally well-tolerated in physiology and even helps the organism adapt and withstand cell stresses. However, excessive mistranslation is pathogenic and implicated in many diseases. Mistranslation may also provide targets for drug and vaccine development, in particular against viral infection and cancer.

Although mRNA mistranslation can be caused by a variety of mechanisms, tRNA misdecoding and tRNA misacylation are the key drivers. The former is largely attributed to the partially missing tRNAs and excessive wobbling decoding. Consequently, mRNA codons can be coupled to cognate or near-cognate tRNAs at position 3, leading to modulation of translation efficiency and misincorporation (33). By furthering wobbling at position 1 and 2, mRNA codon can be falsely decoded by "far-cognate" tRNAs. We speculate that, if wobbling or superwobbling concurrently occurs at all three positions, especially with regard to the codons decoded by the two box tRNA sets, no functional protein would likely be produced.

The development of high-throughput sequencing and ribosome profiling technologies have greatly advanced our understanding of tRNA decoder (71). However, proteomic analysis at single molecular level remains technically infeasible. This hampers a detailed characterization of the protein "quasispecies" pool that results from mistranslation. In the future, deciphering single codon-anticodon decoding will help providing more mechanistic insights as to how tRNA decoding relates to translation fidelity.

References

1. Crick FH. Codon--anticodon pairing: the wobble hypothesis. Journal of molecular biology. 1966;19(2):548-55.
2. Roy B, Leszyk JD, Mangus DA, Jacobson A. Nonsense suppression by near-cognate tRNAs employs alternative base pairing at codon positions 1 and 3. Proceedings of the National Academy of Sciences. 2015;112(10):3038-43.
3. Rozov A, Demeshkina N, Westhof E, Yusupov M, Yusupova G. New Structural Insights into Translational Miscoding. Trends in biochemical sciences. 2016;41(9):798-814.
4. Ribas de Pouplana L, Santos MA, Zhu JH, Farabaugh PJ, Javid B. Protein mistranslation: friend or foe? Trends in biochemical sciences. 2014;39(8):355-62.
5. Garofalo R, Wohlgemuth I, Pearson M, Lenz C, Urlaub H, Rodnina MV. Broad range of missense error frequencies in cellular proteins. Nucleic Acids Res. 2019.
6. Mohler K, Ibba M. Translational fidelity and mistranslation in the cellular response to stress. Nat Microbiol. 2017;2:17117.
7. Robinson R. Which codon synonym is best? It may depend on what's on the menu. PLoS Biol. 2014;12(12):e1002014.
8. Chan CT, Pang YL, Deng W, Babu IR, Dyavaiah M, Begley TJ, et al. Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. Nature communications. 2012;3:937.
9. Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, et al. The RNA Modification Database, RNAMDB: 2011 update. Nucleic Acids Res. 2011;39(Database issue):D195-201.
10. Hanson G, Coller J. Codon optimality, bias and usage in translation and mRNA decay. Nature Reviews Molecular Cell Biology. 2017;19:20.
11. Quax TE, Claassens NJ, Soll D, van der Oost J. Codon Bias as a Means to Fine-Tune Gene Expression. Molecular cell. 2015;59(2):149-61.
12. Endres L, Dedon PC, Begley TJ. Codon-biased translation can be regulated by wobble-base tRNA modification systems during cellular stress responses. RNA biology. 2015;12(6):603-14.
13. Drummond DA, Wilke CO. The evolutionary consequences of erroneous protein synthesis. Nature reviews Genetics. 2009;10(10):715-24.
14. Rubenstein E. Misincorporation of the proline analog azetidine-2-carboxylic acid in the pathogenesis of multiple sclerosis: a hypothesis. Journal of Neuropathology & Experimental Neurology. 2008;67(11):1035-40.
15. Lee JW, Beebe K, Nangle LA, Jang J, Longo-Guess CM, Cook SA, et al. Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. Nature. 2006;443(7107):50-5.
16. Kirino Y, Yasukawa T, Ohta S, Akira S, Ishihara K, Watanabe K, et al. Codon-specific translational defect caused by a wobble modification deficiency in mutant tRNA from a human mitochondrial disease. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(42):15070-5.
17. Rodgers KJ. Non-protein amino acids and neurodegeneration: the enemy within. Exp Neurol. 2014;253:192-6.
18. Santos M, Pereira PM, Varanda AS, Carvalho J, Azevedo M, Mateus DD, et al. Codon misreading tRNAs promote tumor growth in mice. RNA biology. 2018:1-38.

19. Lant JT, Berg MD, Heinemann IU, Brandl CJ, O'Donoghue P. Pathways to disease from natural variations in human cytoplasmic tRNAs. The Journal of biological chemistry. 2019.

20. Agris PF, Eruysal ER, Narendran A, Vare VYP, Vangaveti S, Ranganathan SV. Celebrating wobble decoding: Half a century and still much is new. RNA biology. 2017:1-17.

21. Alkatib S, Scharff LB, Rogalski M, Fleischmann TT, Matthes A, Seeger S, et al. The contributions of wobbling and superwobbling to the reading of the genetic code. PLoS genetics. 2012;8(11):e1003076.

22. Rogalski M, Karcher D, Bock R. Superwobbling facilitates translation with reduced tRNA sets. Nature structural & molecular biology. 2008;15(2):192-8.

23. Chan PP, Lowe TM. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Research. 2016;44(D1):D184-D9.

24. Parisien M, Wang X, Pan T. Diversity of human tRNA genes from the 1000-genomes project. RNA biology. 2013;10(12):1853-67.

25. Inagaki Y, Kojima A, Bessho Y, Hori H, Ohama T, Osawa S. Translation of Synonymous Codons in Family Boxes byMycoplasma capricolumtRNAs with Unmodified Uridine or Adenosine at the First Anticodon Position. Journal of molecular biology. 1995;251(4):486-92.

26. Suzuki T, Nagao A, Suzuki T. Human Mitochondrial tRNAs: Biogenesis, Function, Structural Aspects, and Diseases. Annual Review of Genetics. 2011;45(1):299-329.

27. Novoa Eva Maria, Pavon-Eternod Mariana, Pan Tao, Lluís RdP. A Role for tRNA Modifications in Genome Structure and Codon Usage. Cell. 2012;149(1):202-13.

28. Precup J, Ulrich AK, Roopnarine O, Parker J. Context specific misreading of phenylalanine codons. Molecular and General Genetics MGG. 1989;218(3):397-401.

29. Parker J, Friesen JD. "Two out of three" codon reading leading to mistranslation in vivo. Molecular and General Genetics MGG. 1980;177(3):439-45.

30. Schwartz MH, Pan T. Function and origin of mistranslation in distinct cellular contexts. Critical reviews in biochemistry and molecular biology. 2017;52(2):205-19.

31. Lagerkvist U. "Two out of three": an alternative method for codon reading. Proceedings of the National Academy of Sciences of the United States of America. 1978;75(4):1759-62.

32. Ren D, Zhang J, Pritchett R, Liu H, Kyauk J, Luo J, et al. Detection and identification of a serine to arginine sequence variant in a therapeutic monoclonal antibody. Journal of Chromatography B. 2011;879(27):2877-84.

33. Blanchet S, Cornu D, Hatin I, Grosjean H, Bertin P, Namy O. Deciphering the reading of the genetic code by near-cognate tRNA. Proceedings of the National Academy of Sciences. 2018;115(12):3018-23.

34. Swart EC, Serra V, Petroni G, Nowacki M. Genetic Codes with No Dedicated Stop Codon: Context-Dependent Translation Termination. Cell. 2016;166(3):691-702.

35. Bouadloun F, Donner D, Kurland C. Codon‐specific missense errors in vivo. The EMBO journal. 1983;2(8):1351-6.

36. Edelmann P, Gallant J. Mistranslation in E. coli. Cell. 1977;10(1):131-7.

37. Kimsey IJ, Szymanski ES, Zahurancik WJ, Shakya A, Xue Y, Chu CC, et al. Dynamic basis for dG*dT misincorporation via tautomerization and ionization. Nature. 2018;554(7691):195-201.

38. Kimsey IJ, Petzold K, Sathyamoorthy B, Stein ZW, Al-Hashimi HM. Visualizing transient Watson–Crick-like mispairs in DNA and RNA duplexes. Nature. 2015;519:315.

39. Ling J, Reynolds N, Ibba M. Aminoacyl-tRNA synthesis and translational quality control. Annual review of microbiology. 2009;63:61-78.

40. Demeshkina N, Jenner L, Westhof E, Yusupov M, Yusupova G. A new understanding of the decoding principle on the ribosome. Nature. 2012;484(7393):256-9.

41. Ruan B, Palioura S, Sabina J, Marvin-Guy L, Kochhar S, LaRossa RA, et al. Quality control despite mistranslation caused by an ambiguous genetic code. Proceedings of the National Academy of Sciences. 2008;105(43):16502-7.

42. Guo M, Chong YE, Shapiro R, Beebe K, Yang X-L, Schimmel P. Paradox of mistranslation of serine for alanine caused by AlaRS recognition dilemma. Nature. 2009;462:808.

43. Javid B, Sorrentino F, Toosky M, Zheng W, Pinkham JT, Jain N, et al. Mycobacterial mistranslation is necessary and sufficient for rifampicin phenotypic resistance. Proceedings of the National Academy of Sciences. 2014;111(3):1132-7.

44. Miranda I, Silva-Dias A, Rocha R, Teixeira-Santos R, Coelho C, Gonçalves T, et al. Candida albicans CUG mistranslation is a mechanism to create cell surface variation. mBio. 2013;4(4):e00285-13.

45. Netzer N, Goodenbour JM, David A, Dittmar KA, Jones RB, Schneider JR, et al. Innate immune and chemically triggered oxidative stress modifies translational fidelity. Nature. 2009;462:522.

46. Jenner L, Demeshkina N, Yusupova G, Yusupov M. Structural rearrangements of the ribosome at the tRNA proofreading step. Nature structural & molecular biology. 2010;17(9):1072-8.

47. Ogle JM, Murphy IV FV, Tarry MJ, Ramakrishnan V. Selection of tRNA by the ribosome requires a transition from an open to a closed form. Cell. 2002;111(5):721-32.

48. Kalapis D, Bezerra AR, Farkas Z, Horvath P, Bodi Z, Daraba A, et al. Evolution of Robustness to Protein Mistranslation by Accelerated Protein Turnover. PLoS Biol. 2015;13(11):e1002291.

49. Li L, Boniecki MT, Jaffe JD, Imai BS, Yau PM, Luthey-Schulten ZA, et al. Naturally occurring aminoacyl-tRNA synthetases editing-domain mutations that cause mistranslation in Mycoplasma parasites. Proceedings of the National Academy of Sciences. 2011;108(23):9378-83.

50. Loftfield RB, Vanderjagt D. The frequency of errors in protein biosynthesis. The Biochemical journal. 1972;128(5):1353-6.

51. Yadavalli SS, Ibba M. Selection of tRNA charging quality control mechanisms that increase mistranslation of the genetic code. Nucleic Acids Research. 2013;41(2):1104-12.

52. Liu Y, Satz JS, Vo M-N, Nangle LA, Schimmel P, Ackerman SL. Deficiencies in tRNA synthetase editing activity cause cardioproteinopathy. Proceedings of the National Academy of Sciences. 2014;111(49):17570-5.

53. Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. Microbiology and molecular biology reviews : MMBR. 2012;76(2):159-216.

54. Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. J Virol. 2010;84(19):9733-48.

55. Holmes EC. Error thresholds and the constraints to RNA virus evolution. Trends in microbiology. 2003;11(12):543-6.

56. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature. 2006;439(7074):344-8.

57. Rai DK, Diaz-San Segundo F, Campagnola G, Keith A, Schafer EA, Kloc A, et al. Attenuation of Foot-and-Mouth Disease Virus by Engineered Viral Polymerase Fidelity. Journal of Virology. 2017;91(15):e00081-17.

58. Nibert ML. Mitovirus UGA(Trp) codon usage parallels that of host mitochondria. Virology. 2017;507:96-100.

59. Ma NJ, Hemez CF, Barber KW, Rinehart J, Isaacs FJ. Organisms with alternative genetic codes resolve unassigned codons via mistranslation and ribosomal rescue. eLife. 2018;7.

60. Stavrou S, Ross SR. APOBEC3 Proteins in Viral Immunity. Journal of immunology. 2015;195(10):4565-70.

61. Mohammadzadeh N, Follack TB, Love RP, Stewart K, Sanche S, Chelico L. Polymorphisms of the cytidine deaminase APOBEC3F have different HIV-1 restriction efficiencies. Virology. 2019;527:21-31.

62. Reuman EC, Margeridon-Thermet S, Caudill HB, Liu T, Borroto-Esoda K, Svarovskaia ES, et al. A classification model for G-to-A hypermutation in hepatitis B virus ultra-deep pyrosequencing reads. Bioinformatics. 2010;26(23):2929-32.

63. Topalis D, Gillemot S, Snoeck R, Andrei G. Distribution and effects of amino acid changes in drug-resistant alpha and beta herpesviruses DNA polymerase. Nucleic Acids Res. 2016;44(20):9530-54.

64. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61(5):759-67.

65. Roberts SA, Gordenin DA. Hypermutation in human cancer genomes: footprints and mechanisms. Nat Rev Cancer. 2014;14(12):786-800.

66. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive characterization of cancer driver genes and mutations. Cell. 2018;173(2):371-85. e18.

67. Tomasetti C, Li L, Vogelstein B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. Science. 2017;355(6331):1330-4.

68. Jansen AM, van Wezel T, van den Akker BE, Ventayol Garcia M, Ruano D, Tops CM, et al. Combined mismatch repair and POLE/POLD1 defects explain unresolved suspected Lynch syndrome cancers. Eur J Hum Genet. 2016;24(7):1089-92.

69. Sulima SO, Hofman IJF, De Keersmaecker K, Dinman JD. How Ribosomes Translate Cancer. Cancer Discov. 2017;7(10):1069-87.

70. Truitt ML, Ruggero D. New frontiers in translational control of the cancer genome. Nat Rev Cancer. 2016;16(5):288-304.

71. Goodarzi H, Nguyen HCB, Zhang S, Dill BD, Molina H, Tavazoie SF. Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. Cell. 2016;165(6):1416-27.

72. Ljungstrom V, Cortese D, Young E, Pandzic T, Mansouri L, Plevova K, et al. Whole-exome sequencing in relapsing chronic lymphocytic leukemia: clinical impact of recurrent RPS15 mutations. Blood. 2016;127(8):1007-16.

73. Starrett GJ, Luengas EM, McCann JL, Ebrahimi D, Temiz NA, Love RP, et al. The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. Nature communications. 2016;7.

74. Çağlayan M, Wilson SH. Pol μ dGTP mismatch insertion opposite T coupled with ligation reveals promutagenic DNA repair intermediate. Nature communications. 2018;9(1):4213.

75. Li HD, Cuevas I, Zhang M, Lu C, Alam MM, Fu YX, et al. Polymerase-mediated ultramutagenesis in mice produces diverse cancers with high mutational load. J Clin Invest. 2018;128(9):4179-91.

76. Hartwell LH, Kastan MB. Cell cycle control and cancer. Science. 1994;266(5192):1821-8.

77. Gaillard H, Garcia-Muse T, Aguilera A. Replication stress and cancer. Nat Rev Cancer. 2015;15(5):276-89.

78. Ladang A, Rapino F, Heukamp LC, Tharun L, Shostak K, Hermand D, et al. Elp3 drives Wnt-dependent tumor initiation and regeneration in the intestine. Journal of Experimental Medicine. 2015;212(12):2057-75.

79. Rapino F, Delaunay S, Zhou Z, Chariot A, Close P. tRNA Modification: Is Cancer Having a Wobble? Trends Cancer. 2017;3(4):249-52.

80. Rapino F, Delaunay S, Rambow F, Zhou Z, Tharun L, De Tullio P, et al. Codon-specific translation reprogramming promotes resistance to targeted therapy. Nature. 2018;558(7711):605-9.

81. Rapino F, Close P. Wobble uridine tRNA modification: a new vulnerability of refractory melanoma. Mol Cell Oncol. 2018;5(6):e1513725.

82. Kang M, Lu Y, Chen S, Tian F. Harnessing the power of an expanded genetic code toward next-generation biopharmaceuticals. Current Opinion in Chemical Biology. 2018;46:123-9.

83. Kim CH, Axup JY, Dubrovska A, Kazane SA, Hutchins BA, Wold ED, et al. Synthesis of bispecific antibodies using genetically encoded unnatural amino acids. Journal of the American Chemical Society. 2012;134(24):9918-21.

84. Lu H, Zhou Q, Deshmukh V, Phull H, Ma J, Tardif V, et al. Targeting human C‐type lectin‐like molecule‐1 (CLL1) with a bispecific antibody for immunotherapy of acute myeloid leukemia. Angewandte Chemie International Edition. 2014;53(37):9841-5.

85. Yadav M, Jhunjhunwala S, Phung QT, Lupardus P, Tanguay J, Bumbaca S, et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. Nature. 2014;515(7528):572-6.

86. Gubin MM, Zhang X, Schuster H, Caron E, Ward JP, Noguchi T, et al. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. Nature. 2014;515(7528):577-81.

87. Lu J, Bergert M, Walther A, Suter B. Double-sieving-defective aminoacyl-tRNA synthetase causes protein mistranslation and affects cellular physiology and development. Nature communications. 2014;5:5650.

88. Hoffman KS, Berg MD, Shilton BH, Brandl CJ, O'Donoghue P. Genetic selection for mistranslation rescues a defective co-chaperone in yeast. Nucleic Acids Res. 2017;45(6):3407-21.

89. Raina M, Moghal A, Kano A, Jerums M, Schnier PD, Luo S, et al. Reduced amino acid specificity of mammalian tyrosyl-tRNA synthetase is associated with elevated mistranslation of Tyr codons. The Journal of biological chemistry. 2014;289(25):17780-90.

90. Cantara WA, Murphy FV, Demirci H, Agris PF. Expanded use of sense codons is regulated by modified cytidines in tRNA. Proceedings of the National Academy of Sciences. 2013;110(27):10964-9.

91. Muhlhausen S, Findeisen P, Plessmann U, Urlaub H, Kollmar M. A novel nuclear genetic code alteration in yeasts and the evolution of codon reassignment in eukaryotes. Genome research. 2016;26(7):945-55.

92. Wei J, Yewdell JW. Autoimmune T cell recognition of alternative-reading-frame-encoded peptides. Nature medicine. 2017;23(4):409-10.

Chapter 5

The biological process of lysine-tRNA charging is therapeutically targetable in liver cancer

Ruyi Zhang, Lisanne Noordam, **Xumin Ou**, Buyun Ma, Yunlong Li, Pronay Das, Shaojun Shi, Jiaye Liu, Ling Wang, Pengfei Li, Monique MA Verstegen, D Srinivasa Reddy, Luc JW van Der Laan, Maikel P Peppelenbosch, Jaap Kwekkeboom, Ron Smits, Qiuwei Pan

Abstract

Background & Aims: Mature transfer RNAs (tRNA) charged with amino acids decode mRNA to synthesize proteins. Dysregulation of translational machineries has a fundamental impact on cancer biology. This study aims to map the tRNAome landscape in liver cancer patients and to explore potential therapeutic targets at the interface of charging amino acid with tRNA.

Methods: Resected tumour and paired tumour-free (TFL) tissues from hepatocellular carcinoma (HCC) patients (n = 69), and healthy liver tissues from organ transplant donors (n = 21), HCC cell lines, and cholangiocarcinoma (CC) patient-derived tumour organoids were used.

Results: The expression levels of different mature tRNAs were highly correlated and closely clustered within individual tissues, suggesting that different members of the tRNAome function cooperatively in protein translation. Interestingly, high expression of tRNA-Lys-CUU in HCC tumours was associated with more tumour recurrence (HR 1.1; P = .022) and worse patient survival (HR 1.1; P = .0037). The expression of Lysyl-tRNA Synthetase (KARS), the enzyme catalysing the charge of lysine to tRNALys-CUU, was significantly upregulated in HCC tumour tissues compared to tumourfree liver tissues. In HCC cell lines, lysine deprivation, KARS knockdown or treatment with the KARS inhibitor cladosporin effectively inhibited overall cell growth, single cell-based colony formation and cell migration. This was mechanistically mediated by cell cycling arrest and induction of apoptosis. Finally, these inhibitory effects were confirmed in 3D cultured patient-derived CC organoids.

Conclusions: The biological process of charging tRNA-Lys-CUU with lysine sustains liver cancer cell growth and migration, and is clinically relevant in HCC patients. This process can be therapeutically targeted and represents an unexplored territory for developing novel treatment strategies against liver cancer.

Keywords: cladosporin, liver cancer, lysine, Lysyl-tRNA Synthetase, tRNA-Lys-CUU, tRNAome

Lay summary

Dysregulation of the protein synthesis machinery in cancer has a major impact on pathophysiology, including cancer development and progression. This study found that the biological process of charging lysine to the corresponding tRNA is altered in liver cancer patients. This in turn can be explored for developing potential therapeutic strategies to treat liver cancer.

Introduction

Translation, the synthesis of protein from mRNA, is a fundamental biological process required for virtually all cellular functions.[1] A key step in protein synthesis is the recognition of codons by the transfer RNAs (tRNA) charged with their corresponding amino acids.[2,3] tRNAs are transcribed from genomic DNA, have a length of typically 70-90 base pairs and are subjected to post-transcriptional modification. The addition of a common CCA ribonucleotide sequence to the 3' end of tRNA enables charging with an amino acid. Aminoacyl-tRNA synthetases (ARSs) catalyze the charging of amino acids to their cognate mature tRNAs, thereby providing the substrates for global protein synthesis.[4] Theoretically, 61 types of aminoacyl-tRNAs would be required for decoding the 61 triplet codons that specify 20 amino acids. However, the minimally required tRNA species for decoding in real life is usually less than the theoretically calculated number.[5] From the GtRNA 2.0 database, we have recently retrieved 57 human mature tRNA species constituting as the tRNAome.[6]

Dysregulation of gene translation has been well-recognized in cancer development and progression.[7,8] Involvement of tRNAs has been demonstrated in breast cancer, lung cancer and melanoma.[9] However, despite the biological importance of tRNA, it has hardly been investigated in depth, which includes the setting of cancer research. One of the main challenges is that tRNAs are technically difficult to be quantified due to redundancy in genomic copies, extremely short sequences, rigid secondary structure and post-transcriptional modifications. A recent study has analyzed the global expression landscape of tRNAs across 31 cancer types. They used the Cancer Genome Atlas database and analyzed tRNA profiles from miRNA-sequencing data, as an indirect interpretation of tRNA expression.[10] However, the functional forms are the mature tRNAs charged with amino acids. Based on these unique characteristics, we have developed a simplified qRT-PCR assay to quantify the mature tRNAome.[6]

Several enzymes participate in the modification and maturation of tRNAs. ARSs responsible for charging amino acids have evolutionarily conserved enzymatic properties. As protein synthesis is universally accelerated in proliferating malignant cells, the expression levels of many ARSs are increased.[11] In contrast to tRNAs, the functions of ARSs have been widely investigated and various therapeutic approaches have been explored, including the development of pharmacological inhibitors targeting the catalytic sites.[12] This has provided new avenues for identifying unexplored therapeutic targets and developing effective anti-cancer therapeutics.

Primary liver cancer, mainly consisting of hepatocellular carcinoma (HCC) and cholangiocarcinoma (CC), is among the leading causes of cancer-related deaths with limited treatment options available. In this study, we aim to profile the tRNAome landscape in tumors of HCC patients using our recently developed mature tRNA quantification assay.[6] Secondly,

we aim to assess the therapeutic potential of targeting the interface of amino acid charging to tRNA in experimental models of liver cancer.

Patients, Materials and Methods

Patient tissues and information

Fresh frozen tumor tissues and paired tumor-free liver (TFL) tissues, located at least 2 cm from the tumor, were collected after surgery or retrieved from the archives of the Department of Pathology, Erasmus Medical Center Rotterdam. The included patients underwent hepatic resection (n = 68) or liver transplantation (n = 1) for HCC at the Erasmus Medical Center between February 1995 and September 2017. Diagnosis of HCC was confirmed by histopathological examination and medical records were reviewed for clinicopathological characteristics including date of first recurrence, HCC-specific death, liver transplantation (if applicable), and last follow-up (**Table 1; Supplementary Table 1**). Tissue samples (n = 21) of healthy livers from organ donors that were used for liver transplantation were used as healthy controls. This study was approved by the Erasmus MC Medical Ethical Committee and adhered to the 1975 Declaration of Helsinki.

Reagents

The natural product cladosporin was synthesized at the CSIR-National Chemical Laboratory, India, and the NMR data of cladosporin are shown in Supplementary Figure 1. It was dissolved in dimethyl sulfoxide (DMSO) at a stock concentration of 100 mM. L-Arginine, N-acetylcysteine, gastrin, nicotinamide and 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide (MTT) were purchased from Sigma-Aldrich. Dulbecco's modified Eagle's medium (DMEM, #41965039) and DMEM for SILAC (#88364) were purchased from Life Technologies Europe BV. Advanced DMEM/F-12 (#12634-010) and SILAC Advanced DMEM/F-12 Flex Media, no L-Lysine, no L- Arginine, no glucose, no phenol red (#A2494301) were purchased from Gibco Life Technologies. L-Lysine was purchased from Bio-Connect BV. Matrigel was purchased from BD Bioscience. Cytokines, B27 and N2 were purchased from Invitrogen. EGF, FGF10, and HGF were purchased from Peprotech Company.

Table 1. Patient characteristics	
Characteristics	**HCC patients (n=69)**
Age at treatment (years)	
Mean ± SD	60 ± 16
Median (range)	64 (11-82)
Sex – no. (%)	
Male	41 (59.4)
Female	28 (40.6)
Race – no. (%)	
White	58 (84.1)
African	6 (8.7)
Asian	4 (5.8)
Not reported	1 (1.4)
Etiology – no. (%)	
No known liver disease	20 (29.0)
Alcohol	16 (23.2)
NASH	8 (11.6)
Hepatitis B	8 (11.6)
Hepatitis C + Alcohol	5 (7.2)
Hepatitis B + Alc/HepC/HepD/NASH	4 (5.8)
Hepatitis C	4 (5.8)
Fibrolamellar HCC	3 (4.4)
Other	1 (1.5)
Hepatitis status – no. (%)	
Hepatitis B or C positive	21 (30.4)
Chronic Hepatitis B	12 (17.4)
Chronic Hepatitis C	10 (14.5)
Cirrhosis – no. (%)	
Yes	50 (72.5)
No	19 (27.5)
Differentiation grade – no. (%)	
Good	8 (11.6)

Cell culture and amino acid deprivation

HCC cell lines, including Huh6, Huh7, PLC/PRF/5, SNU398, SNU449, SNU182, Hep3B, HepG2, and human embryonic kidney epithelial cell line 293T (HEK293T) were grown in DMEM (GIBCO Life Technologies), supplemented with 10% (v/v) fetal bovine serum (Hyclone Technologies), 100 units/mL of penicillin and 100 μg/mL of streptomycin. To establish a lysine deprivation model, L-arginine (0.4 mM; the same concentration as in DMEM) was firstly added to "DMEM for SILAC" (L-lysine and L-arginine deficiency) medium, and then different concentrations of L-lysine were added. All the cells were incubated at 37°C in a humidified atmosphere containing 5% CO2. Authentication of all the cell lines was performed using the short tandem repeat genotyping assay at the Department of Pathology, Erasmus Medical Center Rotterdam. The mycoplasma-free status was regularly commercially checked and confirmed by GATC Biotech Konstanz, Germany.

To further test the deprivation of each of the nine essential amino acids, culture medium was prepared with Earle Balanced Salt Solution (H9269, Sigma-Aldrich), Glucose and MEM vitamin solution (100x) (M6895, Sigma-Aldrich). Different amino acids except for the deprived one were then added, according to the concentrations of amino acids in regular complete DMEM medium (#41965, Gibco).[13,14]

Organoid culture and lysine deprivation

The establishment and maintenance of human normal liver and liver tumor organoids were performed as described.[15] Three batches of normal liver organoids (HLO-1, HLO-2, HLO-3) were derived from intrahepatic biliary epithelial progenitor cells. Three batches of liver tumor organoids (CCO-1, CCO-2, CCO-3) were derived from tumor tissue of CC patients. The use of human organoids was approved by the Erasmus MC Medical Ethical Committee. To establish a lysine deprivation model, L-Arginine (0.7 mM; the same concentration as in Advanced DMEM/F-12 DMEM) was firstly added to "Advanced DMEM/F-12 without L-Lysine, L- Arginine and no-phenol red" medium, and then different concentrations of L-Lysine were added. The culture medium for human normal liver organoids and liver tumor organoids was basic medium (Advanced/F12 DMEM with 1% penicillin/streptomycin, 1% L-Glutamine and 10 mM HEPES) supplemented with 1:50 B27 supplement (without vitamin A), 1:100 N2 supplement, 1 mM N-acetylcysteine, 10% (vol/vol) R-spondin 1 (conditioned medium), 10 mM nicotinamide, 10 nM recombinant human [Leu[15]]-gastrin I, 50 ng/ml recombinant human EGF, 100 ng/ml recombinant human FGF10, 25 ng/ml recombinant human HGF, 10 μM Forskolin and 5 μM A83-01. The medium was replaced every 3 days and passage was performed according to the growth of organoids.

Lentiviral shRNA packaging and transduction

To perform gene knockdown, pLKO.1-based lentiviral vectors (Sigma-Aldrich) targeting KARS and a scrambled control vector (shCTR) were obtained from the Erasmus Medical Center for Biomics. Lentiviral pseudoparticles were generated in HEK293T cells. After 2 days of transfection with the lentiviral particles, Huh7 and SNU398 cells were subsequently selected by 3 µg/mL puromycin (Sigma-Aldrich) for one week. The knockdown efficiency was confirmed by qRT-PCR and Western Blot. The target sequences of selected shRNAs are listed in Supplementary Table 2.

RNA isolation and qRT-PCR

RNA was isolated by the NucleoSpin® RNA isolation kit of Macherey-Nagel (Dueren, Germany) and reverse-transcribed into cDNA using the PrimeScript™ RT Master Mix (Perfect Real Time, Takara, cat# RR036A), according to the manufacturer's instructions. mRNA quantification by qRT-PCR was performed using SYBR™ Green PCR Master Mix (ThermoFisher) in a StepOnePlus™ Real-Time PCR System (Applied Biosystems), using 12.5 ng cDNA per reaction, with the following conditions: 50°C for 2 min, 95°C for 2 min, then 38 cycles of 95°C for 15 s, 58°C for 15 s, 72°C for 1 min, followed by the Melt Curve stage of 95°C for 15 s, 60°C for 1 min and a 0.7°C step-wise increase until 95°C was reached. Means of technical replicates were used for calculation of gene expression by $2^{-\Delta\Delta T}$ method. The geometric mean of three housekeeping genes (*HPRT1*, *GUSB*, *PMM1*) was used to normalize gene expression in patient samples. For experimental models, *GAPDH* was used as a reference gene to normalize target gene expression. Primer sequences used for qPCR of mRNA are included in Supplementary Table 3. For tRNA quantification, we used the isolation protocol and primer sequences that we previously described.[6]

Western blotting

Cells were lysed in Laemmli sample buffer with 0.1 M DTT and heated for 5-10 min at 95°C, followed by loading and separation on an 8-15% sodium dodecyl sulphate-polyacrylamide gels. After 90 min running at 100V, proteins were electrophoretically transferred onto a polyvinylidene difluoride membrane (Invitrogen) for 1.5 h with an electric current of 250 mA. Subsequently, the membrane was blocked with blocking buffer (Li-COR, Lincoln, USA) mixed with PBST in the ratio of 1:1. This was followed by overnight incubation with rabbit anti-KARS (Sanbio BV, #14951-1-AP), anti-Cleaved-Caspase 3 (CST, #9664S), and mouse anti-β-actin (Santa Cruz, #SC-47778) at 4°C. The membrane was washed 3 times with PBST, which was followed by incubation for 1 h with anti-rabbit (1:10000) or anti-mouse (1:5000) IRDye conjugated secondary antibodies (Li-COR, Lincoln, USA) at room temperature. Blots were assayed for actin content as standardization of sample loading and scanned and quantified by Odyssey infrared imaging (Li-COR, Lincoln, USA). The results were analyzed with Image Studio software.

Analysis of cell cycle

Huh7 and SNU398 cells ($5×10^5$/well) were plated in 6-well plates and incubated overnight to attach to the bottom, and then serial concentrations of Lysine were added. After incubated 3 days, cells were harvested and fixed in cold 70% ethanol for 2 h at 4°C. The cells were washed twice with PBS and incubated with 50 µL RNase (100 µg/ml) at 37°C for 30 min. Finally, cells were incubated with 250 µL Propidium Iodide (PI, 50 µg/ml) at room temperature for 5 min. The samples were analyzed immediately by flow cytometry using a FACS Canto (BD Biosciences), and the cell cycle was analyzed by using FlowJo_V10 software

Colony and organoid formation assay

For colony formation assay, HCC cells were harvested and suspended in medium, then seeded into 6-well plates (1000 cells/well) and the medium was refreshed every 4 days. After two weeks of culture, formed colonies were washed with PBS and fixed by 70% ethanol, followed by counterstaining with crystal violet (Sigma-Aldrich) and washed with PBS. Colony sizes were measured microscopically by digital image analysis.

For the organoid formation assay, intact organoids were first digested into single cells, and then seeded into 48-well plates (3000-5000 cells/well). After two weeks, the sizes and numbers of formed organoids were calculated.

MTT and Alamar Blue assays

HCC cells were seeded in 96-well plates at a concentration of $3×10^3$ cells/well in 100 µL medium. Upon indicated treatment, cell viability was analyzed by adding 5 mg/mL MTT for 4 h. After discarding the cell supernatant, 100 µL DMSO was added followed by 10 min shaking. Absorbance was determined using the microplate absorbance reader (Bio-Rad, Hercules, CA, USA) at a wavelength of 490 nm.

Organoids were split in the ratio of 1:10 for daily culture and seeded in 24-well plates. On the third or seventh day of indicated treatment, organoids were incubated with Alamar Blue (Invitrogen, 1:20 in Advanced/F12 DMEM) for 4 h, and the medium was collected for measurement of the metabolic activity of the organoids. Absorbance was determined by a fluorescence plate reader (CytoFluor® Series 4000, Perseptive Biosystems) at the excitation of 530/25 nm and emission of 590/35 nm.

Cell migration assay

Cells ($5×10^5$ /well) were collected and suspended in 300 µL of serum-free medium after the indicated treatment for 48 h. Cells were placed into the upper chamber of a 24-transwell (8-µm pore size; Corning, #353097), and 700 µL of DMEM containing 20% FBS was added into the lower chamber. Cells were allowed to migrate to the bottom chamber for 24 h. Next, cells that remained on the apical side of the chamber were gently scraped off using wetted cotton

swabs. Subsequently, cells were fixed with 4 % paraformaldehyde solution for 30 min and stained with 0.1% crystal violet for 30-60 min. Finally, the cells were observed with a microscope, and images were obtained.

Statistics

All statistical analyses were performed using Graphpad (Version 5 for Windows, San Diego, CA) and R Statistical software (Version 3.6.1 for Windows, Foundation for Statistical Computing, Vienna, Austria). The expression level of genes between HCC and TFL patient tissues were compared using the Wilcoxon matched pairs test. For functional experiments, potential differences between the experimental groups were analyzed using the Mann-Whitney test. The correlation analysis was performed in RStudio with the "corplot" package, using Spearman's correlation coefficient. For creating heatmaps, RStudio was used with the "gplots" and "pheatmap" packages. Survival analysis was performed by the Kaplan-Meier method and the Cox proportional hazards model. Time to the event, either HCC-specific death or HCC-recurrence, was calculated from the day of surgery. If the event of interest did not occur, data were censored at the time of the last follow-up, or if a patient underwent liver transplantation, at the time of liver transplantation. Used statistical tests are indicated in the figures. P-values < 0.05 were considered significant, and indicated with a single asterisk. P-values < 0.01 were indicated with double asterisks and P-values < 0.001 were indicated with triple asterisks.

Results

Profiling the tRNAome landscape in patient HCC tumors identified the potential clinical significance of tRNA-Lys-CUU

Mature tRNAs with a CCA tail are charged with amino acids to elongate protein synthesis and are thus biologically most relevant. To focus on this functional species of tRNAs, we recently developed a specific qRT-PCR quantification assay involving the removal of attached amino acids and a universal adaptor ligated to the CCA tail.[6] We profiled the tRNAome consisting of 57 mature tRNA species in 69 pairs of tumors and matched TFL tissues of HCC patients. The heatmap depicted in **Figure 1A** shows the relative expression levels of 57 mature tRNAs in 69 pairs of tumor and TFL tissues. Significantly differentially expressed tRNAs between tumor and TFL tissues were listed in Supplementary Table 4, showing that many tRNAs were down-regulated in tumors. Interestingly, the patterns of tRNA expression were closely clustered and correlated based on the individual tissues (**Supplementary Figure 2A**), suggesting that different members of the tRNAome function cooperatively in protein translation. Next, the expression of each individual tRNA in tumor tissues was further analyzed to identify potential associations with clinical features. We found that tRNA-Lys-CUU has potential clinical relevance. Kaplan–Meier analysis showed that high expression of tRNA-Lys-CUU in tumor appears to associate with more tumor recurrence (**Figure 1B**) and worse patient survival late after tumor resection (**Figure 1C**), although not statistically significant. Interestingly,

univariate Cox regression analysis indicated that high levels of tRNA-Lys-CUU in tumor were significantly associated with increased rates of tumor recurrence (HR 1.1; p = 0.022) and HCC-specific death (HR 1.1; p = 0.0037) (**Figure 1B,C**; lower panels). In multivariate analysis together with clinicopathological characteristics, tRNA-Lys-CUU expression in tumor was an independent negative prognostic factor for both HCC-specific survival and HCC recurrence (**Supplementary Figure 2B**). As TFL tissues from HCC patients often show various signs of liver pathology, we also included healthy controls of liver tissues obtained from organ donor livers. The expression of tRNA-Lys-CUU was significantly higher in HCC tumor tissues compared with the healthy liver tissues (**Supplementary Figure 2C**).

Elevated expression of Lysyl-tRNA synthetase in patient HCC tissues

Lysyl-tRNA synthetase (KARS), encoded by *KARS1*, catalyzes the charging of lysine to tRNA-Lys-CUU and tRNA-Lys-UUU.[16] We found that mRNA expression levels of *KARS1* were significantly higher in HCC tumor tissues compared with paired TFL tissues (n = 59) (**Figure 2A; Supplementary Figure 3A**), and this was further confirmed by analysis of the publically available GEPIA database (**Supplementary Figure 3B**). When analyzing the association with clinical outcomes, we found no evidence for a relation with tumor recurrence (**Figure 2B**). However, there was a trend of association with patient survival. In univariate Cox regression analysis patients with high *KARS1* expression seemed to have an increased death rate (HR 1.4; **Figure 2C**), although statistically not significant (p = 0.078). Importantly, multivariate analysis revealed a significantly increased risk of death for high KARS1 expressers (HR 1.66; p=0.021) (**Supplementary Figure 3C**). Altogether, these associations between the expression levels of tRNA-Lys-CUU and *KARS1* in tumor tissues and prognosis of HCC patients indicate a potential clinical relevance for the interface of charging lysine with its tRNA in liver cancer. In addition, we found a significant positive correlation of tRNA-Lys-CUU and *KARS1* in tumor tissues (**Figure 2F**), but not in healthy liver tissues or TFL tissues (**Figure 2D,E**). This promoted us to further explore their functions in experimental models.
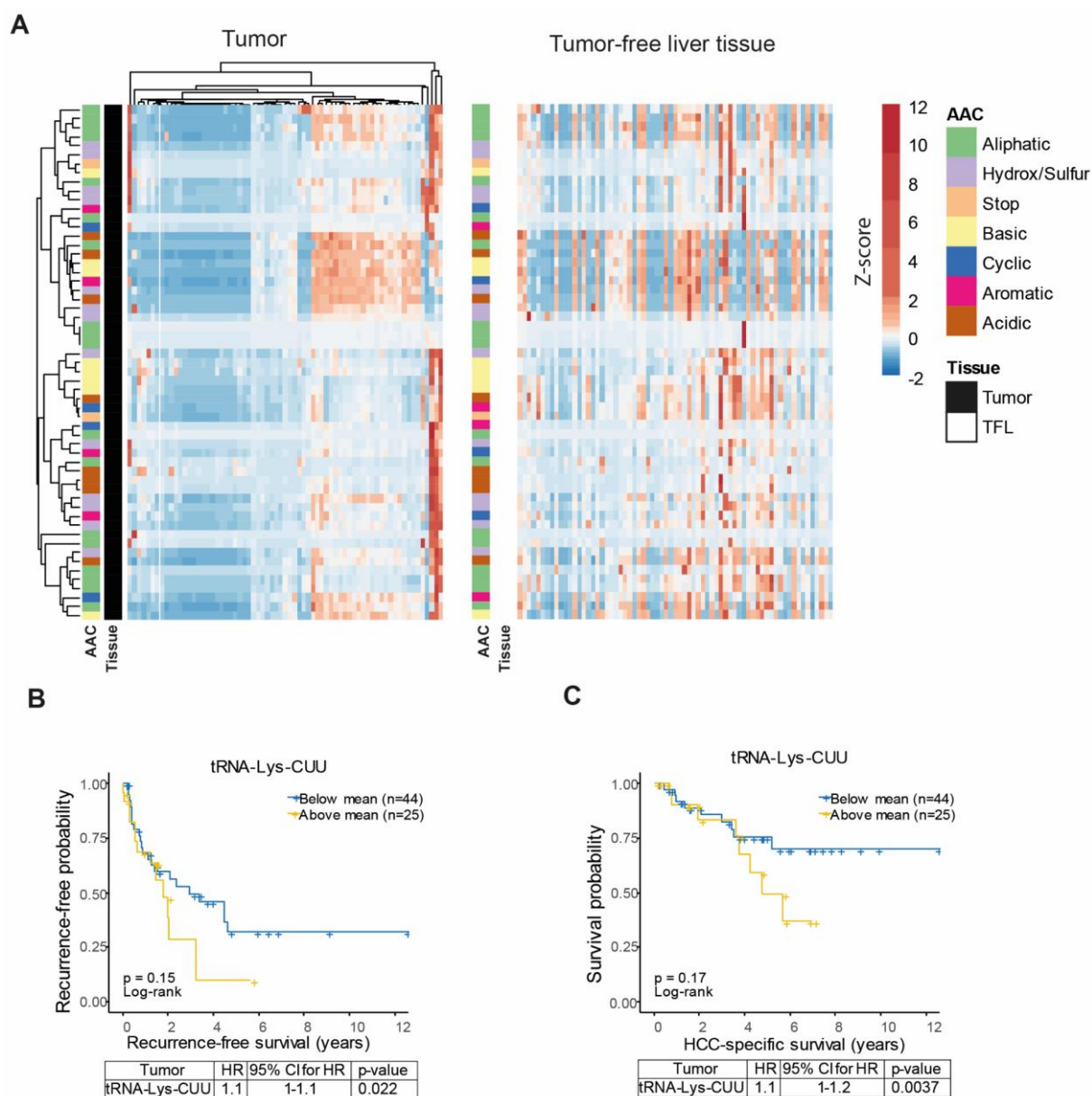
Figure 1 Profiling tRNAome in HCC patients identified tRNA-Lys-CUU having potential clinical significance. (**A**) Heatmap showing the mean expression(z-score normalized per tRNA species, represented in color, as indicated in the legend) of 57 tRNA in tumor and TFL tissues, respectively. Both rows (tRNA) and columns (patient samples) of tRNAs expressed in tumors (left panel) are hierarchically clustered using correlation distance and complete linkage. Rows and columns of tRNAs expressed in TFL (right panel) were ordered according to the rows and columns in the tumor heatmap.(**B-C**) Kaplan Meier analysis and Cox regression analysis of tumor recurrence (**B**) and HCC-specific survival (**C**) in relation to tRNA-Lys-CUU expression. The panels below the survival graphs show the results of univariate Cox regression analysis. HR: Hazard Ratio; AAC: amino acid class. HCC: hepatocellular carcinoma; TFL: tumor free liver; CI: confidence interval.

Knockdown of KARS inhibits HCC cell growth

To study the functionality of KARS in liver cancer cells, we first examined *KARS1* expression at mRNA and protein levels (**Figure 3A,B**), and tRNA-Lys-CUU expression across eight HCC cell lines (**Supplementary Figure 4A**). All HCC cell lines expressed higher *KARS1* RNA levels, and all except one expressed higher levels of tRNA-Lys-CUU than healthy liver organoids. Huh7 and SNU398 cell lines with high levels of *KARS1* and tRNA-Lys-CUU expression were selected for further experimentation. We used the lentiviral RNA interference approach to investigate the effects of silencing KARS on HCC cells, and selected the two shRNAs with optimal efficiency for KARS knockdown (**Supplementary Figure 4B,C; Figure 3C,D**). KARS knockdown attenuated the overall growth of Huh7 and SNU398 cells (**Figure 3E**), and dramatically inhibited single cell-based colony formation (**Figure 3F**). Silencing of KARS arrested cell cycling at the G0/G1 phase (**Figure 3G**) and induced cell apoptosis as indicated by drastically increased levels of cleaved-Caspase 3 protein (**Figure 3C**).

Lysine deprivation inhibits HCC cell growth and migration

Lysine is the substrate which is ligated by KARS to tRNA-Lys-CUU. We first examined the effects of complete deprivation of each of the nine essential amino acids in SNU398 cells. We found that deprivation of any of the nine amino acids had some effects on cell growth, but lysine deprivation exerted the strongest inhibitory effect (**Supplementary Figure 5A**). In single cell-based colony formation assay, deprivation of lysine, methionine or valine but not any of the other six amino acids had significant inhibition, and lysine deprivation again exerted the strongest effect (**Supplementary Figure 5B**). These results collectively indicated lysine is most essentially required. Next, compared to the normal concentration (0.8 mM) of lysine in cell culture medium, lowering the concentration to 0.16 mM had minimal effects on HCC cell growth (**Figure 4A**). However, complete deprivation of lysine dramatically inhibited overall cell growth (**Figure 4A**) and single cell-based colony formation (**Figure 4B**). This was associated with increased G0/G1 cell cycle arrest (**Figure 4C**), and increased cell apoptosis indicated by the dramatic increase of Caspase 3 cleavage (**Figure 4D**). Interestingly, lysine deprivation also attenuated HCC cell migration in the transwell system, an assay recapitulating some features of cancer cell invasion and metastasis (**Figure 4E, F**).
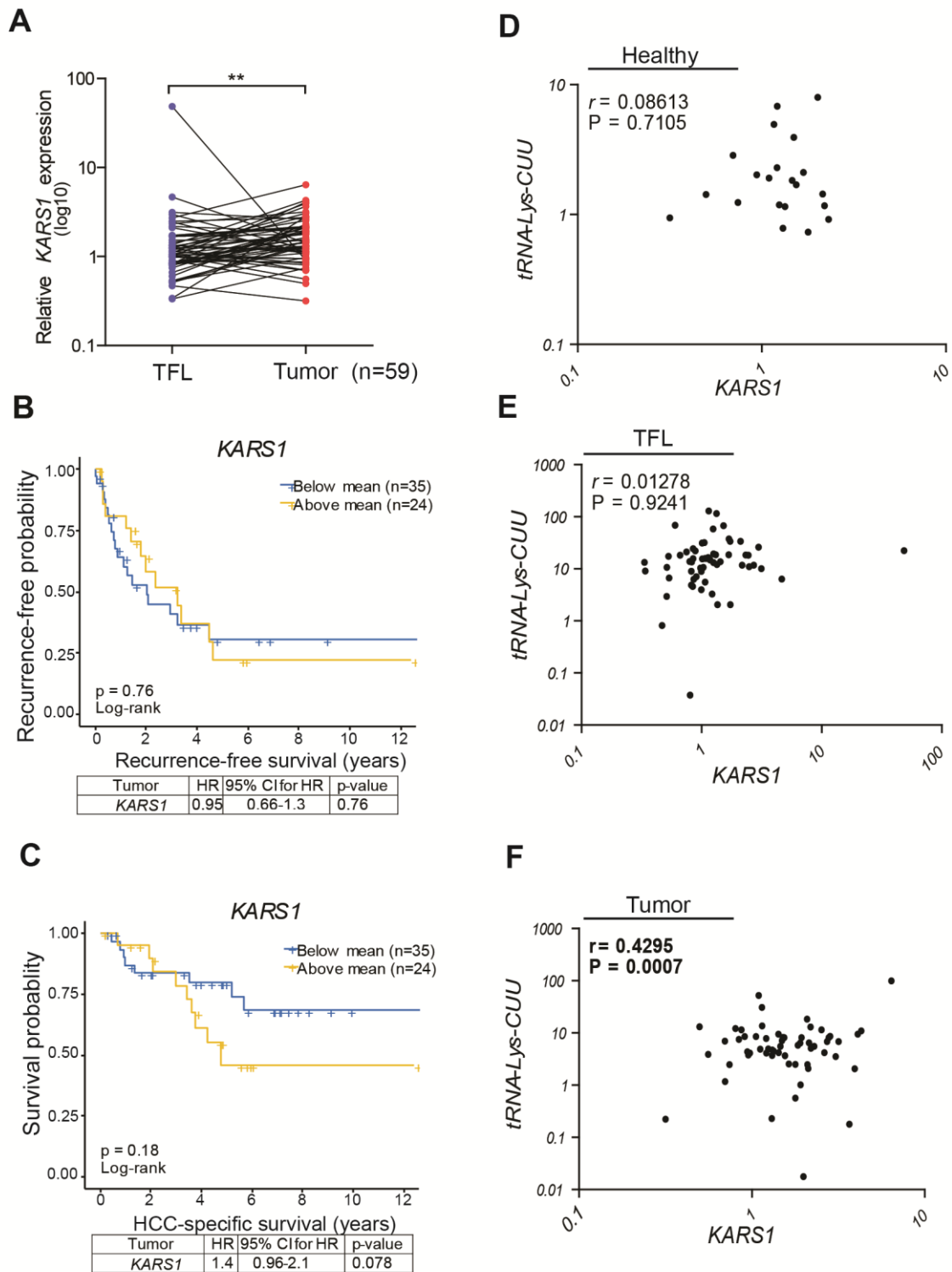
Figure 2 KARS1 expression and clinical relevance in HCC patients. (A) mRNA expression of KARS1 in HCC tumors compared to paired TFL tissues (n = 59). (B-C) Kaplan Meier analysis of tumor recurrence (B) and HCC-specific survival (C) in relation to KARS1 expression in tumors. The panels below the survival graphs show the results of univariate Cox regression analysis. (D-F) Correlation analysis between the expression levels of tRNA-Lys-CUU and KARS1 in healthy organ donor liver tissues (D, n = 21), TFL (E) and paired HCC tissues (F, n = 59),

respectively. **P <0.01, by the Wilcoxon matched pairs test. KARS1: Lysyl-tRNA Synthetase; HCC: hepatocellular carcinoma; TFL: tumor free liver; HR: Hazard Ratio; CI: confidence interval.

Cladosporin, a pharmacological inhibitor of KARS, inhibits HCC cell growth and migration

As the biological process of charging lysine with tRNA sustains HCC cells, we next examined whether this can be therapeutically targeted. We evaluated the effects of the well-established KARS pharmacological inhibitor cladosporin, an antifungal antibiotic isolated from *Cladosporium cladosporioides* and *Aspergillus flavus*. Recently we accomplished a scalable synthesis of the active natural product cladosporin, producing more than 2 grams of the compound.[17] We found the 50% inhibitory concentrations of cladosporin to be 104 µM and 70 µM in Huh7 and SNU398 cells, respectively (**Supplementary Figure 6**). Treatment with different concentrations of cladosporin (0-200 µM) dose-dependently inhibited Huh7 and SNU398 cell growth (**Figure 5A**). Consistently, cladosporin significantly inhibited single cell-based colony formation (**Figure 5B**), induced Caspase 3 cleavage (**Figure 5C**), and affected cell migration (**Figure 5D, E**). Combination of cladosporin treatment and *KARS1* gene silencing further augmented the inhibitory effects (**Supplementary Figure 7A, B**).
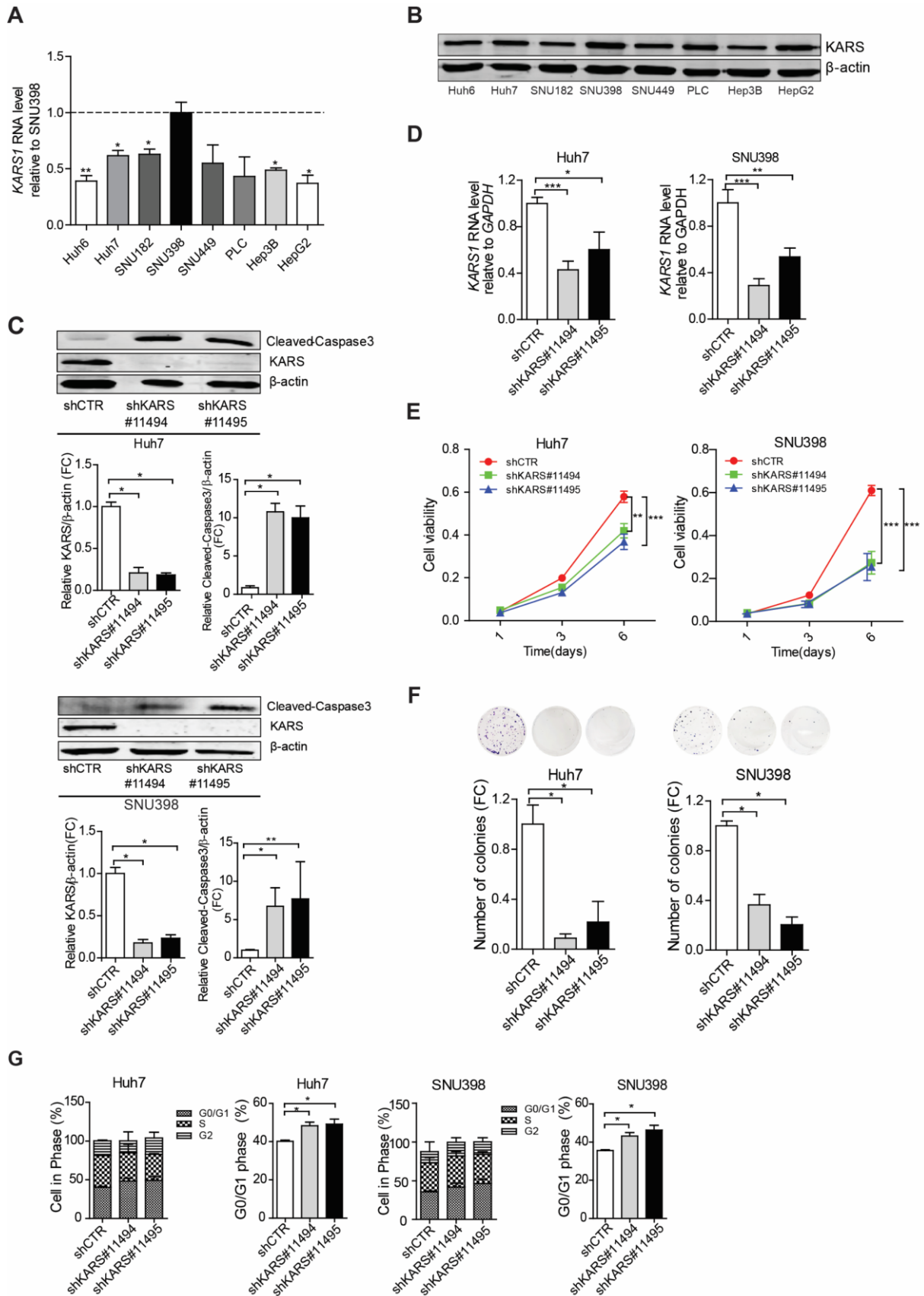
Figure 3 The effects of KARS knockdown on HCC cell lines. (A) Relative expression of *KARS1* mRNA in HCC cell lines compared to SNU398 (n = 4-6). (B) KARS protein expression in HCC cell lines. (C) Protein expression levels of KARS and cleaved-Caspase 3 in Huh7 and SNU398 HCC

cells following shRNA mediated knockdown of KARS. The intensity was quantified relative to β-actin (n =4-8). (D) *KARS1* mRNA expression following shRNA mediated knockdown quantified by qRT-PCR (n= 4-10). (E) Effects of KARS knockdown on cell growth measured by MTT assay following 1, 3 and 6 days of culture (n = 11, 22). (F) KARS knockdown affects the number of single cell formed colonies as assayed 2 weeks following seeding (n = 5). (G) The effect of KARS knockdown on cell cycling was measured by Propidium Iodide staining, and the fraction of cells in G0/G1 was quantified (n = 4-6). Quantification of C, D, and F data of knockdown groups were relative to the shCTR group. Data are presented as mean ± SEM. *P < 0.05; **P <0.01; ***P <0.001, by the Mann-Whitney test. HLO: healthy liver organoids; HCC: hepatocellular carcinoma; FC: fold change; *KARS1*: Lysyl-tRNA Synthetase.

Clinical relevance of KARS expression in CC patients and inhibition of patient CC organoids growth by lysine deprivation or treatment with cladosporin

We have included three batches of healthy human liver organoids and three batches of CC organoids. Because organoids are very difficult to be cultured from HCC tissues, we used liver tumor organoids cultured from CC patients,[18] of which the *KARS1* expression levels were shown in **Figure 6A**. Importantly, similar to what we found in HCC patients, *KARS1* expression was significantly higher in patient CC tumor tissues based on the GEPIA database. However, in this small CC cohort, *KARS1* expression was not significantly associated with patient's prognosis (**Supplementary Figure 8**). Thus, we tested the effects of lysine deprivation or cladosporin on patient-derived CC organoids in 3D culture. We found significant growth inhibition of three independent batches of CC organoids, with the effects being more robust after 7 days of treatment (**Figure 6B,C; Supplementary Figure 9A,B**). In parallel, we examined the effects on three independent batches of healthy human liver organoids. Although lysine deprivation or cladosporin treatment also inhibited growth of the healthy liver organoids, the effects were significantly stronger on the CC organoids (**Figure 6D,E; Supplementary Figure 9C,D**). Next, we tested the effects on organoid initiation. Lysine deprivation or cladosporin treatment significantly inhibited the number and the size of initiated CC organoids from single cells (**Figure 6F,G; Supplementary Figure 9E-H**). Interestingly, the responsiveness varies among organoids derived from different patients (**Figure 6**). Taken together, the biological process of charging tRNA with lysine sustains HCC cell growth and can be therapeutically targeted as demonstrated in cell culture and patient organoid models.
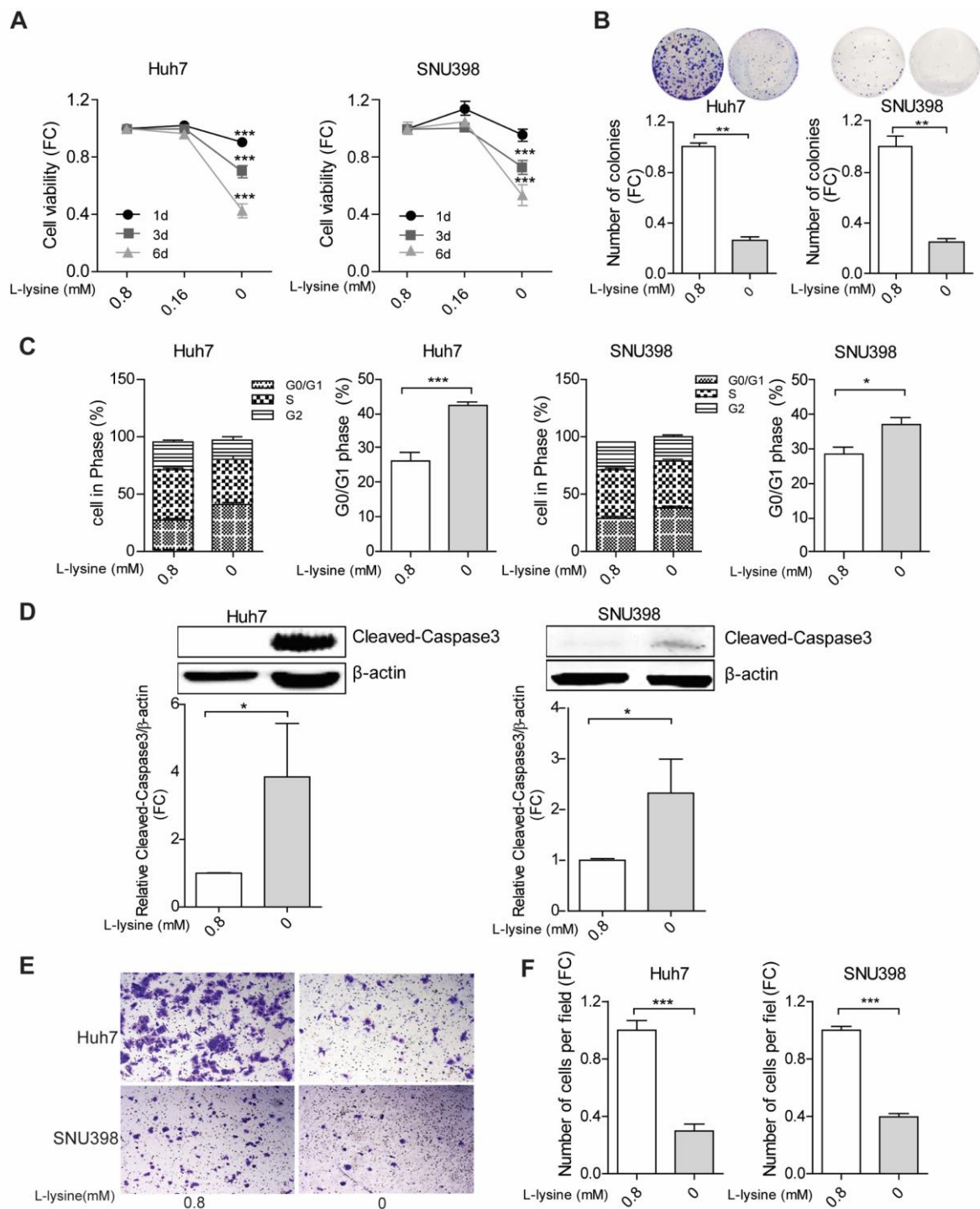
Figure 4 The effects of lysine on HCC cell lines. (A) The effects of lysine on cell growth measured by MTT assay following 1, 3 and 6 days of culture (n = 20). (B) Lysine affects the number of single cell-derived colonies as assayed 2 weeks following seeding (n = 6). (C) The effects of lysine on cell cycling was measured by Propidium Iodide staining, and the fraction of cells in G0/G1 was quantified (n = 6-9). (D) Protein expression levels of cleaved-Caspase 3 in Huh7 and SNU398 HCC cells. The intensity was quantified relative to β-actin (n = 5). Huh7 and SNU398 cells were cultured with or without lysine for 3 days. (E) Representative images of migrating cells with or without lysine and (F) quantification of the number of migrating cells (n = 10-20). Quantification of A, B, D and F data without lysine groups were relative to with

lysine group. Data are presented as mean ± SEM. *P<0.05; **P<0.01; ***P<0.001, by the Mann-Whitney test. HCC: hepatocellular carcinoma; FC: fold change.
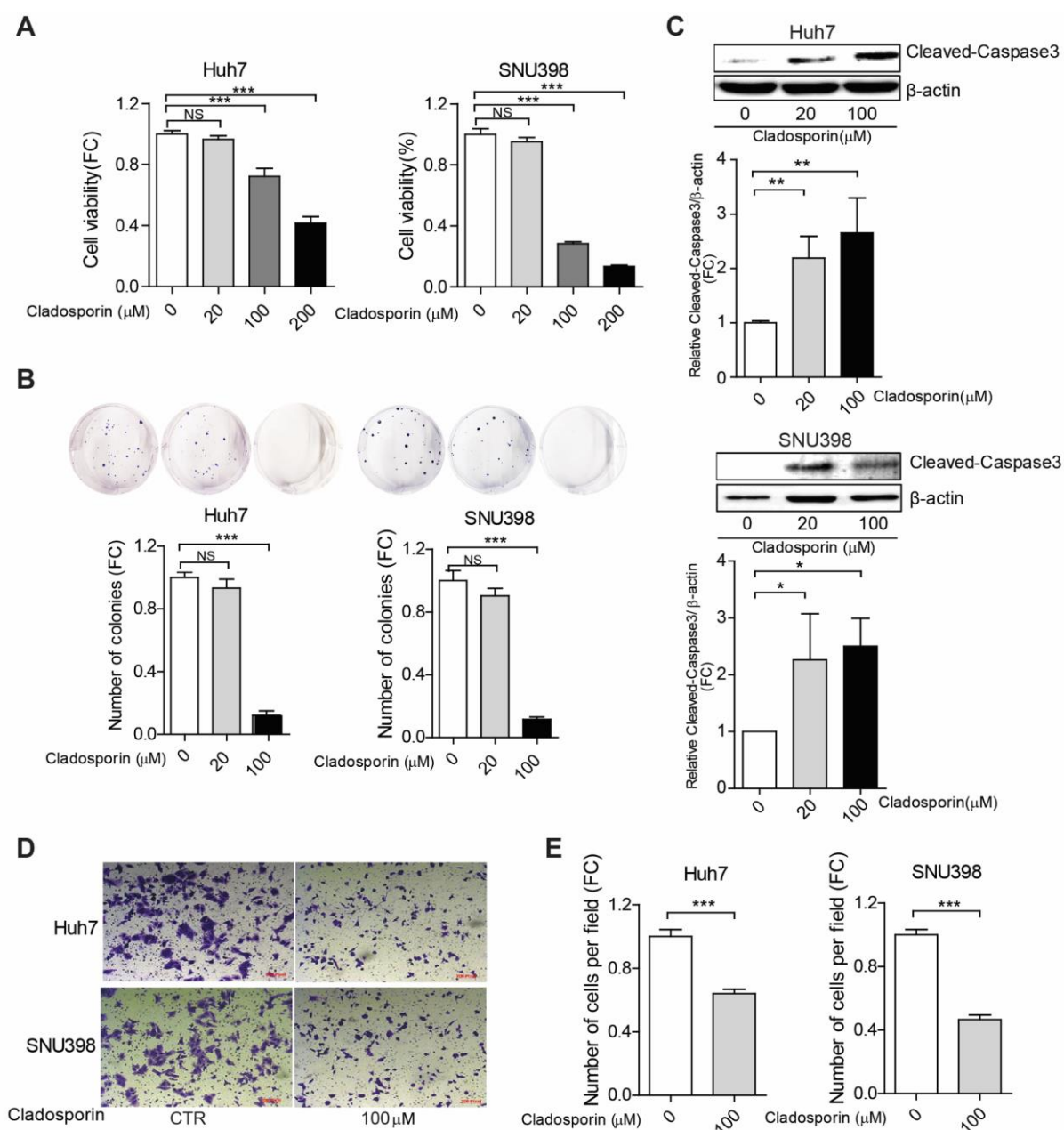
Figure 5 The effects of cladosporin treatment on HCC cell lines. (A) Huh7 and SNU398 cells were treated with different concentrations of cladosporin for 3 days. Cell viability was measured by MTT assay (n = 12). (B) Cladosporin affects the number of single cell-derived colonies as assayed 2 weeks following seeding (n = 6-8). (C) Protein expression levels of cleaved-Caspase 3 in Huh7 and SNU398 cells, and the intensity was quantified relative to β-actin (n = 5). Huh7 and SNU398 were treated with cladosporin for 3 days. (D) Representative images of migrating cells with cladosporin and (E) quantification of the number of migrating cells (n = 20-25). Quantification of all data were relative to the negative control group. Data are presented as mean ± SEM. *P<0.05; **P<0.01; ***P<0.001, by the Mann-Whitney test. HCC: hepatocellular carcinoma; FC: fold change.
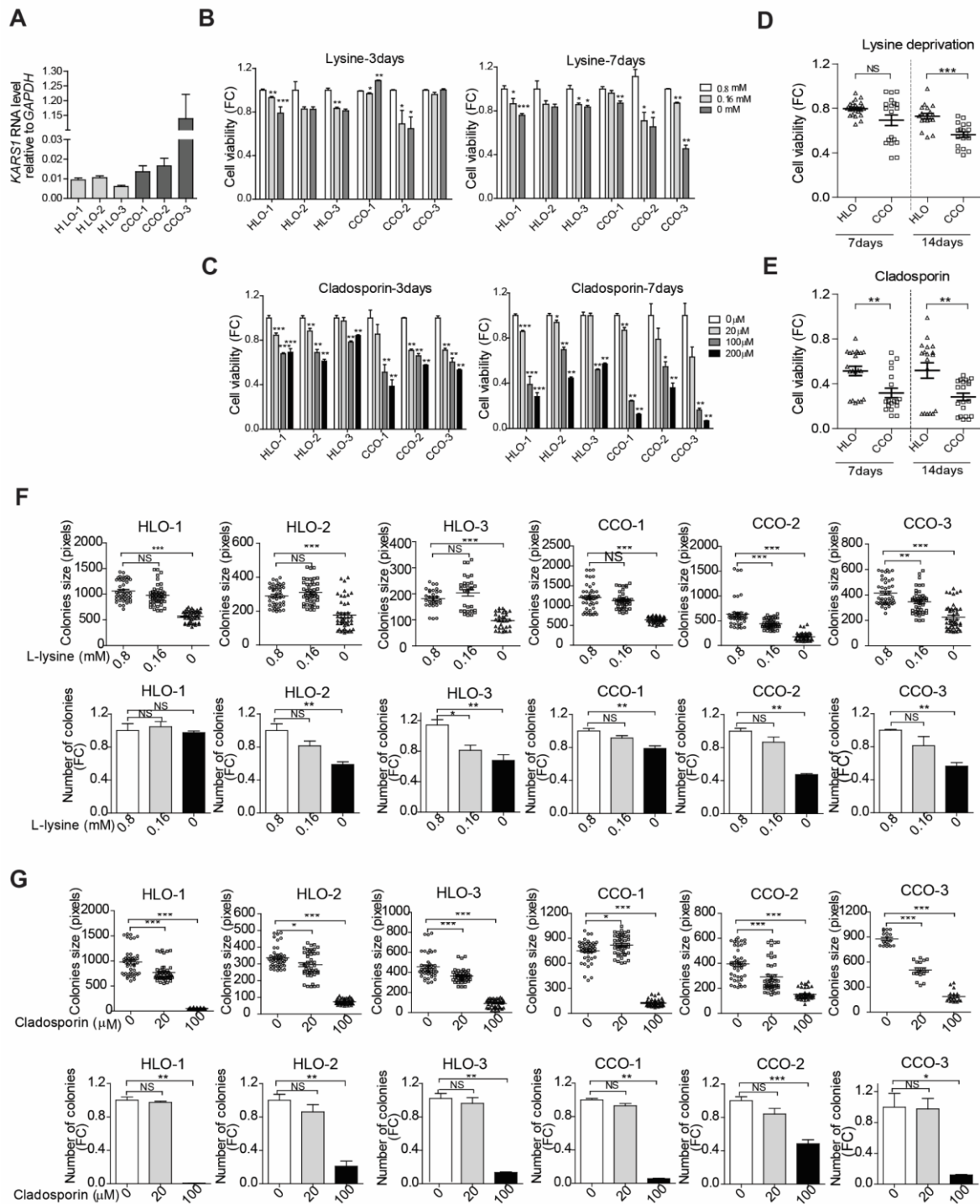
Figure 6 Lysine deprivation or treatment with cladosporin affects the growth of patient-devied CC organoids. (A) Relative expression of *KARS1* mRNA in patient CC organoids and human healthy liver organoids (HLO) (n = 5-11). Effects of (B) lysine deprivation or (C) cladosporin treatment on cell viability were measured by Alamar blue assay after culturing 3 or 7 days (n = 4-12). Organoids were treated with (D) lysine deprivation or (E) cladosporin (100 µM) for 7 or 14 days. The results depicts the mean viability of CC organoids (CCO-1,CCO-2, CCO-3) compared with the mean viability of healthy liver organoids (HLO-1, HLO-2, HLO-3) (n = 18-24). (F) Lysine deprivation or (G) cladosporin treatment affects organoid initiation as assayed 2 weeks following seeding. The number (n = 4-10) and size (n = 18-45) of organoids

after culturing for 14 days were calculated. Quantification of the data target groups were relative to the negative control group. Data are presented as mean ± SEM and dots represent individual organoid cultures. *P<0.05, **P<0.01, ***P<0.001, by the Mann-Whitney test. CCO: cholangiocarcinoma organoid; HLO: human liver healthy organoid; FC: fold change.

Discussion

tRNAs perform central functions in protein synthesis. Intuitively, dysregulation of tRNA expression is expected to associate with pathogenesis, such as cancer development.[19-21] However, due to the intrinsic complexity and lack of easy techniques for quantification, the role of tRNAs in cancer has rarely been studied. The recently launched tRNAscan-SE program that can accurately identify genomic tRNA sequences has facilitated the development of tRNA-based quantification methods.[22] High-throughput sequencing techniques have been used to quantify the tRNAome at transcriptional level,[23] but they are highly dependent on programming and very specialized reagents. We have recently developed a novel qPCR method that specifically quantifies mature tRNAs.[6] We think this form of tRNAs is functionally most relevant as these mature tRNAs are directly charged with amino acids and thus facilitate protein synthesis.

In this study, we first profiled the tRNAome consisting of 57 mature tRNA species in 69 pairs of tumors and matched TFL tissues of HCC patients. We surprisingly observed that many tRNAs are down-regulated in HCC tumor compared to TFL. This is different from previous studies comparing tRNA expression between cancer cell lines and noncancerous cell lines or human cancer and healthy non-cancer tissues.[21] Of note, TFL tissues represent a unique intermediate state between healthy liver tissue and tumor tissue. In many HCC patients, TFL is chronically inflamed and cirrhotic (**Table 1**). Indeed, expression levels of tRNA-Lys-CUU in both tumors and tumor-free liver tissues of HCC patients were significantly enhanced compared to healthy livers. Interestingly, we found that the expression patterns of tRNAs were closely clustered and correlated within individual tissues, suggesting that they function cooperatively and collectively in protein synthesis. Among these tRNAs, we found that high expression of tRNA-Lys-CUU in HCC tissue appears to associate with poor clinical outcomes, although the effects were moderate and our cohort is too small to draw firm conclusions.

Recently, amino acid deprivation therapy has been widely explored as potential anti-cancer strategy. Among different approaches in interrupting amino acid metabolism in cancer cells, enzymatic depletion is the most promising strategy. This results in deprivation of these exogenously supplied nutrients, and thus inhibits tumor growth.[24] Deprivation of amino acids including asparagine,[25] arginine,[26] glutamine,[27] methionine,[28,29] and phenylalanine[30] has been tested for cancer treatment in experimental models. As we have observed that high expression levels of tRNA-Lys-CUU potentially associates with poor outcomes in HCC patients, we tested whether deprivation of the substrate lysine functionally affects HCC cells in experimental models. Lysine deprivation inhibited overall cell growth, migration and single cell-based colony formation, and induced cell cycling arrest and cell apoptosis in two HCC cell

line models. These effects were further confirmed in 3D cultured patient CC organoids. Importantly, we found HCC cells are more sensitive to deprivation of lysine compared to other essential amino acids. Our findings together with those of previous studies[13,31] suggest the potential of developing anti-cancer strategies based on amino acid deprivation. However, the dependency of particular amino acid varies among different cancer types, and normal cells also require these amino acids to survive. Therefore, which amino acid to be targeted and which therapeutic modality to be used should be further studied.[32]

The structures and functions of ARSs that catalyze the charging of amino acids to cognate tRNAs have been widely studied. A systematic analysis of the expression of ARSs has identified that many of them are highly expressed in tumors and form networks with cancer-associated genes. This also supports the fact that protein synthesis is universally accelerated in proliferating malignant cells.[33,34] We found that the expression of *KARS1* was significantly upregulated in HCC compared to TFL tissues. Similar results were also observed in CC patients. We postulate that the abundance of KARS will enhance the efficiency of catalyzing lysine to charge the tRNAs including tRNA-Lys-CUU. This will accelerate protein translation of genes enriched with cognate codons, therefore promoting cancer cell growth. This hypothesis is supported by our findings in two HCC cell lines that knockdown of KARS inhibited cell growth, migration and colony formation, and induced cell cycle arrest and apoptosis.

ARSs targeted drug development has mainly been examined in the field of infectious diseases by exploiting species-specific structural diversity and catalytic activity.[35] For example, ARS-inhibitors mupirocin and AN2690 are FDA-approved for treatment of gram-positive bacterial skin infections and fungal nail infections, respectively.[12] With respect to oncology, ARSs represent overlooked targets for therapeutic development against cancer.[34] In this study, we tested a specific pharmacological KARS inhibitor cladosporin, an antifungal antibiotic isolated from *Cladosporium cladosporioides* and *Aspergillus flavus*.[36,37] We found that cladosporin can effectively inhibit the growth of HCC cell lines and CC organoids, consistent with the results from KARS knockdown and lysine deprivation. Interestingly, the responsiveness to lysine deprivation or cladosporin varies among CC organoids derived from different patients. However, the activity of cladosporin towards human KARS is 100-fold lower than towards KARS of the malaria parasite.[38-40]. Cladosporin is currently explored as a potential treatment for malaria, but whether it is also applicable for treating cancer remains uncertain.[17] Another important consideration is that ARSs are essentially required for maintaining the physiological functions of normal cells. Targeting human ARSs for treating cancer is expected to cause substantial side effects. Indeed, in healthy human liver organoids, we observed that both lysine deprivation and cladosporin treatment resulted in notable inhibitory effects, although to a lesser extent than those in CC organoids. Based on our HCC cell lines and CC organoids, cells expressing higher level of KARS appear to be more sensitive to KARS or lysine targeted inhibition. Collectively, our results support the notion that ARSs are viable targets for anti-cancer drug development, but this is just the start in unveiling their biological functions in cancer and exploring possible therapeutic modalities.

In summary, we have profiled the tRNAome landscape in human HCC tumors and identified that high expression of tRNA-Lys-CUU in tumor is potentially associated with poor clinical outcomes. KARS expression is upregulated in both patient HCC and CC tumor tissues. In experimental models of HCC cell lines and CC patient-derived organoids, biological or pharmacological targeting the interface of charging lysine to tRNA-Lys-CUU inhibits cancer cell growth and migration. These findings bear important implications of exploring such unexplored territories for developing novel therapeutics against liver cancer.

References

1.   Gingold H, Pilpel Y. Determinants of translation efficiency and accuracy. *Mol Syst Biol.* 2011;7:481.

2.   Wu Q, Medina SG, Kushawah G, et al. Translation affects mRNA stability in a codon dependent manner in human cells. *Elife.* 2019;8.

3.   Chatterjee S, Yadav S. The Origin of Prebiotic Information System in the Peptide/RNA World: A Simulation Model of the Evolution of Translation and the Genetic Code. *Life (Basel).* 2019;9(1):25.

4.   Yao P, Fox PL. Aminoacyl-tRNA synthetases in medicine and disease. *EMBO Mol Med.* 2013;5(3):332-343.

5.   Ou X, Cao J, Cheng A, Peppelenbosch MP, Pan Q. Errors in translational decoding: tRNA wobbling or misincorporation? *PLoS Genet.* 2019;15(3):e1008017.

6.   Ou X, Ma B, Zhang R, et al. A simplified qPCR method revealing tRNAome remodeling upon infection by genotype 3 hepatitis E virus. *FEBS Lett.* 2020.

7.   Hernandez-Alias X, Benisty H, Schaefer MH, Serrano L. Translational efficiency across healthy and tumor tissues is proliferation-related. *bioRxiv.* 2019:782227.

8.   Goodarzi H. Charting the "unknown unknowns" of cancer progression. *Sci Transl Med.* 2017;9(400).

9.   Huang SQ, Sun B, Xiong ZP, et al. The dysregulation of tRNAs and tRNA derivatives in cancer. *J Exp Clin Cancer Res.* 2018;37(1):101.

10.   Zhang Z, Ye Y, Gong J, et al. Global analysis of tRNA and translation factor expression reveals a dynamic landscape of translational regulation in human cancers. *Commun Biol.* 2018;1:234.

11.   Musier-Forsyth K. Aminoacyl-tRNA synthetases and tRNAs in human disease: an introduction to the JBC Reviews thematic series. *J Biol Chem.* 2019;294(14):5292-5293.

12.   Kwon NH, Fox PL, Kim S. Aminoacyl-tRNA synthetases as therapeutic targets. *Nat Rev Drug Discov.* 2019.

13.   Tang X, Wu J, Ding CK, et al. Cystine Deprivation Triggers Programmed Necrosis in VHL-Deficient Renal Cell Carcinomas. *Cancer Res.* 2016;76(7):1892-1903.

14.   Campisano S, Bertran E, Caballero-Díaz D, La Colla A, Fabregat I, Chisari AN. Paradoxical role of the NADPH oxidase NOX4 in early preneoplastic stages of hepatocytes induced by amino acid deprivation. *Biochim Biophys Acta Gen Subj.* 2019;1863(4):714-722.

15.   Broutier L, Andersson-Rolf A, Hindley CJ, et al. Culture and establishment of self-renewing human and mouse adult liver and pancreas 3D organoids and their genetic manipulation. *Nat Protoc.* 2016;11(9):1724-1743.

16.   Melnikov SV, Soll D. Aminoacyl-tRNA Synthetases and tRNAs for an Expanded Genetic Code: What Makes them Orthogonal? *Int J Mol Sci.* 2019;20(8).

17.   Das P, Mankad Y, Reddy DS. Scalable synthesis of cladosporin. *Tetrahedron Letters.* 2019;60(12):831-833.

18.   Li M, Wang L, Wang Y, et al. Mitochondrial Fusion Via OPA1 and MFN1 Supports Liver Tumor Cell Metabolism and Growth. *Cells.* 2020;9(1):121.

19.   Sulima SO, Hofman IJF, De Keersmaecker K, Dinman JD. How Ribosomes Translate Cancer. *Cancer Discov.* 2017;7(10):1069-1087.
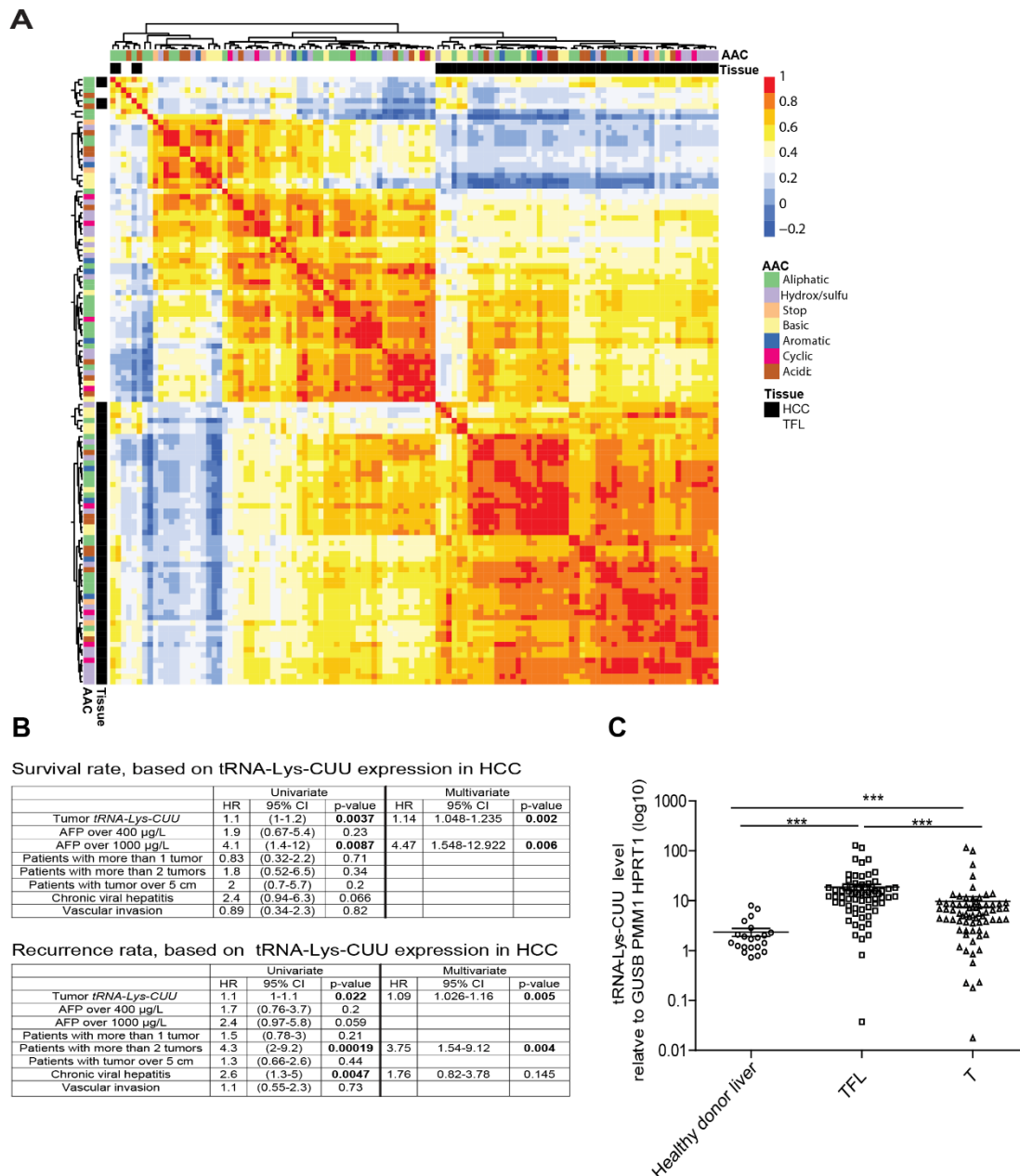
20. Truitt ML, Ruggero D. New frontiers in translational control of the cancer genome. *Nat Rev Cancer.* 2016;16(5):288-304.

21. Santos M, Fidalgo A, Varanda AS, Oliveira C, Santos MAS. tRNA Deregulation and Its Consequences in Cancer. *Trends Mol Med.* 2019.

22. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol Biol.* 2019;1962:1-14.

23. Schwartz MH, Wang H, Pan JN, et al. Microbiome characterization by high-throughput transfer RNA sequencing and modification analysis. *Nature Communications.* 2018;9(1):5353.

24. Dhankhar R, Gupta V, Kumar S, Kapoor RK, Gulati P. Microbial enzymes for deprivation of amino acid metabolism in malignant cells: biological strategy for cancer treatment. *Appl Microbiol Biotechnol.* 2020;104(7):2857-2869.

25. Fernandes HS, Silva Teixeira CS, Fernandes PA, Ramos MJ, Cerqueira NM. Amino acid deprivation using enzymes as a targeted therapy for cancer and viral infections. *Expert Opin Ther Pat.* 2017;27(3):283-297.

26. Zou S, Wang X, Liu P, Ke C, Xu S. Arginine metabolism and deprivation in cancer therapy. *Biomed Pharmacother.* 2019;118:109210.

27. Gwangwa MV, Joubert AM, Visagie MH. Effects of glutamine deprivation on oxidative stress and cell survival in breast cell lines. *Biol Res.* 2019;52(1):15.

28. Jeon H, Kim JH, Lee E, et al. Methionine deprivation suppresses triple-negative breast cancer metastasis in vitro and in vivo. *Oncotarget.* 2016;7(41):67223-67234.

29. Chaturvedi S, Hoffman RM, Bertino JR. Exploiting methionine restriction for cancer treatment. *Biochem Pharmacol.* 2018;154:170-173.

30. Yang J, Tao R, Wang L, et al. Thermosensitive Micelles Encapsulating Phenylalanine Ammonia Lyase Act as a Sustained and Efficacious Therapy Against Colorectal Cancer. *J Biomed Nanotechnol.* 2019;15(4):717-727.

31. Fernandes HS, Silva Teixeira CS, Fernandes PA, Ramos MJ, Cerqueira NMFSA. Amino acid deprivation using enzymes as a targeted therapy for cancer and viral infections. *Expert Opinion on Therapeutic Patents.* 2017;27(3):283-297.

32. Fung MKL, Chan GC. Drug-induced amino acid deprivation as strategy for cancer therapy. *J Hematol Oncol.* 2017;10(1):144.

33. Hyeon DY, Kim JH, Ahn TJ, Cho Y, Hwang D, Kim S. Evolution of the multi-tRNA synthetase complex and its role in cancer. *The Journal of biological chemistry.* 2019;294(14):5340-5351.

34. Kim S, You S, Hwang D. Aminoacyl-tRNA synthetases and tumorigenesis: more than housekeeping. *Nat Rev Cancer.* 2011;11(10):708-718.

35. Francklyn CS, Mullen P. Progress and challenges in aminoacyl-tRNA synthetase-based therapeutics. *The Journal of biological chemistry.* 2019;294(14):5365-5385.

36. Baragana B, Forte B, Choi R, et al. Lysyl-tRNA synthetase as a drug target in malaria and cryptosporidiosis. *Proc Natl Acad Sci U S A.* 2019;116(14):7015-7020.

37. Wang X, Wedge DE, Cutler SJ. Chemical and Biological Study of Cladosporin, an Antimicrobial Inhibitor: A Review. *Natural Product Communications.* 2016;11(10):1934578X1601101039.

38. Hoepfner D, McNamara Case W, Lim Chek S, et al. Selective and Specific Inhibition of the Plasmodium falciparum Lysyl-tRNA Synthetase by the Fungal Secondary Metabolite Cladosporin. *Cell Host & Microbe.* 2012;11(6):654-663.

39. Khan S, Garg A, Camacho N, et al. Structural analysis of malaria-parasite lysyl-tRNA synthetase provides a platform for drug development. *Acta Crystallogr D Biol Crystallogr.* 2013;69(Pt 5):785-795.

40. Khan S, Sharma A, Belrhali H, Yogavel M, Sharma A. Structural basis of malaria parasite lysyl-tRNA synthetase inhibition by cladosporin. *J Struct Funct Genomics.* 2014;15(2):63-71.

Supplementary information

¹H NMR of Cladosporin in 400 MHz

¹H NMR (400 MHz, CDCl₃) δ 11.08 (s, 1H), 6.29 (s, 1H), 6.16 (s, 1H), 4.68-4.66 (m, 1H), 4.11 (s, 1H), 4.00 (s, 1H), 2.89 – 2.76 (m, 2H), 1.90 (dt, *J* = 23.0, 12.8 Hz, 2H), 1.68 (dd, *J* = 21.9, 12.2 Hz, 4H), 1.36 -1.34 (m, 2H), 1.23 (d, *J* = 6.4 Hz, 3H).
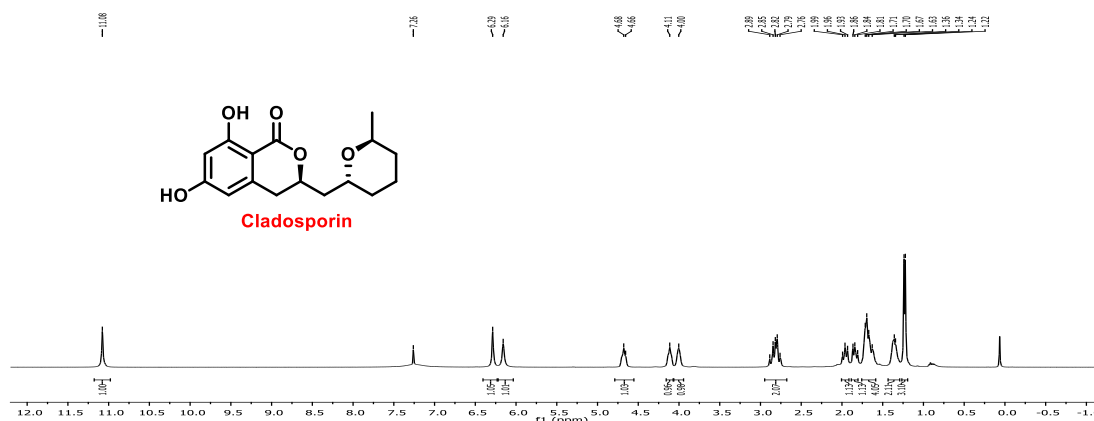
¹³C NMR of Cladosporin in 100 MHz

¹³C NMR (100 MHz, CDCl₃) δ 169.9, 164.3, 163.0, 141.8, 106.7, 102.0, 101.6, 76.3, 68.0, 66.5, 39.3, 33.6, 30.9, 29.7, 18.9, 18.1
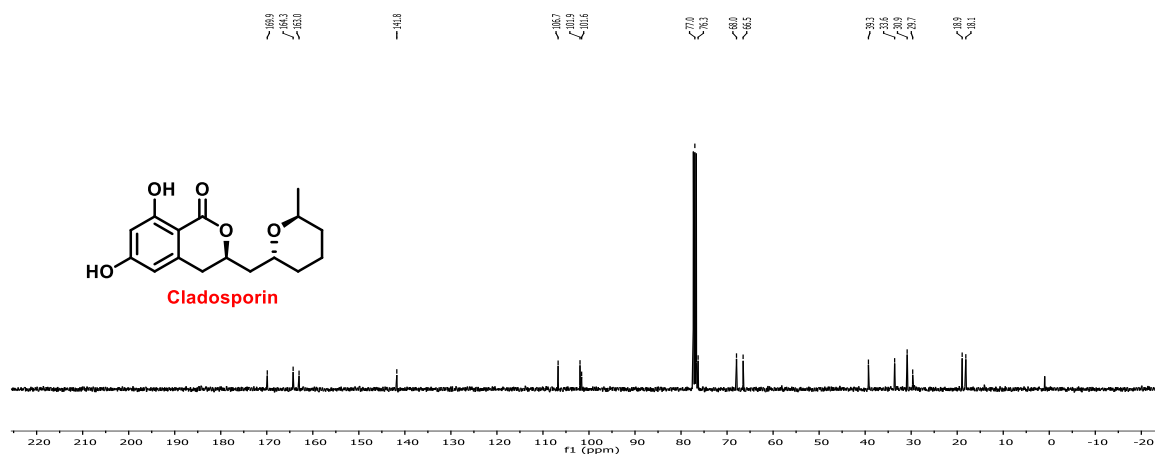
Supplementary Figure 1 NMR for Cladosporin

A

B

Survival rate, based on tRNA-Lys-CUU expression in HCC

	Univariate			Multivariate		
	HR	95% CI	p-value	HR	95% CI	p-value
Tumor *tRNA-Lys-CUU*	1.1	(1-1.2)	**0.0037**	1.14	1.048-1.235	**0.002**
AFP over 400 µg/L	1.9	(0.67-5.4)	0.23			
AFP over 1000 µg/L	4.1	(1.4-12)	**0.0087**	4.47	1.548-12.922	**0.006**
Patients with more than 1 tumor	0.83	(0.32-2.2)	0.71			
Patients with more than 2 tumors	1.8	(0.52-6.5)	0.34			
Patients with tumor over 5 cm	2	(0.7-5.7)	0.2			
Chronic viral hepatitis	2.4	(0.94-6.3)	0.066			
Vascular invasion	0.89	(0.34-2.3)	0.82			

Recurrence rata, based on tRNA-Lys-CUU expression in HCC

	Univariate			Multivariate		
	HR	95% CI	p-value	HR	95% CI	p-value
Tumor *tRNA-Lys-CUU*	1.1	1-1.1	**0.022**	1.09	1.026-1.16	**0.005**
AFP over 400 µg/L	1.7	(0.76-3.7)	0.2			
AFP over 1000 µg/L	2.4	(0.97-5.8)	0.059			
Patients with more than 1 tumor	1.5	(0.78-3)	0.21			
Patients with more than 2 tumors	4.3	(2-9.2)	**0.00019**	3.75	1.54-9.12	**0.004**
Patients with tumor over 5 cm	1.3	(0.66-2.6)	0.44			
Chronic viral hepatitis	2.6	(1.3-5)	**0.0047**	1.76	0.82-3.78	0.145
Vascular invasion	1.1	(0.55-2.3)	0.73			

C

Supplementary Figure 2 (A) Heatmap showing the correlation patterns of 57 tRNA species in 69 pairs of human HCC and TFL tissues. (B) The expression of tRNA-Lys-CUU in organ donor liver tissues (n = 21), TFL and paired HCC tissues (n = 69), respectively. (C) Univariate and multivariate analysis of associations between clinicopathological risk factors and tRNA-Lys-CUU expression in tumor tissue and HCC-specific patient survival rate and cancer recurrence rate (n=69) in our HCC patient cohort. ***P <0.001, by the Mann-Whitney test. TFL: tumor free liver; T:HCC tumor.

A

B
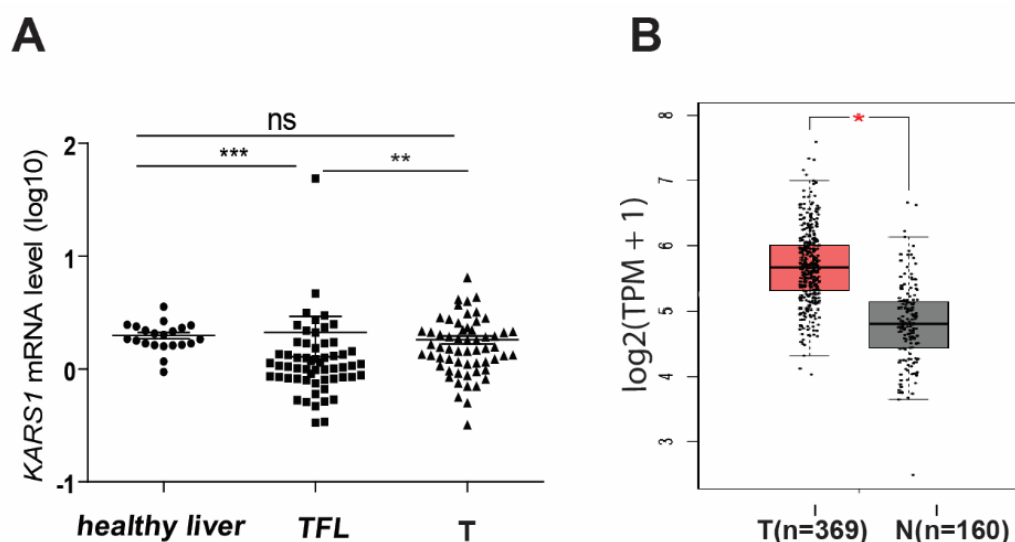
C

Survival rate, based on *KARS1* expression in HCC

	Univariate			Multivariate		
	HR	95% CI	p-value	HR	95% CI	p-value
Tumor *KARS1*	1.4	0.96-2.1	**0.078**	**1.66**	1.08-2.55	**0.021**
AFP over 400 µg/L	1.9	(0.67-5.4)	0.23			
AFP over 1000 µg/L	4.1	1.4-12	**0.008**	5.77	1.86-17.94	**0.002**
Patients with more than 1 tumor	0.78	(0.3-2.1)	0.62			
Patients with more than 2 tumors	1.8	(0.5-6.3)	0.38			
Patients with tumor over 5 cm	2.1	(0.74-6)	0.16			
Chronic viral hepatitis	2.4	(0.92-6.2)	0.073			
Vascular invasion	1.2	(0.48-3.2)	0.66			

Recurrence rata, based on *KARS1* expression in HCC

	Univariate		
	HR	95% CI	p-value
Tumor *KARS1*	0.95	(0.66-1.3)	0.76
AFP over 400 µg/L	1.6	(0.7-3.8)	0.26
AFP over 1000 µg/L	2.3	(0.88-6.1)	0.09
Patients with more than 1 tumor	1.4	(0.68-2.8)	0.38
Patients with more than 2 tumors	3.6	(1.6-8.3)	0.0025
Patients with tumor over 5 cm	1.4	(0.68-2.9)	0.37
Chronic viral hepatitis	2.5	(1.2-5)	0.011
Vascular invasion	0.97	(0.48-2)	0.94

Supplementary Figure 3 The mRNA expression levels of KARS1, and their relation with clinical outcomes. (A) The mRNA expression levels of *KARS1* in organ donor liver tissues (n = 21), TFL and paired HCC tissues(n = 59).(B)The mRNA expression data of *KARS1* from GEPIA database, and red boxplot showed HCC tumor samples (n = 369), black boxplot showed Normal healthy liver samples (n = 160). (C) Univariate and multivariate analysis of associations between clinicopathological risk factors and *KARS1* expression in tumor tissue and HCC-specific patient survival rate and cancer recurrence rate (n=59) in our HCC patient cohort. *P<0.05; **P<0.01; ***P<0.001, by the Mann-Whitney test. TFL: tumor free liver; T:HCC tumor.
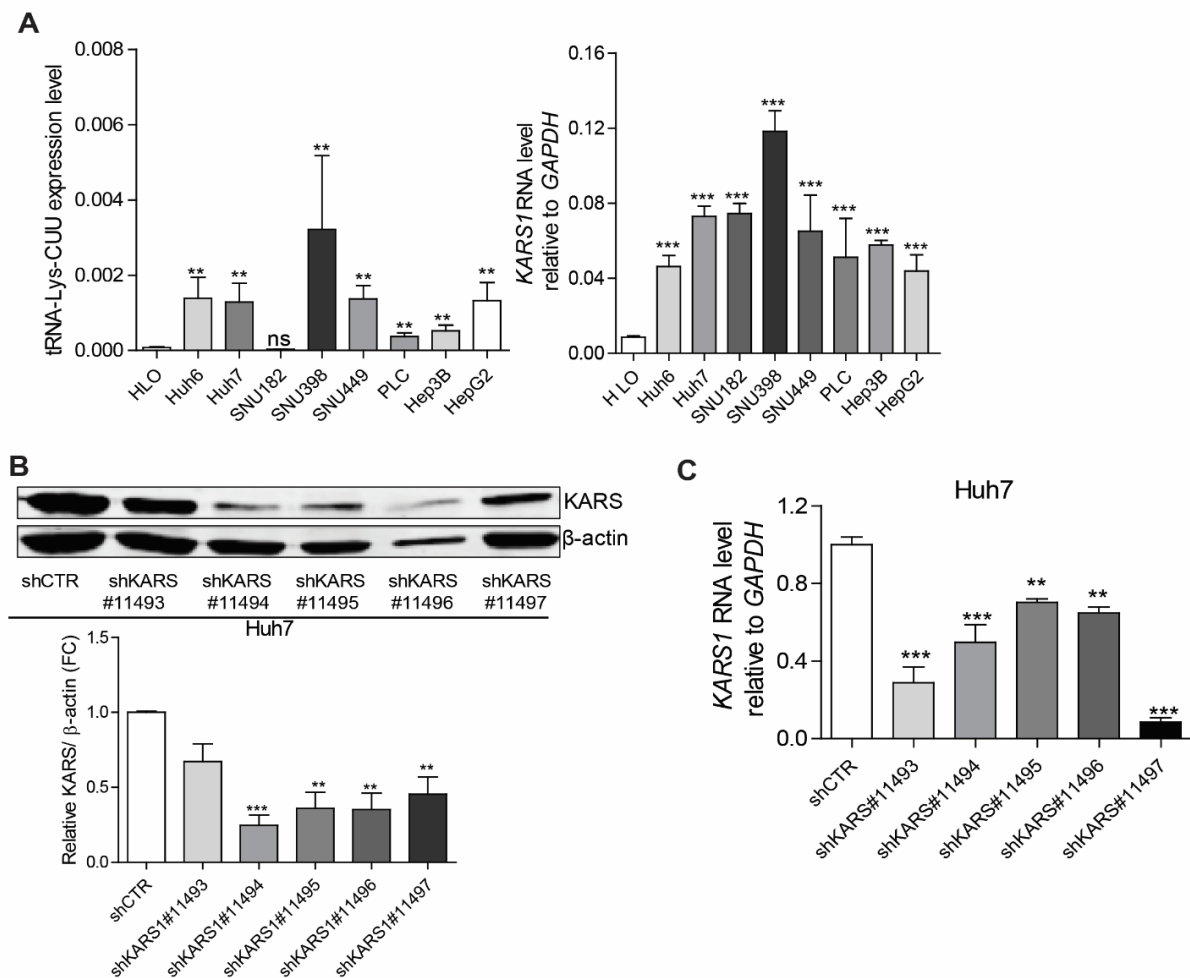
Supplementary Figure 4 The expression level of tRNA-Lys-CUU and *KARS1*, and KARS knockdown efficiency. (A) The expression level of tRNA-Lys-CUU (left graph) and *KARS1* (right graph) respectively, relative to GAPDH in HCC cell lines compared to normal healthy liver organoids (n = 4-7). (B-C) Knockdown in Huh7 and SNU398 was analyzed in (B) by western blot (upper graph), and the intensity was quantified relative to β-actin (lower graph) by image studio software. In (C) the expression of *KARS1* mRNA upon knockdown was determined by qRT-PCR (mean ± SEM, n = 4-9). **P<0.01; ***P<0.001, by Mann-Whitney test. HLO: healthy liver organoids.
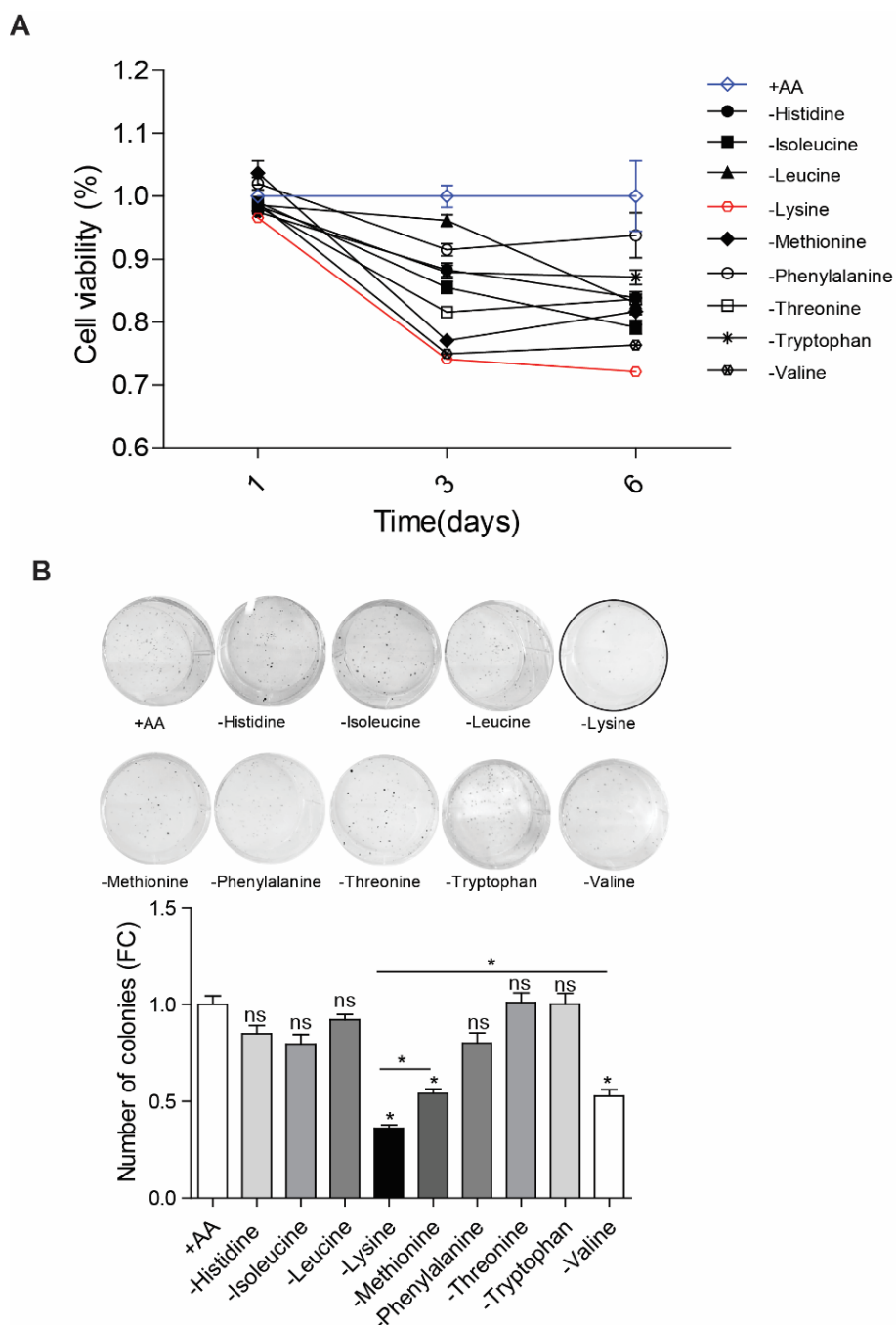
Supplementary Figure 5 Lysine deprivation exerts the strongest inhibitory effects on HCC cell lin growth. (A) The effects of nine essential amino acid deprivation on SNU398 cell growth measured by MTT assay following 1, 3 and 6 days of culture (n = 8). (B) The effects of amino acid deprivation in SNU398 single cell-derived colony formation as assayed 2 weeks after seeding (n = 4). *P<0.05, by the Mann-Whitney test. AA: amino acid; "-" means deprivation.
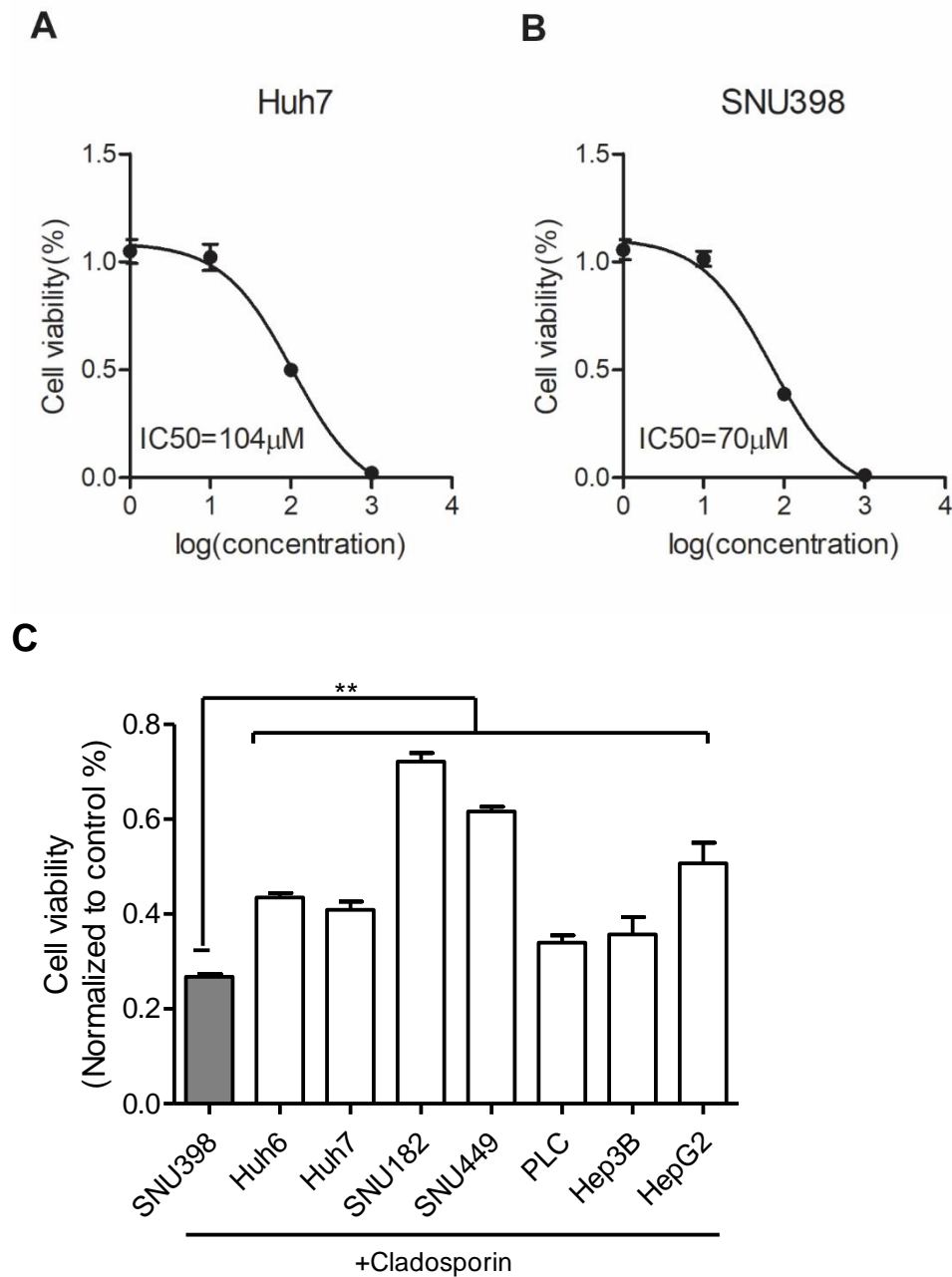
Supplementary Figure 6 The 50% inhibitory concentration of cladosporin was determined by MTT for Huh7 (A) and SNU398 (B) cells following 3 days of treatment. The drug concentrations were 1, 10, 100 and 1000 µM. Data are presented as mean ± SEM.(C) SNU398 cells are more sensitive than other HCC cell lines to treatment of 100 µM cladosporin (n = 6). **P<0.01, by the Mann-Whitney test.
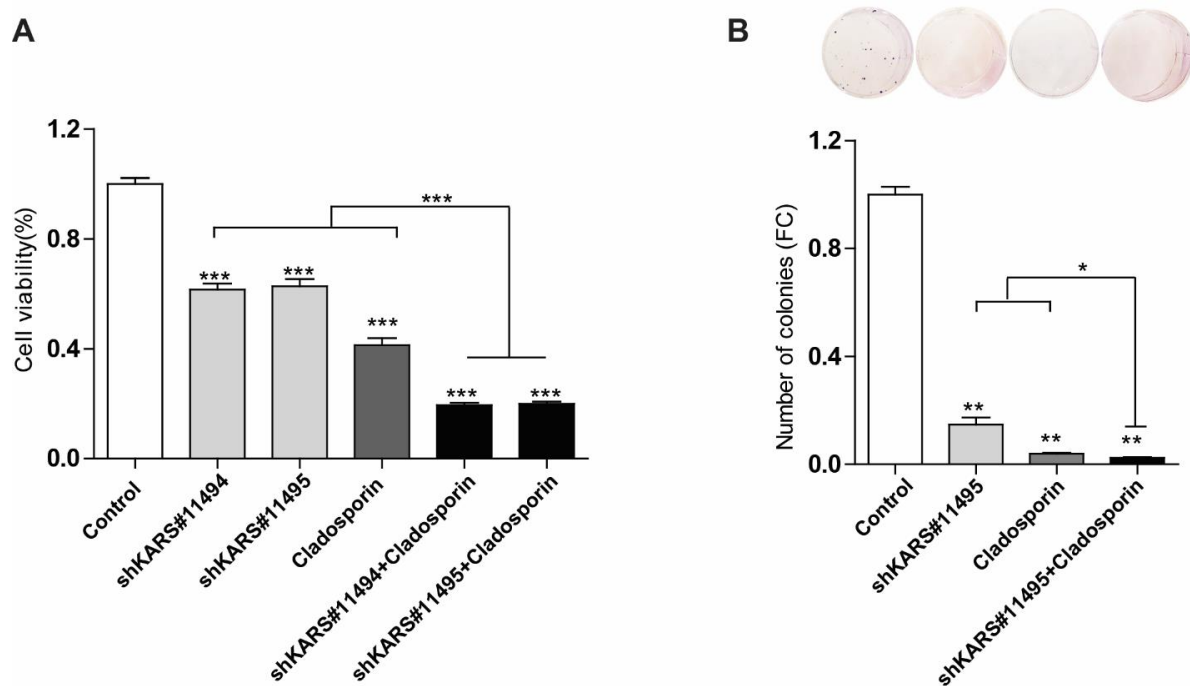
Supplementary Figure 7 The effects of combining cladosporin treatment with KARS silencing in SNU398 cells. Cell viability (A) and single-colony formation (B) were quantified when combining *KARS1* knockdown and cladosporin treatment (n = 4 -16). *P<0.05; **P<0.01; ***P<0.001, by the Mann-Whitney test.
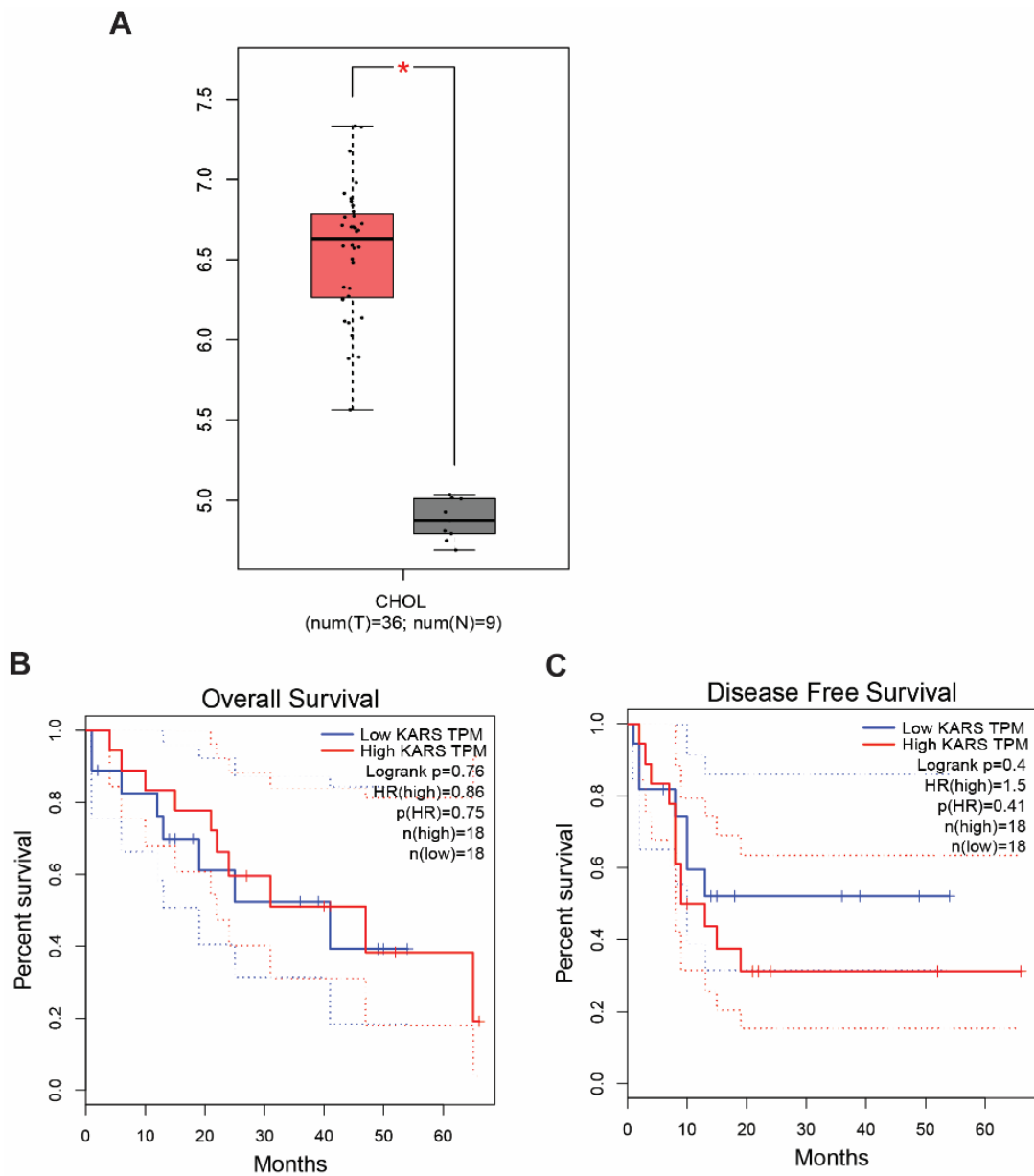
Supplementary Figure 8 The expression level of *KARS1* and association with patient survival in cholangiocarcinoma from the GEPIA database. (A)The boxplot shows expression level of KARS in CHOL (red color, n = 36). (B) and (C) indicated overall suvival and disease free survival in CHOL, respectivity. *P<0.05; CHOL: cholangiocarcinoma.
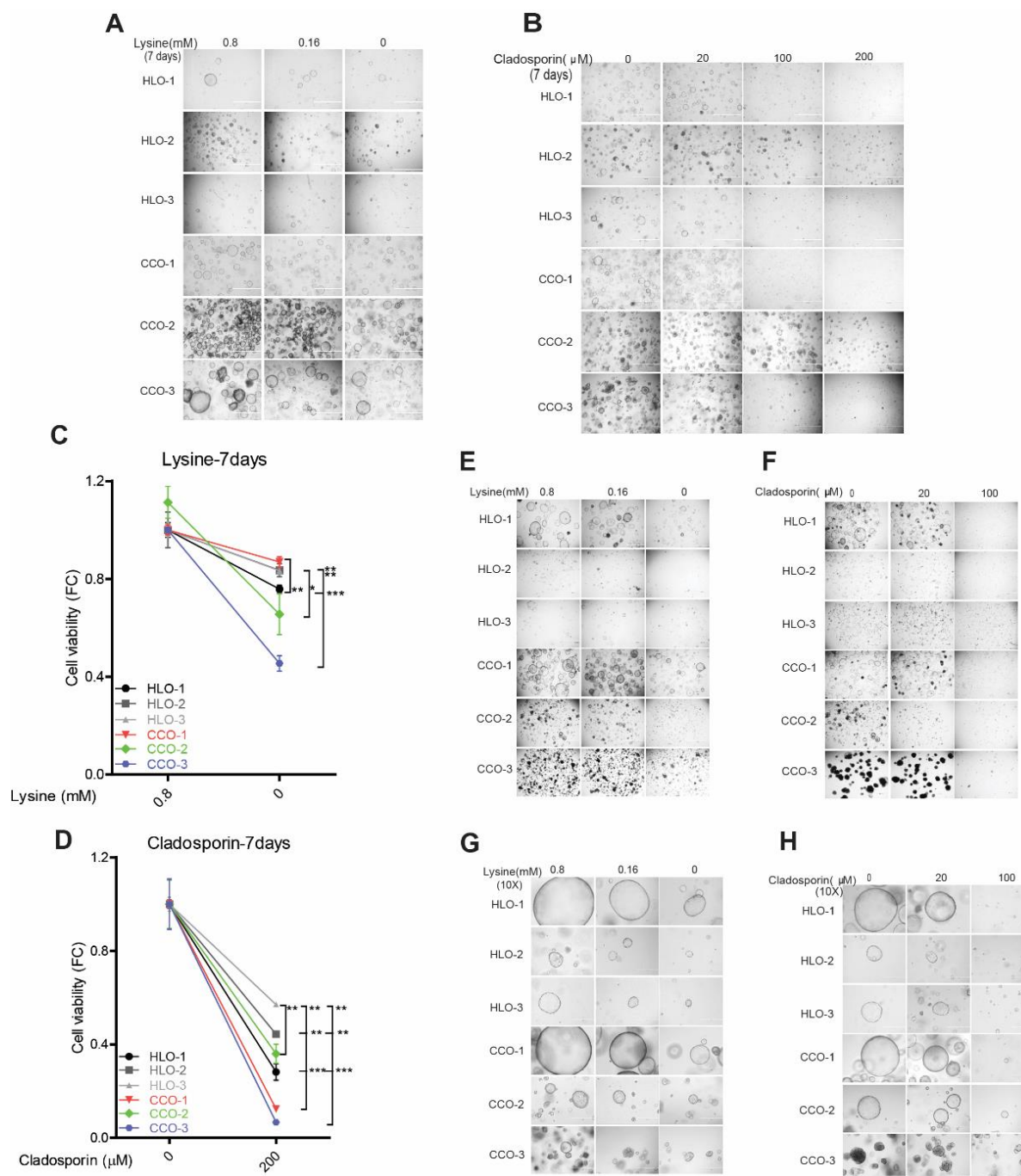
Supplementary Figure 9 Lysine deprivation or treatment with cladosporin affects the growth of patient CC organoids. (A-B)Images of organoids were treated with different concentrations of lysine (A) or cladosporin (B) for 7 days. (C-D) For cell viability of HLOs and CCOs were treated with lysine deprivation (C) or cladosporin 200 μM (D), high expression of *KARS1,* CCO-3 more sensitive. (E-H) Images were taken after 14 days of culture. (G) and (H) show a 10x magnification highlighting the reduced organoid size after prolonged culture with lysine deprivation or cladosporin. CCO: cholangiocarcinoma organoid; HLO: human liver healthy organoid. *P<0.05；**P<0.01; ***P<0.001, by the Mann-Whitney test.

Supplementary Table 1 Patient characteristics		
Characteristic	**HCC patients (n=69)**	**Healthy controls (n=21)**
Age at treatment (years)		
Mean ± SD	60 ± 16	51 ± 16.2
Median (range)	64 (11-82)	52 (13-88)
Sex – no. (%)		
Male	41 (59.4)	12 (57.1)
Female	28 (40.6)	9 (42.9)
Race – no. (%)		
White	58 (84.1)	-
African	6 (8.7)	-
Asian	4 (5.8)	-
Not reported	1 (1.4)	21 (100)
Etiology – no. (%)		
No known liver disease	20 (29.0)	NA
Alcohol	16 (23.2)	NA
NASH	8 (11.6)	NA
Hepatitis B	8 (11.6)	NA
Hepatitis C + Alcohol	5 (7.2)	NA
Hepatitis B + Alc/HepC/HepD/NASH	4 (5.8)	NA
Hepatitis C	4 (5.8)	NA
Fibrolamellar HCC	3 (4.4)	NA
Other	1 (1.5)	NA
Hepatitis status – no. (%)		
Hepatitis B or C positive	21 (30.4)	0 (0)
Chronic Hepatitis B	12 (17.4)	0 (0)
Chronic Hepatitis C	10 (14.5)	0 (0)
Cirrhosis – no. (%)		
Yes	50 (72.5)	0 (0)
No	19 (27.5)	21 (100)
Differentiation grade – no. (%)		
Good	8 (11.6)	NA
Moderate	35 (50.7)	NA
Poor	14 (20.3)	NA
Unknown	12 (17.4)	NA
Number of lesions – no. (%)		
1	38 (55.1)	NA

>1	31 (44.9)	NA
Median (range)	1 (1-11)	NA
Size of largest lesion (cm)		
Mean ± SD	7.5 ± 5.1	NA
Median (range)	6.2 (1-24)	NA
AFP level before resection (ug/l)		
Mean ± SD	46679.2 ± 375310.5	NA
Median (range)	9 (1-3118700)	NA

Supplementary Table 2 shRNA target sequences of lysyl-tRNA synthetase.

targetSeq	Oligoseq	Gene ID	Clone name	Vector
CGGCGAATCAA CATGGTAGAA	CCGGCGGCGAATCAACATGGTAGAACTCG AGTTCTACCATGTTGATTCGCCGTTTTTG	NM_005548.1	1214s1c1	pLKO.1
GCCTTTCATCAC TTATCACAA	CCGGGCCTTTCATCACTTATCACAACTCGA GTTGTGATAAGTGATGAAAGGCTTTTTG	NM_005548.1	883s1c1	pLKO.1
CCTGGAAGTGA CTTGCATCAA	CCGGCCTGGAAGTGACTTGCATCAACTCGA GTAGTATTGATTTGGGTCCACGTTTTTG	NM_005548.1	1393s1c1	pLKO.1
CGTGGACCCAAT CAA4TACTA	CCGGCGTGGACCCAAATCAATACTACTCGA GTAGTATTGATTTGGGTCCACGTTTTTG	NM_005548.1	250s1c1	pLKO.1
CCAGAGATACTT GGACTTGAT	CCGGCCAGAGATACTTGGACTTGATCTCGA GATCAAGTCCAAGTATCTCTGGTTTTTG	NM_005548.1	724s1c1	pLKO.1

Supplementary Table 3 Primer sequences

Genes	Forward primers (5'-3')	Reverse primer (5'-3')
KARS1	GACGCACAATCCTGAGTTCACC	AGGTGACCTTGTAACTGCCTGTA
GUSB	CAGGTGATGGAAGAAGTGG	GTTGCTCACAAAGGTCACAG
HPRT1	GCTATAAATTCTTTGCTGACCTGCTG	AATTACTTTTATGTCCCCTGTTGACTGG
PMM1	CGAGTTCTCCGAACTGGAC	CTGTTTTCAGGGCTTCCAC
GAPDH	GTCTCCTCTGACTTCAACAGCG	ACCACCCTGTTGCTGTAGCCAA

Supplementary Table 4 List of tRNA expression comparison Differences of tRNA expression between tumor and TFL patient tissues (n=69) were analyzed using the Wilcoxon matched pairs test. *P<0.05, **P<0.01, ***P<0.001. TFL:tumor-free liver. Red: signficant up-regulation in tumor; Green: signficant down-regulation in tumor; Black: no signficant difference.

Gene	Mean		Wilcoxon matched pairs test
	TFL	Tumor	
tRNA-Leu-CAA	1.522	1.277	***
tRNA-Leu-CAG	5.106	3.551	***
tRNA-Ile-AAU	1.37	1.254	***
tRNA-Ile-GAU	1.076	0.8677	***
tRNA-iMet-CAU	13.49	12.49	***
tRNA-Met-CAU	6.762	5.325	***
tRNA-Stop-UUA	5.32	4.172	***
tRNA-His-GUG	16.01	13.99	***
tRNA-Phe-GAA	1.095	0.801	***
tRNA-Val-AAC	1.8	0.6382	***
tRNA-Val-CAC	2.248	0.7724	***
tRNA-Pro-GGG	0.6167	0.4805	***
tRNA-Pro-CGG	4.063	3.23	***
tRNA-Thr-AGU	0.748	0.5531	***
tRNA-Thr-UGU	2.64	1.922	***
tRNA-Thr-CGU	1.107	0.7374	***
tRNA-Ala-CGC	10.14	4.745	***
tRNA-Ala-GGC	3.448	1.932	***
tRNA-Ala-UGC	8.042	4.49	***
tRNA-Ala-AGC	1.779	1.098	***
tRNA-Tyr-AUA	1.429	1.227	***
tRNA-Tyr-GUA	5.609	2.763	***
tRNA-Stop-CUA	5.562	2.651	***
tRNA-Asn-AUU	2.535	1.669	***
tRNA-Lys-UUU	18.79	9.738	***
tRNA-Lys-CUU	**6.867**	**3.389**	***
tRNA-Asp-AUC	4.942	2.412	***
tRNA-Asp-GUC	0.8559	0.6074	***
tRNA-Glu-UUC	0.9868	0.753	***
tRNA-Cys-GCA	4.178	3.506	***
tRNA-Arg-ACG	0.8352	0.6351	***
tRNA-Arg-CCG	2.109	1.783	***
tRNA-Ser-ACU	1.497	1.283	***
tRNA-seCys-UCA	3.942	3.471	***
tRNA-Val-UAC	1.783	0.4871	**
tRNA-Ser-CGA	5.424	5.389	**
tRNA-Pro-AGG	1.317	1.246	**
tRNA-Gln-CUG	18.07	12.53	**
tRNA-Cys-ACA	0.8275	0.4544	**
tRNA-Arg-UCG	0.682	0.5212	**
tRNA-Ile-UAU	2.046	1.529	ns
tRNA-Gln-UUG	18.85	14.06	ns
tRNA-Asn-GUU	0.4828	0.3918	ns
tRNA-Arg-UCU	0.7006	0.6279	ns
tRNA-Gly-CCC	0.4193	0.3827	ns

Gene	Mean		Wilcoxon matched pairs test
	TFL	Tumor	
tRNA-Gly-GCC	197.2	603.3	***
tRNA-Ser-AGA	2.547	2.894	***
tRNA-Leu-AAG	6.162	8.093	**
tRNA-Leu-UAG	8.082	8.548	**
tRNA-Pro-UGG	3.571	3.923	**
tRNA-Trp-CCA	4.311	5.186	**
tRNA-Glu-CUC	2.303	3.733	*
tRNA-Gly-UCC	1.371	3.731	*
tRNA-Leu-UAA	4.735	5.492	ns
tRNA-Ser-UGA	4.997	5.848	ns
tRNA-Ser-GCU	3.642	4.028	ns
tRNA-Arg-CCU	1.474	1.573	ns

PART II

The implications of translational decoding in viral cross-species infection

Chapter 6

The neglected avian hepatotropic virus induces acute and chronic hepatitis in ducks:

an alternative model for hepatology

Xumin Ou, Sai Mao, Jingyu Cao, Yunchao Ma, Guangpeng Ma, Anchun Cheng, Mingshu Wang, Dekang Zhu, Shun Chen, Renyong Jia, Mafeng Liu, Kunfeng Sun, Qiao Yang, Ying Wu, Xiaoyue Chen

Abstract

Duck Hepatitis A Virus (DHAV) belongs to the Avihepatovirus, which is also classified into Picornaviridae with Hepatovirus, Hepatitis A Virus (HAV). In humans, the pathogenesis of HAV is not well understood because of limited work with animal models. Here, we investigated the progress of duck viral hepatitis caused by DHAV and their potential for dissecting the pathogenesis of HAV. During the course of infection, the duck model had undergone hepatocellular lesions (vacuolation, acidophilic degeneration and steatosis), lymphocytes recruitment (neutrophil granulocytes, heterophilic granulocytes and T cells or plasm cells) and repair (activation of hepatic stellate cells, fibrosis and regeneration). Coincident with liver injury, the serum biomarkers, aspartate aminotransferase and alanine transaminase were significantly increased. Moreover, comparatively lower CD4+ and CD8+ T-cells were recruited to the liver, which might lead to a persistent infection (40 wk). Because DHAV and HAV have similar genomic structure, biological phenotypes and can easily replicate in liver. And half of fibrosis-related genes had high homology between humans and ducks. Considering these similarities in pathological and virological phenotypes, we proposed that the ducks might be an alternatively small animal model that would provide insight into the pathogenesis of viral hepatitis, fibrosis and liver regeneration.

Keywords: Duck Hepatitis A virus; viral hepatitis; liver injury; pathogenesis; viral-host interaction

Introduction

Hepatitis A is a worldwide human liver disease caused by Hepatitis A Virus (HAV), especially in areas of lower socio-economic status and reduced sanitary conditions. HAV, a member of the Hepatovirus genus in the family of Picornaviridae, has an estimated 1.4 million cases occur each year [1]. However, there is a similar liver disease in ducks caused by Duck Hepatitis A virus (DHAV) [2, 3]. DHAV, a member of the Avihepatovirus genus within the same family Picornaviridae, causes a highly fatal, rapidly infectious disease in ducklings, characterized by swelling livers mottled with haemorrhages [3-6]. This type of viral hepatitis in ducklings should be classified as acute viral hepatitis, but the progressive pathogenesis caused by DHAV in mature ducks has been largely unexplored.

Chronic Viral Hepatitis (CVH) is mainly caused by Hepatitis B Virus (HBV), Hepatitis C Virus (HCV), and Hepatitis D Virus (HDV), but humans infected with HAV do not progress to CVH [7-10]. Briefly, CVH is transformed from Acute Viral Hepatitis (AVH) and is related to inflammatory responses, parenchymal cell necrosis or apoptosis and regeneration. Many human and animal model studies have revealed a correlation between inflammatory responses and fibrosis, which is caused by proliferation of hepatic stellate cells (HSCs) [11, 12]. Additionally, approximately 5%-10% of adult patients with CVH may progress to liver cirrhosis [13]. Hepatocytes are the major sources of matrix metalloproteinases (MMP), and their inhibitors are involved in liver extracellular cell matrix (ECM) deposition, which is involved in the pathogenesis of liver cirrhosis in CCL4-induced liver cirrhosis in rats [14]. The endothelial fenestrae is 150-175 nm in diameter and acts as a dynamic filter facilitating the exchange of nutrition and particles between parenchymal cells and the hepatic sinusoid. Many hepatitis viruses could result in defenestration and a decrease in the number of fenestrae [15]. Apoptosis or necrosis is the most common consequence of hepatocellular injury, followed by KCs engulfing apoptotic bodies and death ligands, including the Fas ligand and TNF-α, thereby promoting inflammation and fibrogenesis [16, 17]. Damaged hepatocytes will then induce activation of HSCs to stimulate the fibrogenic actions through release of reactive oxygen species (ROS) and fibrogenic mediators [18]. Hepatocyte apoptosis was also associated with expression of death receptors and activation of NF-κB in patients with steatohepatitis [19]. In addition, plentiful cytokines were involved in regulation of liver fibrosis, and most of them were widely used to attenuate the apoptotic and fibrotic effects [9, 15, 18, 20]. IFN-α was shown to reduce apoptotic effects on activated HSCs, and HSCs were also inactivated by IFN-β and then decreased the synthesis of α-smooth muscle actin (SMA) and collagen through inhibition of the TGF-β and PDGF pathways [21]. Moreover, the activation of HSCs is also inhibited by IFN-γ to decrease ECM deposition through inhibiting TGFβ1/Smad3 signalling pathways [22, 23]. On the contrary, interleukin 1 (IL-1), a proinflammatory cytokine, produced

by KCs and sinusoidal endothelial cells (SECs) can activate HSCs to produce MMP-9, MMP-13 and TIMP-1, resulting in liver fibrogenesis [11]. IL-6 can attenuate hepatocyte apoptosis and promote their regeneration through NF-κB signalling and the Ras-MAPK pathway [24, 25].

Due to the crucial role of the animal model in understanding the pathogenesis and the development of therapeutic strategies for liver injury, fibrosis and cirrhosis, many types of mammal models have been developed in mice, rats, chimpanzees, rabbits, pigs and horses to mimic the process of liver injury [26-28]. However, owing to the proposed listing of the endangered primates and the inability to include them in research, optional duck models are essential, much more accessible, easier for laboratory management and also available for serial sampling of tissue in the volume required for detailed studies [29]. Previously, ducks have been widely used in Hepatitis B Virus preclinical trials and to elucidate their life cycle [30, 31]. Duck Hepatitis B Virus (DHBV) infection cannot induce apparent liver injury and will process to hepatocellular carcinoma (HCC) [32]. However, avian hepatotropic virus (DHAV) can cause obvious liver disease characterized by swelling livers mottled with haemorrhages in ducklings. In this study, the course of hepatitis and host immune responses in mature ducks induced by DHAV were investigated to mimic the process of liver injury, fibrosis and regeneration.

Materials and methods

Ethics Statement

The animal study was also performed in strict accordance with the recommendations in the ARRIVE guidelines (http://www.nc3rs.org.uk/arrive-guidelines). The animal experiment has been approved by the committee of experiment operational guidelines and animal welfare of Sichuan Agricultural University, China (the approved permit number is XF2014-18). All ducks were handled in compliance with the animal welfare regulations and maintained according to standard protocols.

Virus and primary duck embryo liver cell preparation

The DHAV-H strain (GenBank: JQ301467.1) was propagated in 9- to 11-day-old duck embryos by standard procedures. The embryos that died at 36-72 hpi were harvested. The homogenate of the allantoic fluid of duck embryos was stored at -80°C until use. The virus, at a concentration of 4.56×10^8 copies/ml as determined by quantitative real-time PCR (qPCR), was used to infect ducks [48].

Duck embryo liver cells were prepared using 9- to 11-day-old duck embryonated eggs according to definitions within the program. The dispersed cells were dispensed into flat-bottomed, 12-well culture plastic plates (Corning Inc., New York, USA) with cover slips. The cells were cultured in Dulbecco's modified Eagle's medium with 10 % foetal bovine serum.

Then, the cells were maintained in another medium containing 3 % foetal bovine serum. Cells were maintained at 37°C with 5 % CO2.

Experimental design

One hundred and five female ducks were randomly divided into 21 groups. Group 1 to group 20 received 1 ml of DHAV-H strain (Genebank: JQ301467) (4.56×10^8 copies/ml) by intramuscular injection, and group 21 was injected with an equal volume of 0.85% physiological saline as a negative control. The livers of 20 experimental groups were collected according to the time post-infection, including 1/2d, 1d, 2d, 4d, 6d, 8d, 10d, 12d, 14d, 21d, 28d, 56d, 84d, 112d, 140d, 168d, 196d, 224d, 252d, and 280d. One hundred milligrams of liver specimens were weighed and then immediately cryopreserved in liquid nitrogen until processing for RNA isolation. The adjacent tissues were used for histopathological examination and immunohistochemistry.

To identify the hepatocellular lesions caused by DHAV, duck embryos were also used to address this question. Twelve 6-day-old duck embryos were randomly divided into 4 groups. Group 1 to group 3 received 0.2 ml of virus (4.56×10^8 copies/ml) at the allantoic cavity, and group 4 was injected with an equal volume of 0.85% physiological saline as a negative control.

The duck embryo liver cells were cultured in the 12-well culture plastic plates divided into 6 groups. Group 2 to group 6 received 0.1 ml of virus (4.56×10^8 copies/ml), and group 1 received an equal volume of 0.85% physiological saline as a negative control.

qPCR

Total cellular RNA was isolated from 100 mg of liver tissue using the RNAiso plus Reagent (TaKaRa, Japan) according to the manufacturer's protocols. The RNA isolated from each specimen needed to detect immune-related genes was used to carry out reverse transcription to produce cDNA with the PrimeScriptTM RT Reagent Kit according to the manufacturer's instructions (TaKaRa, Janpan). Viral copies were detected by previously established methods in our laboratory [48]. Seventeen immune-related genes (IL-1β, IL-2, IL-4, IL-6, IFN-α, IFN-β, IFN-γ, MHC-I, MHC-II, CCL19, CCL-21, BAFF, TLR3, TLR7, β-defensin, RIG-1 and MDA5) and a housekeeping gene (glyceraldehyde-3-phosphate dehydrogenase (GAPDH)) were also detected according to the previously established protocols [33].

HE and Masson staining

After administering sodium pentobarbital anaesthesia, the livers from the same samples used for transcriptional analysis were fixed in 4% paraformaldehyde, dehydrated, embedded in paraffin, sectioned into 4-μm-thick sections and stained with haematoxylin and eosin (HE) and Masson using standard procedures. Additionally, the duck embryos and duck embryo liver

cells infected with the DHAV-H strain were fixed in 4% paraformaldehyde first, and then their liver tissues were also used for HE staining using standard procedures.

Flow Cytometry Analysis

To validate the identity of CD4+ and CD8a+ positive staining, the liver tissues used for IHC assays were also detected by flow cytometry (BD FACS VerseTM, San Jose, CA, USA). Uninfected liver tissues were used as a negative control. Briefly, the positive cells were incubated with mouse anti-duck CD4 or CD8a monoclonal antibody (8μg/ml) followed by FITC-coupled goat anti mouse secondary antibody solutions (1:500 dilution) (Beyotime Biotechnology). Then, the percentage of FITC-positive cells were determined by flow cytometry.

Immunohistochemistry

Paraffin-embedded liver tissues were deparaffinized in xylene and rehydrated in graded alcohols. Briefly, the standard protocol were previously established for detecting viral capsid and 3D polymerase and also double IHC staining for detecting viral capsid and CD4+ or CD8+ T cells.[33].

Indirect immunofluorescence

Following the 12 h, 24 h, 36 h, 48 h and 60 h incubation at 37°C after infection of the DHAV-H strain (100 μL), the cells were washed with phosphate buffered saline solution (PBS; 0.15 mol/L, pH 7.2) and fixed for 30 min at room temperature (RT) with 4% paraformaldehyde. After fixation, the paraformaldehyde was removed and the cells were permeabilized with 0.5% Triton X-100 for 10 min at RT and incubated in blocking buffer (PBS, 5% BSA) for 30 min at RT. The primary rabbit anti DHAV polyclonal antibody was diluted in 0.5% BSA and incubated with cells for 1 h at RT. After being washed with PBS, the coverslips were incubated with goat anti-rabbit Alexa Fluor® 488 conjugate (Life technologies, USA) diluted in 0.5% BSA for 1 h at RT. After counterstaining, coverslips were washed three times with PBS, mounted with fluorescence mounting media (KPL, Gaithersburg, MD), and examined with a Nikon Microscope. We used cells infected with PBS as a negative control.

Phylogenetic analysis of DHAV and HAV and the fibrosis-related genes in humans and ducks

The DHAV-related genus in the family of Picornavirdae was used to identify the neighbour genus. The sequences of fibrosis-related genes in human and duck were also used to compare their evolutionary relationship. The phylogenetic tree was constructed by Mega 6.0 with the neighbour-joining (NJ) method and 1000 bootstrap analysis [49]. The members of picornavirus

and fibrosis-related genes in ducks and humans used for phylogenetic analysis are listed in Table S3.

Statistical analyses

Relative gene expression data were analyzed using the 2-ΔΔCt method by comparing with the control group injected with 1 ml normal saline (NS) [50], and ΔCt values were determined by subtracting average Ct values of the endogenous control gene GAPDH from average Ct values of target genes. The photographs were generated using GraphPad Prism 5 software. In order to understand the impact of virulence on immune networks, the correlations of each pair of immune related genes were calculated using correlation analysis (Pearson) using SPSS software. Those correlated pairs of immune related genes were visualized as immune network by Cytoscape software. The significant levels in this study determined using Student's t test (Two-tail, α=0.05).

RESULTS

Post-infection with DHAV in mature ducks was associated with acute and chronic hepatitis

Liver tissues from infected breeding ducks had developed apparent liver injury with acute hepatitis at early infection and chronic hepatitis at later infection characterized by histopathological changes. At the early stage of infection, liver lesions were characterized by acidophilic degeneration, vacuolation and infiltration of neutrophil granulocytes (**Figure 1**). The liver had also developed apparent serious steatosis at 28-dpi (**Figure 1**). However, at the later stage of infection, acidophilic degeneration disappeared instead of serious vacuolation (ballooning degeneration). The intensity of steatosis, interface hepatitis and neutrophil granulocytes were temporarily elevated and gradually decreased in the last stage of infection. Of note, the changes of knodell Histological Activity Index (HAI) were highly coincident with the severity of liver injury (**Table 1**). Many more heterophilic granulocytes and plasma cells were strongly recruited to the liver until the occurrence of liver cell regeneration. Accompanying serious liver injury and inflammatory response, the fibrosis secreted by fibroblasts also took place from 10-dpi to the last stage of infection to make up the hepatocellular lesions (**Figure S1**).

Table 1: The hepatitis activity during the course of infection.

Phases / Factors	Early hepatocellular injury						Later hepatocellular injury					
	Con	2d	6d	10d	21d	28d	56d	112d	140d	168d	224d	280d
I, Periportal +/-bridging necrosis	0	1	1	1	1	1	3	4	5	4	5	1
II, Intralobular degeneration and focal necrosis	0	1	1	3	4	4	4	4	4	4	4	1
III, Portal inflammation	0	1	3	3	3	3	4	4	4	3	4	1
IV, Fibrosis	0	0	0	1	1	3	1	1	1	3	4	4
Total scores	0	3	5	7	9	11	12	13	14	14	17	7

The hepatitis activity scored by knodell Histological Activity Index (HAI) score system. The total score was calculated by four indexes, such as periportal +/- bridging necrosis, intralobular degeneration and focal necrosis, portal inflammation and fibrosis. The HAI scores were assessed by three diagnose pathologists.

Additionally, the biochemical analysis of serum from infected ducks indicated that serum aspartate aminotransferase (AST), alanine transaminase (ALT) and total bilirubin (T-Bil) were increased and decreased successively in the course of infection (P< 0.05) (**Figure 2**). Moreover, the highly elevation of triacylglycerol levels at 28dpi was accord with serious steatosis identified by HE staining (**Figure 1-2**). The significantly decreased albumin indicated the liver protein synthesis was seriously impaired during the infection. Simultaneously, the increased globulin indicated humoral immunity might not be impaired by DHAV (**Figure 2**).

Comparatively lower T-cells in liver was associated with chronic infection.

T-cell immunity is crucial for viral clearance and also humoral immunity. At early stage of infection, the numbers of CD4+ T -cells (Ths) and CD8+ T-cells (Tcs) significantly increased at 10dpi (**Figure 3A/B/C**). After that, both of them vanished from 28-dpi to 140-dpi and reoccurred at later stage of infection (**Figure 3A/B/C**). The numbers between Tcs and Ths at each interval were no significance (P>0.05) (**Figure 3D**). In addition, immunohistochemistry assays indicated that both Tcs and helper T-cells Ths were located at the portal area of liver (**Figure S2/S3**). In consistent with these observation, the dynamic changes in viral capsid load and viral RNA polymerases were also gradually increased and decreased during the later stage of infection (**Figure S4/S5**) and also consistent with the decline of viral RNA copies (**Figure 4A**). These results indicated that recruitment of the T-cells in liver was associated with viral clearance. However, the comparatively lower CD4+ and CD8+ T-cells and persistent disappearance of them might lead to a chronic infection (40wk).
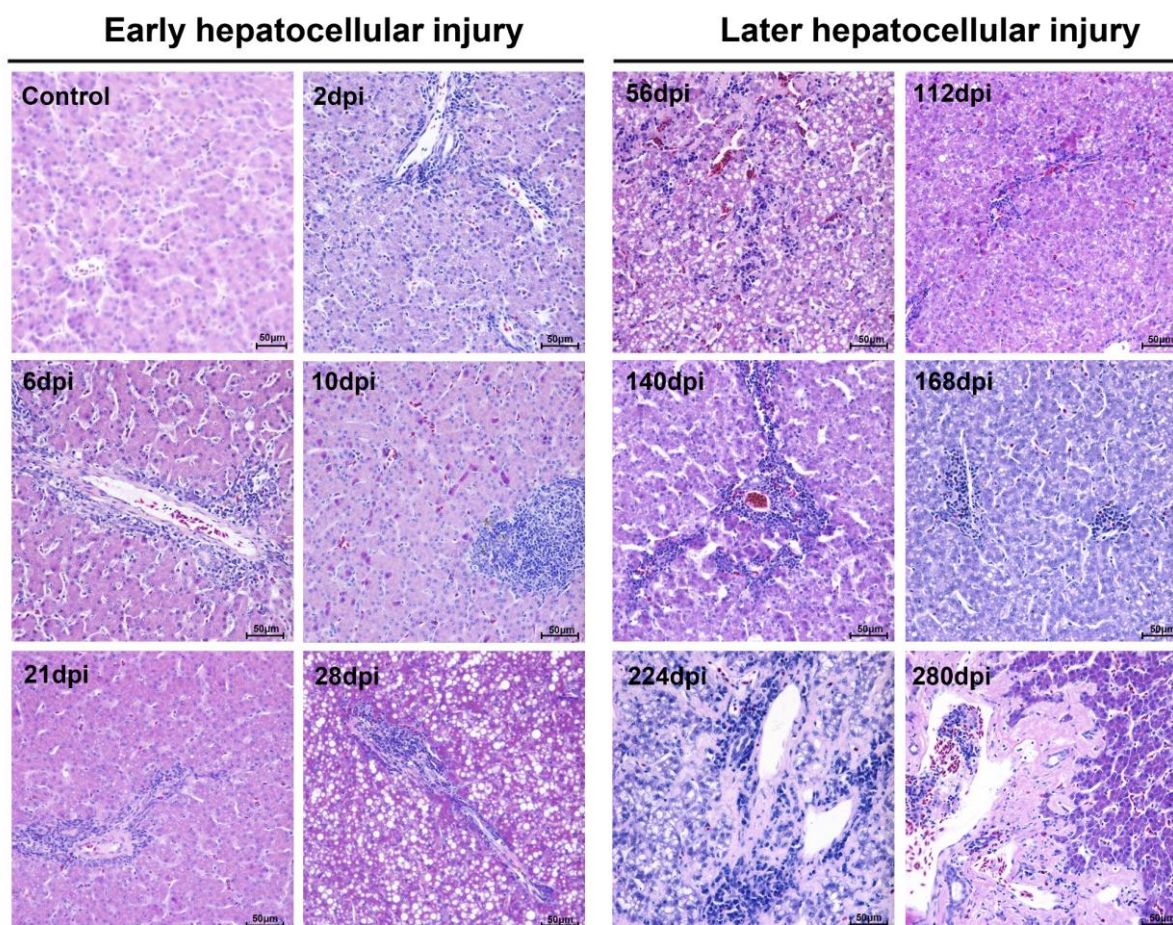
Figure 1: Microscopic Lesions in mature ducks experimentally infected with DHAV-1 H strain. Representative liver HE staining from 1d to 280d post-infection with DHAV are displayed from two stages, early infection and later infection (n=5). (Con) represents pre-inoculation liver photomicrograph. The brightness and contrast are slightly modified to create a uniform background.
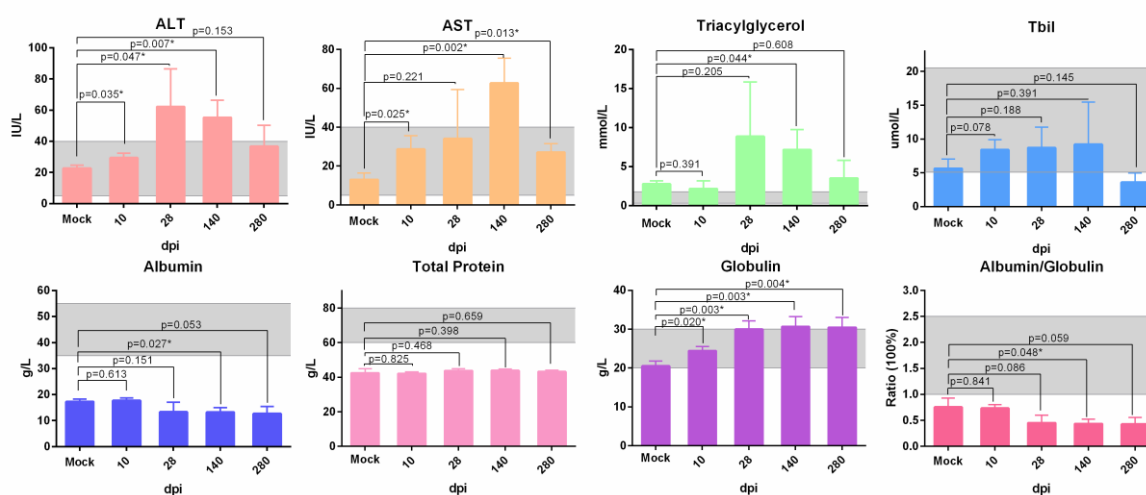
Figure 2: Biochemical detection of serum from ducks infected with DHVA-H. Serum biochemical markers, ALT, AST, triglyceride, total Bilrubin, total protein, albumin, globulin and

ratio of albumin and globulin, were analysed by auto biochemical detector. The scope between two dash lines were the normal biochemical reference range in human. Data are presented as means ± SD (n=3). *P<0.05 indicates significant difference using student's T-test.
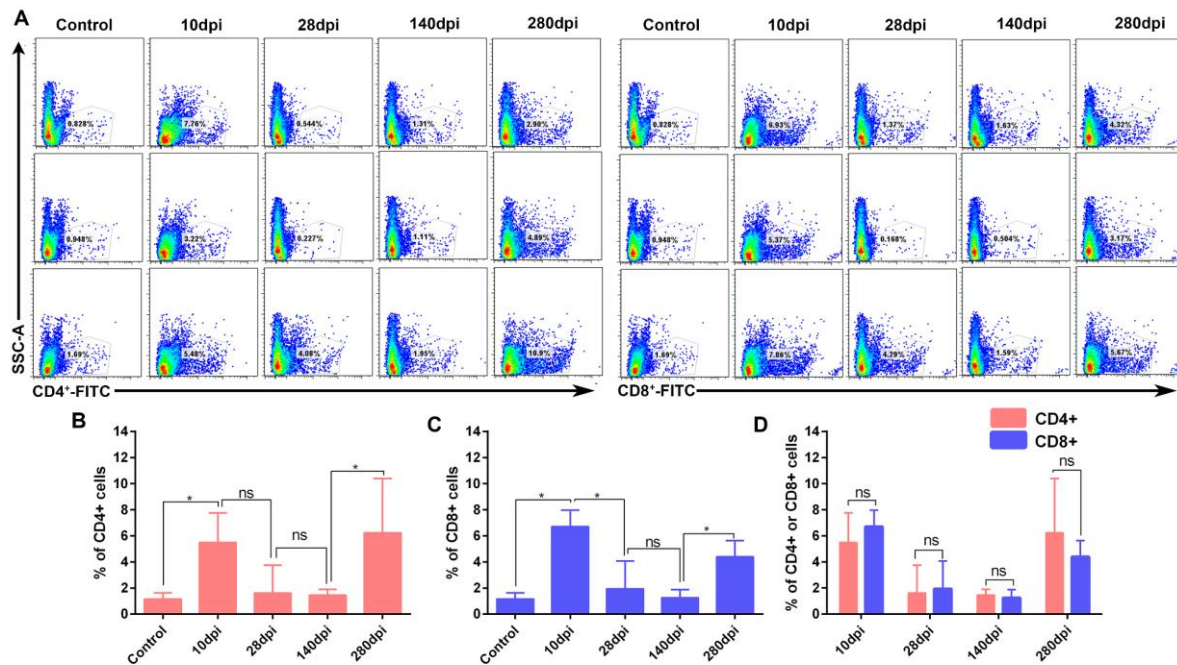
Figure 3: Flow cytometry analysis of CD4+ and CD8+ T-cells in liver (n=3). The percentage of CD4+ and CD8+ T-cells were detected by FITC-conjugated goat anti mouse IgG. A. Flow cytometry analysis of CD4+ and CD8+ T-cells from 10 dpi to 280dpi. B. Significant difference analysis of positive CD4+ T cells in the course of infection. C. Significant difference analysis of positive CD8+ T cells in the course of infection. D. Significant difference analysis of between positive CD4+ and CD8+ T cells in the course of infection. Data are presented as means ± SD (n=3). *P<0.05 indicates significant difference using student's T-test. ns, no significance.

Cytokine storms might be also associated with viral clearance

Humoral and cellular immunity are strongly modulated by immune-related cytokines. In this study, the interferons (IFN-α/β/γ) and interleukins (IL-1β/2/4/6) were highly up-regulated from 2-dpi to 6-dpi (**Figure 4A**). Interestingly, the period of cytokine storms was similar to that in kidney infected with DHAV-1 [33]. Simultaneously, viral copies in liver were decreased under the course of cytokine storms (**Figure 4A**). Subsequently, the cytokine storms were transiently quitted down at 10-dpi that might be caused by newly viral replications (**Figure S5**). Consistently, the viral copies were significantly increased after the cytokine storms (**Figure 4A**). Additionally, Pearson correlation analysis indicated that the changes of interferons and interleukins were negatively correlated with the viral RNA copies, especially for IFN-γ (P<0.05) (**Figure 4B**). In this study, TLR7 was highly upregulated from 2-dpi to 6-dpi, whereas other members of the Pattern Recognition Receptors (PRRs), TLR3, MDA5 and RIG-1, were slightly upregulated from 8-dpi to 14-dpi. Those observations indicated that TLR7 might be involved in the early cytokine storms and the other PRRs might be involved in the later cytokine expressions. Comparative analysis indicated that IL-4 and IFN-γ were the top two highly up-

regulated cytokines from 12-dpi to 28-dpi (**Figure S6**). While, the expression patterns of the Major Histocompatibility Complex (MHC) and chemokines were quite different. MHC-I was slightly upregulated from the time of infection to 6-dpi. CCL21 but not CCL19 was strongly elevated from 1-dpi to 6-dpi and disappeared from 8-dpi to 28-dpi. Additionally, BAFF, a member of the tumour necrosis factor (TNF) family, was temporarily upregulated at 8-dpi. The expression pattern of β-defensin was highly similar to that of interferons and interleukins (**Figure 4A**).

In order to understand their relationship, Pearson correlation analysis was used in this study. 27 pairs of genes were highly correlated and enriched in IFNs and interleukins (P<0.05 at least) (**Figure 4B**). And we also identified that TLR7 was the top one PRRs correlated with both IFNs and interleukins (P<0.05) (**Figure 4C**). Those results indicated that the activation of TLR7 pathway might be involved in early cytokine storms and viral clearance.

Serious liver lesions in duck embryos were caused by infection of DHAV

The duck embryos infected with DHAV progressed to liver lesion from apoptosis, cellular swelling to vacuolar degeneration (ballooning degeneration) and necrosis (**Figure 5A**). The normal liver cells had no histopathological change with many more mitotic cells. With prolonged time, then, the hepatocellular lesions were characterized by ballooning degeneration, and many more heterophilic granulocytes and neutrophile granulocytes were recruited to the liver at 8-hpi. However, they swelled, became round and even progressed to apoptosis after infection at 24-hpi. At the last stage of infection, all of the liver cells were characterized by necrosis, and plentiful neutrophil granulocytes were recruited to the liver. The histopathological changes were also accompanied with rapidly viral reproduction (**Figure 5B/C**). Of note, the expression levels of viral polymerase (3D protein) was apparently higher in those serious infected cells than adjacent cells (**Figure 5D/E**).

Activation of primary duck embryo HSCs was associated with hepatocellular lesions

Cultured cells used for certain virus infections provided the potential for researchers to address the pathogenesis caused by the virus in vivo. Fortunately, primary duck embryo liver cells were available for DHAV infection. Meanwhile, infection of primary duck embryo liver cells could also induce activation of HSCs. As is shown in **Figure 6**, morphogenesis of quiet HSCs was significantly transformed to an activated type after hepatocellular lesion caused by DHAV. This process was associated with virus infection such that virions were adsorbed at the cell membrane at 12-hpi and then produced much more progeny virus in the cytoplasm from 36-hpi to 60-hpi. Additionally, virus replication was also accompanied by serious cytopathic changes characterized by activation of HSCs and disappeared hepatocellular islands (**Figure 6A/B**). It is interesting that activated HSCs numbers and viral replication were both gradually

increased (**Figure 6C/D**), which may be involved in liver repair when undergo the infection of DHAV.
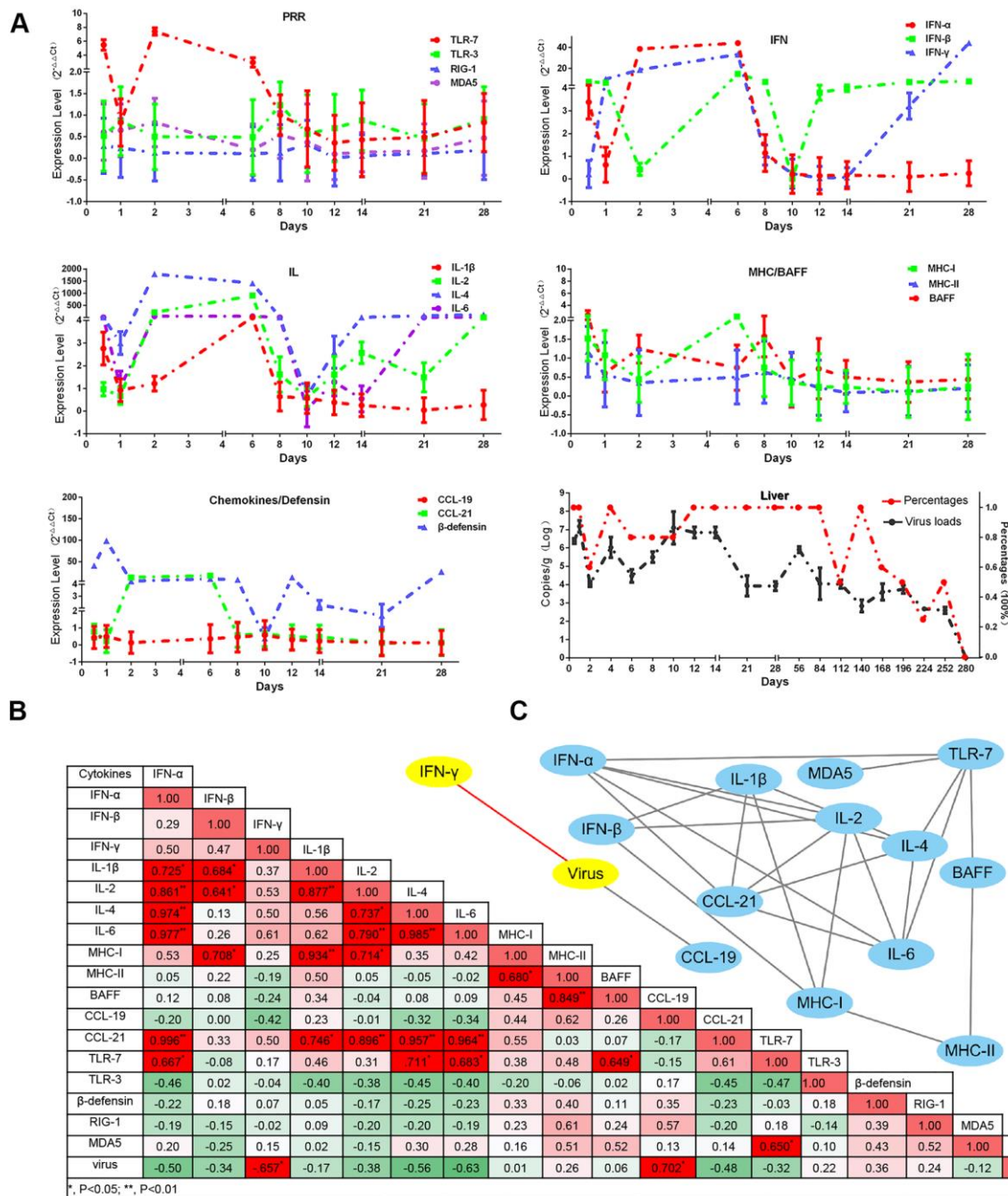
Figure 4: Dynamic immune-related gene expression profiles induced by DHAV-H strain and also the viral loads in liver. A. Immune-related genes are displayed in each part according to their biological functions, such as the activation of innate immune responses (TLR7, TLR3, RIG-1 and MDA5), effective interferons (IFN-α/β/γ) and interleukins (IL-1β/2/4/6), chemokines (CCL19/21), MHC-I and MHC-II, β-defensin, and B cell activating factor (BAFF). The relative values for each immune related genes were calculated by $2^{-\Delta\Delta Ct}$ method by comparing with the non-infected group. The viral RNA copies (Log10/g) and their percentage in each group are displayed at the bottom right. Data are represented as mean +/- SD (n=5). B**. In order to understand the impact of virulence on immune networks, the correlations of each pair of

immune related genes (P<0.05 at least) were calculated using correlation analysis (Pearson). **C.** Those correlated pairs of immune related genes were visualized as immune network by Cytoscape software.
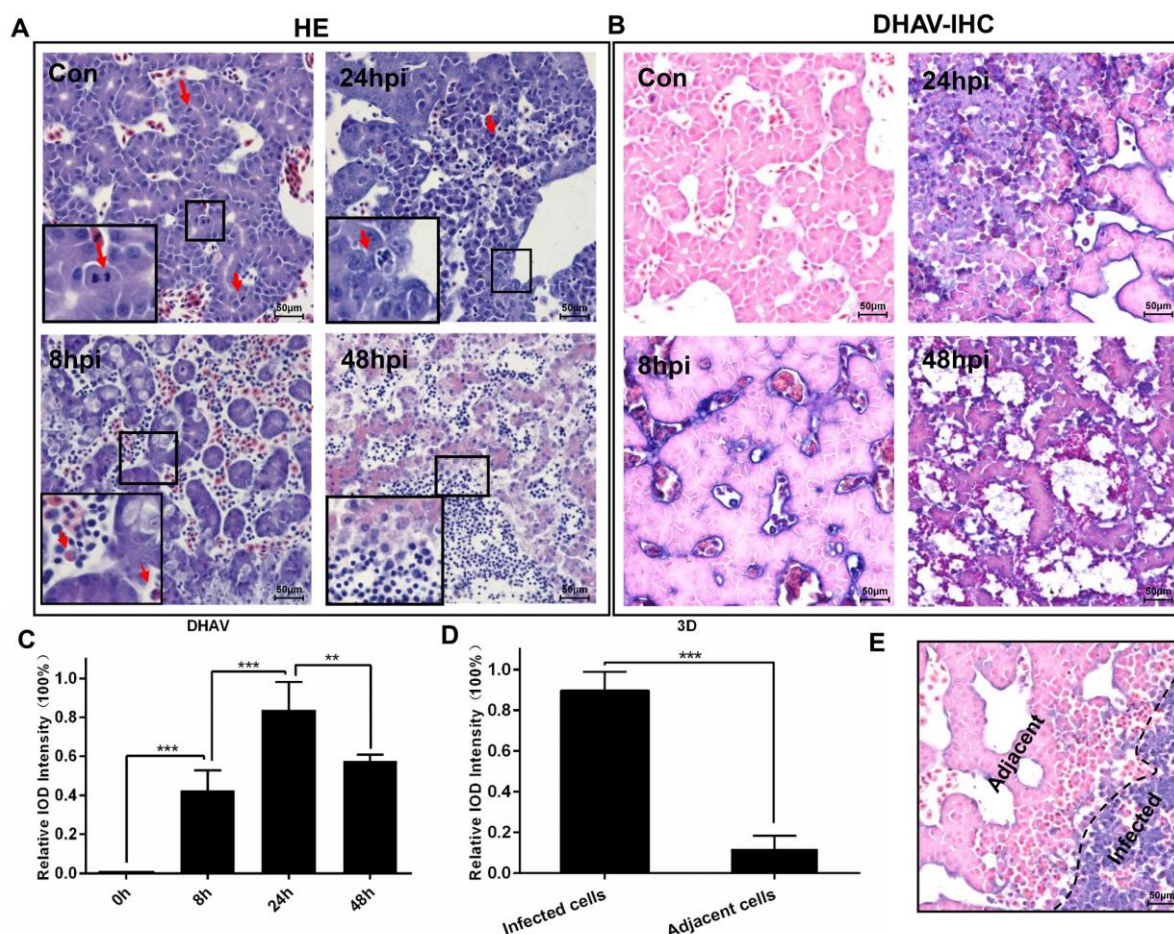
Figure 5: Microscopic lesions in the liver of duck embryos experimentally infected with DHAV-1 H strain. A. Representative liver HE staining from 8-hpi to 48-hpi of DHAV is displayed (n=3). Typical histopathological changes are exemplified at the corner of each photograph. (Con) represents pre-inoculation liver photomicrograph. **B.** Representative liver DHAV staining by IHC. **C.** The relative intensity of positive DHAV staining was calculated by Image-Pro Plus software. 100% Integral Optical Density (IOD) was defined as the most positive staining in the course of infection. The significant levels of their expression were analysed by Student's T-test, two way. **D.** The positive 3D^{pro} staining intensity of infected cells was significantly higher than the adjacent cells. **E.** Representative 3D protein staining at 24-hpi. The infected cells and adjacent cells were labelled. NS, No significance; *, P<0.05; **, P<0.01; ***, P<0.001. The brightness and contrast are slightly modified to create a uniform background.
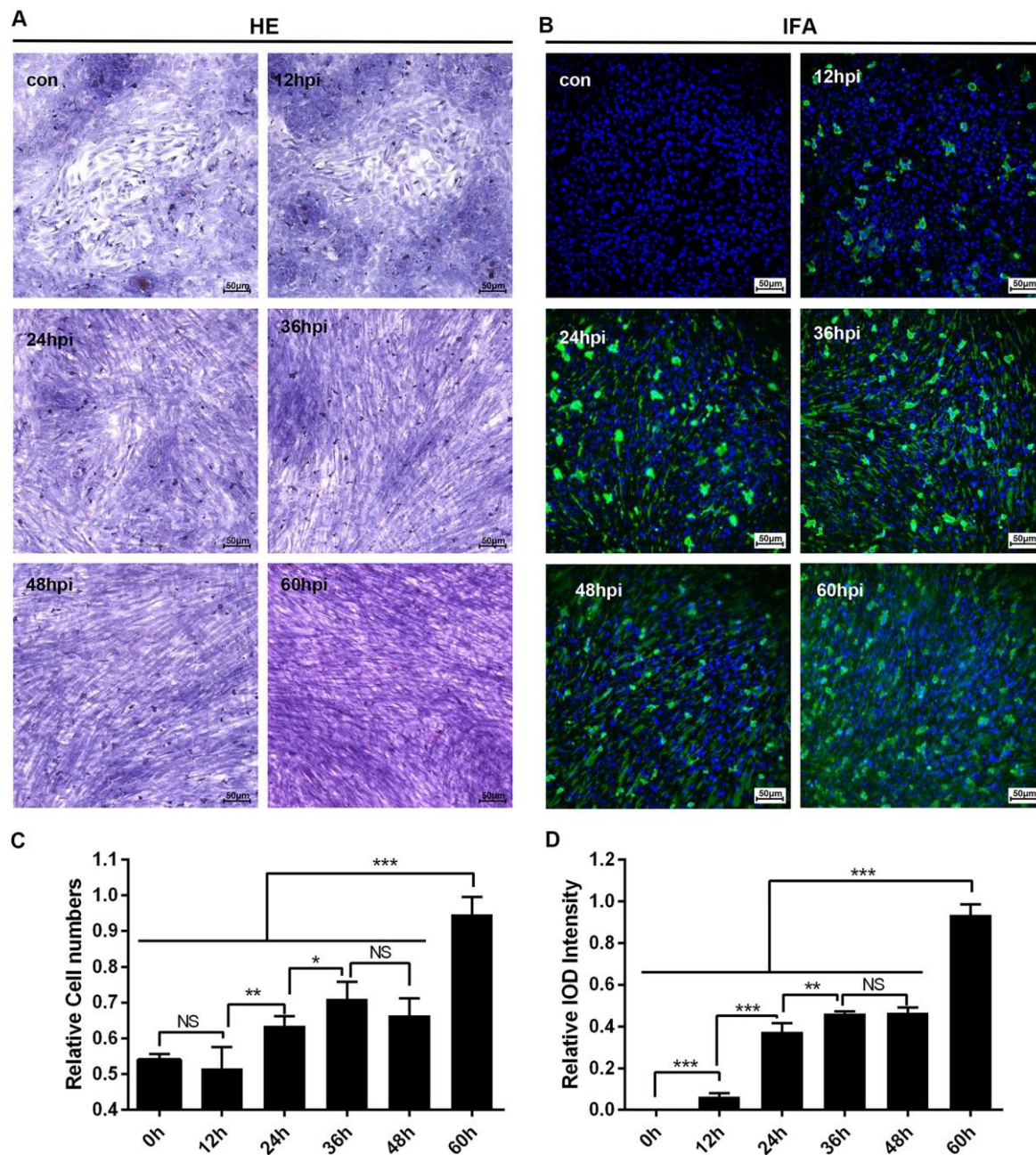
Figure 6: Microscopic lesions and virus replication of duck embryo HSCs infected with DHAV-1 H strain. A. Representative duck embryo HSC HE staining (n=2). B. Virus detection through indirect immunofluorescence after DHAV infection from 12-hpi to 60-hpi are displayed on the left and right (n=2). C. The relative cell numbers were calculated by Image-Pro Plus software. D.The intensity of positive staining were also calculated by Image-Pro Plus software. The significant levels of their expression or numbers were analysed by Student's T-test. NS, No significance; *, $P<0.05$; **, $P<0.01$; ***, $P<0.001$. The relative cell numbers was calculated by comparing the slides with maximum cells (1.0). The 100% Integral Optical Density (IOD) intensity was defined as the slides with the most positive staining. The brightness and contrast are slightly modified to create a uniform background.

Phylogenetic analysis of DHAV and HAV genomes and fibrosis-related genes in humans and ducks

To identify their sequence similarity, multiple alignments of DHAV and HAV were used for this purpose (https://blast.ncbi.nlm.nih.gov/Blast.cgi). They shared 50% nucleotide sequence identity with 20% protein sequence identity. Phylogenetic analysis provided curial information to understand the genetic relationship between different genera. In this analysis, though DHAV was related more closely to the genus *Parechovirus*, the avian hepatotropic virus, *Avihepatovius*, was also very close to the human Hepatitis A virus [34] (**Figure 7A**). Therefore, DHAV should be recognized as an HAV-related animal virus in birds.

To find the potential of ducks as an animal model for liver fibrosis, fibrosis-related genes in the genomes of humans and ducks were used to construct a phylogenetic tree. It has been shown that approximately 50% (16/32) of fibrosis-related genes had high homology between humans and ducks (**Figure 7B**). This result indicated that the mechanism of fibrogenesis might be similar in mammal and avian species, at least in this study.

DISCUSSION

Non-primate hepatitis animal models would be beneficial due to the lower cost to maintain, the availability to house within laboratories, and especially for the great concern about the shortage of primate models and phase out of use of them in research. Optional rodent models for hepatotropic virus research are necessary. Unfortunately, the currently developed humanized mouse models have limited life cycles or no viral replication [35]. The most recently discovered horse model provided an alternative large animal model for chronicity studies over several years. However, their higher expense and the shortage of immunological and biochemical reagents were the major obstacles faced by researchers [27]. Here, we provide an alternative small animal model that remedied irreparable defects in the HBV-related duck model owing to limited cytopathic effect. Fortunately, mature ducks infected with DHAV can induce acute and chronic hepatitis, which can be used to mimic the process of liver injury in the long term. Those advantages include the ability to perform acute hepatitis and chronic studies for fibrosis genesis, cirrhosis and hepatocellular regeneration and enough mass for serial liver biopsies.
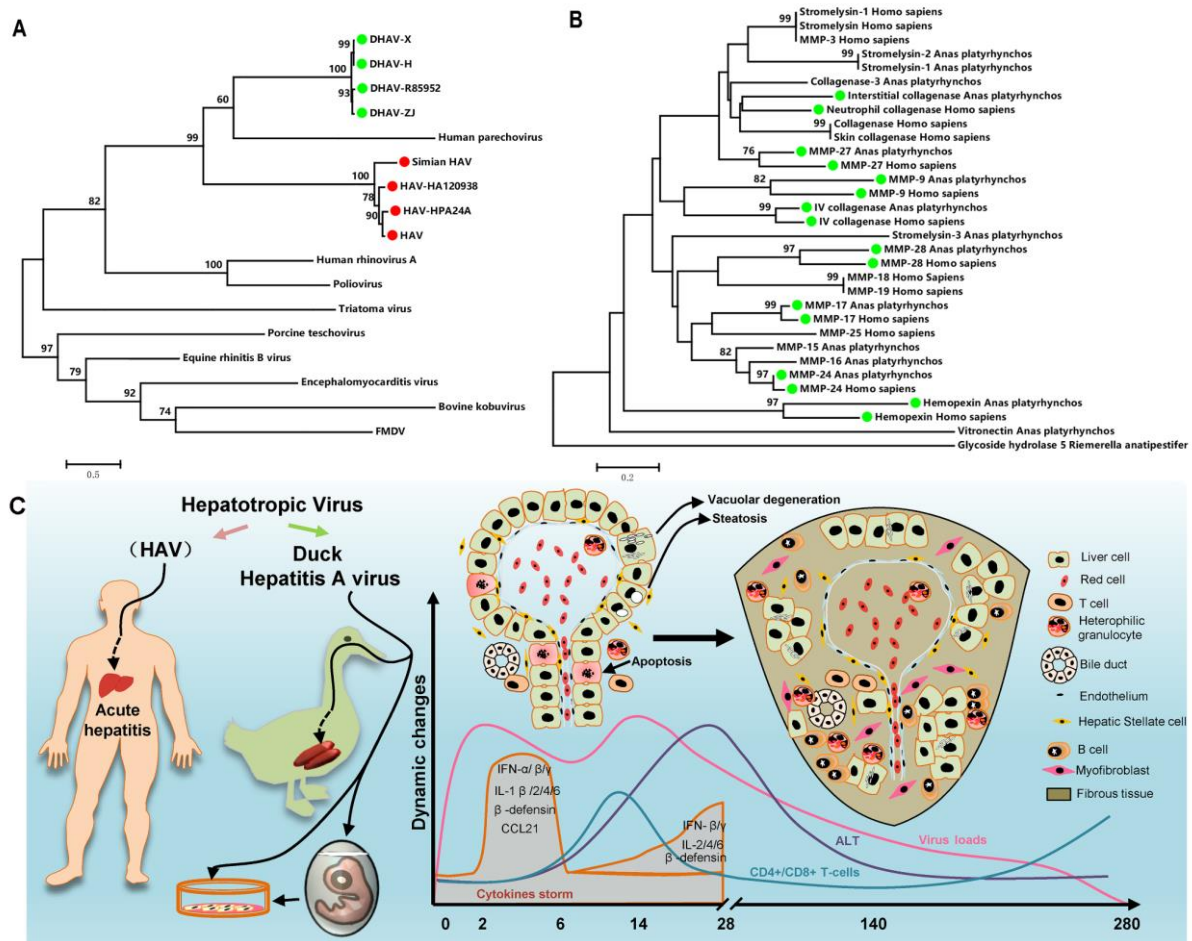
Figure 7: Comparatively phylogenetic relationship with "Hepatitis A virus" and fibrosis-related genes in humans and ducks and the process of liver injury and repair during DHAV infection. A. The phylogenetic tree of the DHAV-related genes was constructed from the complete genomic sequences of the picornaviruses using the neighbour-joining method. **B.** The complete amino acid sequences of the fibrosis-related genes in the genomes of humans and ducks were used to construct a phylogenetic tree using the neighbour-joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown above the branches. **C.** The duck has a similar liver disease caused by avian hepatropic virus (DHAV). This virus can also be cultured in duck embryos or its primary cells. During the period of infection, virus rapidly replicates to a high level, and then the transient cytokine storm reduces the virus to a comparatively lower level and gradually disappears until 280dpi. Hepatocellular injury was characterized as vacuolar degeneration, steatosis and apoptosis at early stage of infection. While those injured cells were replaced by collagenous fiber at later stage of infection. Those fibres may be secreted by myofibroblast when hepatic stellate cells were activated by injured liver cells. Meanwhile, strong recruitment of plasmocytes and heterophilic granulocytes were identified by HE assays. However, the comparatively lower CD4+ and CD8+ T-cells recruited to liver might lead to a persistent infection. Those immune cells may combine with interferons and interleukins to eradicate viral replication from liver.

In this longitudinal study, experimental inoculation of mature ducks resulted in a course of liver lesions, repair, and regeneration similar to HAV in primates, especially in the early stage of infection [20, 36]. During the early stage of infection, liver injury was characterized by

vacuolation, acidophilic degeneration, steatosis and infiltration of neutrophil granulocytes. Coincident with liver injury, the serum biomarkers, ALT and AST, were significantly increased (P<0.05). Similarly, human acute viral hepatitis also resulted in hepatocellular acidophilic degeneration and recruitment of a plentiful of lymphocytes [17]. The histological lesions in ducks associated with acute viral hepatitis were very similar to that in EHCV and HAV infection [20, 27]. Viral RNAs can be sensed by cytosolic PRRs, including Toll-like receptor 7 (TLR7), TLR3, retinoic acid-inducible gene I (RIG-I) and melanoma differentiation-associated gene 5 (MDA5) [37]. Activation of both types of PRRs by ssRNA or dsRNA was subjected to induce type I interferon and other inflammatory cytokines [38, 39]. However, except for TLR7, the other three members of the PRRs were only slightly up-regulated at 8-dpi to 14-dpi. This finding was supported by that the TLR3 signalling pathway was also disrupted through the cleavage of TRIF directed by a 3CD intermediate [40]. HAV infection also suppresses the production of Type I IFNs through the RIG-I/MDA5 pathway, through which signal transduction is blocked through the cleaving activity of 3ABC protease to Mitochondrial Antiviral Signalling Protein (MAVS) [40, 41]. The infiltration of neutrophil granulocytes was accompanied by strong elevation of CCL21 but not CCL19. This result was consistent with larger and more organized infiltrates directed by mouse ectopic expression of CCL21 [42]. Expression of CCL21 was also involved in the organization of inflammatory lymphoid tissue and the recruitment of T-cells, which may promote fibrogenesis via activation of chemokine receptor 7 (CCR7) on HSCs during chronic hepatitis C [43]. Previous studies indicated that recruitment of Tc and Th cells to the liver could increase the efficiency of viral clearance [44, 45]. In this study, however, the comparatively lower T-cell response may lead to a chronic DHAV infection.

During the later stage of infection, liver lesions were characterized by serious vacuolation, steatosis, interface hepatitis and progressive fibrosis. Many T-cells, plasm cells and heterophilic granulocytes were recruited to the liver to resist viral infection until the last stage of virus clearance. During the sustained battle between virus and host, the injured liver cells were replaced either by fibrosis or regenerated parenchyma cells. Our data indicated that the time of sustained infection in the liver (40 wk) was very similar to that of HAV-infected chimpanzees (35 to >48 wk) and much longer than HCV infection (10-20 wk) [20]. The sustained expression of viral RNA polymerases reflects ongoing viral replication (**Figure S5**), which was consistent with viral RNA copies (**Figure 2**). Combined with these evidences, the ducks infected DHAV could progress to chronic viral hepatitis. Interestingly, viral RNA copies peaked at 2 wk, dramatically declined at 4 wk, and then gradually vanished until 280-dpi in liver. The dynamic changes were negatively related to the intensity of Tc responses, interferons and interleukins and positively correlated with liver injury. To further investigate the role of cytokines in fibrogenesis, conjoint analysis indicated that Th1/Th2 cytokines (IFN-γ/IL-4) were inclined to Th2 immune responses from 14-dpi to 28-dpi, which was consistent with the progress of fibrosis (**Figure S1/S6**). These results indicated that the IL-4 might be the predominantly profibrogenic cytokines, which is consistent with CCl_4-induced fibrosis in mice [12]. Those characteristics of liver injury, fibrosis and regeneration should be recognized as

chronic viral hepatitis. To understand the potential of Peking ducks as a small model for research on human acute and chronic hepatitis, phylogenetic analysis of the homology of avian hepatotropic virus (DHAV) and human hepatovirus indicated that although DHAV was close to the genus *Parechovirus*, it was also highly related to HAV in the family *Picornaviridae* [34]. Additionally, phylogenetic analysis of ECM deposition-related genes in ducks and humans suggested that half of them shared higher homology (**Figure 7B**). Those data indicated that DHAV, the HAV-related avian hepatotropic virus, could induce very similar liver injury in human caused by HBV or HCV. This small model may benefit the pathogenesis and development of therapeutic strategies for liver injury in humans [10, 46].

Additionally, DHAV could be cultured in both duck embryos and primary duck embryo liver cells, which provided much more availability to address pathogenesis in vivo and in vitro. Serious liver injury caused by DHAV in duck embryos was characterized by apoptosis, cellular swelling to vacuolar degeneration (ballooning degeneration), apoptosis or necrosis, which was mainly caused by viral non-structural proteins, such as $2A^{pro}$ and $3C^{pro}$ or their precursors [40, 41, 47]. In addition, duck embryo HSCs could be stimulated and activated by DHAV infection (**Figure 6**). The activation of HSCs is involved in differentiation and proliferation of fibroblasts. Our data indicated that DHAV can be widely cultured in duck embryos and primary liver cells and can infect mature ducks progressing from acute hepatitis to chronic hepatitis.

In summary, the avian hepatotropic virus (DHAV) can induce apparently liver injury in both mature ducks and their embryos. Due to similar phenotype of liver injury, the pathogenesis of avian hepatotropic virus might be similar to some of the human hepatotropic virus. This small animal model provides the potential to address the process of acute and chronic viral hepatitis, fibrosis and hepatocellular regeneration, in a manner that was economical and easy to acquire and house in the laboratory (**Figure 7C**). Comparative and correlation studies between the viral life cycle and liver injury and repair are likely to provide insight into the mechanisms of viral persistence or clearance in chronic HBV or HCV infection.

References

1. WHO. (2013). Hepatitis A.
2. Gu CQ, Xie CQ, Hu XY, Zhang WP, Bi DR, Cheng GF. Cytokine gene expression in the livers of ducklings infected with duck hepatitis virus-1 JX strain. Poult Sci. 2012; 91: 583-91. doi: 10.3382/ps.2011-01743.
3. Salmon DE. (2013). Diseases of Poultry. In: Woolcock PR, ed. Duck Hepatitis: Wiley, pp. 422-31.
4. Anchun C, Mingshu W, Hongyi X, Dekang Z, Xinran L, Haijuen C, Renyong J, Miao Y. Development and application of a reverse transcriptase polymerase chain reaction to detect Chinese isolates of duck hepatitis virus type 1. Journal of Microbiological Methods. 2009; 77:332-336. doi: http://dx.doi.org/10.1016/j.mimet.2008.07.018.
5. Kim MC, Kwon YK, Joh SJ, Kim SJ, Tolf C, Kim JH, Sung HW, Lindberg AM, Kwon JH. Recent Korean isolates of duck hepatitis virus reveal the presence of a new geno- and serotype when compared to duck hepatitis virus type 1 type strains. Arch Virol. 2007; 152: 2059-72. doi: 10.1007/s00705-007-1023-0.
6. Tseng CH, Knowles NJ, Tsai HJ. Molecular analysis of duck hepatitis virus type 1 indicates that it should be assigned to a new genus. Virus Res. 2007; 123: 190-203. doi: 10.1016/j.virusres.2006.09.007.
7. Cacciola I, Pollicino T, Squadrito G, Cerenzia G, Orlando ME, Raimondo G. Occult Hepatitis B Virus Infection in Patients with Chronic Hepatitis C Liver Disease. New England Journal of Medicine. 1999; 341: 22-6. doi: 10.1056/NEJM199907013410104.
8. Dandri M, Burda MR, Zuckerman DM, Wursthorn K, Matschl U, Pollok JM, Rogiers X, Gocht A, Köck J, Blum HE. Chronic infection with hepatitis B viruses and antiviral drug evaluation in uPA mice after liver repopulation with tupaia hepatocytes. Journal of hepatology. 2005; 42: 54-60. doi: http://dx.doi.org/10.1016/j.jhep.2004.09.021.
9. Abbas Z, Memon MS, Mithani H, Jafri W, Hamid S. Treatment of chronic hepatitis D patients with pegylated interferon: a real-world experience. Antivir Ther. 2014; 19: 463-8. doi: 10.3851/IMP2728.
10. Winer BY, Ding Q, Gaska JM, Ploss A. In vivo models of hepatitis B and C virus infection. FEBS Lett. 2016; 590: 1987-99. doi: 10.1002/1873-3468.12157.
11. Kamari Y, Shaish A, Vax E, Shemesh S, Kandel-Kfir M, Arbel Y, Olteanu S, Barshack I, Dotan S, Voronov E, Dinarello CA, Apte RN, Harats D. Lack of Interleukin-1α or Interleukin-1β Inhibits Transformation of Steatosis to Steatohepatitis and Liver Fibrosis in Hypercholesterolemic Mice. Journal of hepatology. 2011; 55: 1086-94. doi: 10.1016/j.jhep.2011.01.048.
12. Weng S-Y, Wang X, Vijayan S, Kim Y, Kaps L, Tang Y, Molokanova O, Crosby J, Mccaleb M, Brombacher F. SAT-453-IL-4 and IL13 promote liver fibrosis signaling through M2 macrophages, but not Th2 T cells or hepatocytes. Journal of Hepatology. 2017; 66: S655. doi: 10.1016/S0168-8278(16)01375-1.
13. Wasmuth RMCBJ-C. (2016). Hepatitis B. In: Wedemeyer SMTBJRCSH, ed. Hepatology, pp. 37-44.
14. Garciade Leon Mdel C, Montfort I, Tello Montes E, Lopez Vancell R, Olivos Garcia A, Gonzalez Canto A, Nequiz-Avendano M, Perez-Tamayo R. Hepatocyte production of modulators of extracellular liver matrix in normal and cirrhotic rat liver. Exp Mol Pathol. 2006; 80: 97-108. doi: 10.1016/j.yexmp.2005.03.008.

15. Iredale JP. Models of liver fibrosis: exploring the dynamic nature of inflammation and repair in a solid organ. The Journal of clinical investigation. 2007; 117: 539-48. doi: 10.1172/JCI30542.

16. Zhou W-C. Pathogenesis of liver cirrhosis. World Journal of Gastroenterology. 2014; 20: 7312. doi: 10.3748/wjg.v20.i23.7312.

17. Guidotti LG, Chisari FV. Immunobiology and pathogenesis of viral hepatitis. Annu Rev Pathol Mech Dis. 2006; 1: 23-61. doi: 10.1146/annurev.pathol.1.110304.100230.

18. Bataller R, Brenner DA. Liver fibrosis. Journal of Clinical Investigation. 2005; 115: 209-18. doi: 10.1172/jci24282.

19. Ribeiro PS, Cortez-Pinto H, Solá S, Castro RE, Ramalho RM, Baptista A, Moura MC, Camilo ME, Rodrigues CM. Hepatocyte apoptosis, expression of death receptors, and activation of NF-κB in the liver of nonalcoholic and alcoholic steatohepatitis patients. The American journal of gastroenterology. 2004; 99: 1708-17. doi: 10.1111/j.1572-0241.2004.40009.x.

20. Lanford RE, Feng Z, Chavez D, Guerra B, Brasky KM, Zhou Y, Yamane D, Perelson AS, Walker CM, Lemon SM. Acute hepatitis A virus infection is associated with a limited type I interferon response and persistence of intrahepatic viral RNA. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108: 11223-8. doi: 10.1073/pnas.1101939108.

21. Rao HY, Wei L, Wang JH, Fei R, Jiang D, Zhang Q, Chen HS, Cong X. Inhibitory effect of human interferon-beta-1a on activated rat and human hepatic stellate cells. J Gastroenterol Hepatol. 2010; 25: 1777-84. doi: 10.1111/j.1440-1746.2010.06264.x.

22. Baroni GS, D'Ambrosio L, Curto P, Casini A, Mancini R, Jezequel AM, Benedetti A. Interferon gamma decreases hepatic stellate cell activation and extracellular matrix deposition in rat liver fibrosis. Hepatology. 1996; 23: 1189-99. doi: 10.1002/hep.510230538.

23. Du S, Li H, Cui Y, Yang L, Wu J, Huang H, Chen Y, Huang W, Zhang R, Yang J, Chen D, Li Y, Zhang S, et al. Houttuynia cordata inhibits lipopolysaccharide-induced rapid pulmonary fibrosis by up-regulating IFN-gamma and inhibiting the TGF-beta1/Smad pathway. Int Immunopharmacol. 2012; 13: 331-40. doi: 10.1016/j.intimp.2012.03.011.

24. Nasir GA, Mohsin S, Khan M, Shams S, Ali G, Khan SN, Riazuddin S. Mesenchymal stem cells and Interleukin-6 attenuate liver fibrosis in mice. Journal of translational medicine. 2013; 11: 1. doi: 10.1186/1479-5876-11-78.

25. Kishimoto T. IL-6: from its discovery to clinical applications. International immunology. 2010: dxq030. doi: 10.1093/intimm/dxq030.

26. Bukh J. Animal models for the study of hepatitis C virus infection and related liver disease. Gastroenterology. 2012; 142: 1279-87. e3. doi: 10.1053/j.gastro.2012.02.016.

27. Ramsay JD, Evanoff R, Wilkinson TE, Jr., Divers TJ, Knowles DP, Mealey RH. Experimental transmission of equine hepacivirus in horses as a model for hepatitis C virus. Hepatology. 2015; 61: 1533-46. doi: 10.1002/hep.27689.

28. Chayama K, Hayes CN, Hiraga N, Abe H, Tsuge M, Imamura M. Animal model for study of human hepatitis viruses. Journal of gastroenterology and hepatology. 2011; 26: 13-8. doi: 10.1111/j.1440-1746.2010.06470.x.

29. News NIoH. (2013 (June 26)). NIH to reduce significantly the use of chimpanzees in research.

30. Mason W, Seal G, Summers J. Virus of Pekin ducks with structural and biological relatedness to human hepatitis B virus. Journal of virology. 1980; 36: 829-36. doi: PMC353710.

31. Reaiche GY, Le Mire MF, Mason WS, Jilbert AR. The persistence in the liver of residual duck hepatitis B virus covalently closed circular DNA is not dependent upon new viral DNA synthesis. Virology. 2010; 406: 286-92. doi: 10.1016/j.virol.2010.07.013.

32. Kremsdorf D, Soussan P, Paterlini-Brechot P, Brechot C. Hepatitis B virus-related hepatocellular carcinoma: paradigms for viral-related human carcinogenesis. Oncogene. 2006; 25: 3823-33. doi: 10.1038/sj.onc.1209559.

33. Ou X, Mao S, Jiang Y, Zhang S, Ke C, Ma G, Cheng A, Wang M, Zhu D, Chen S, Jia R, Liu M, Sun K, et al. Viral-host interaction in kidney reveals strategies to escape host immunity and persistently shed virus to the urine. Oncotarget. 2016; 8: 7336-49. doi: 10.18632/oncotarget.14227.

34. Kim MC, Kwon YK, Joh SJ, Lindberg AM, Kwon JH, Kim JH, Kim SJ. Molecular analysis of duck hepatitis virus type 1 reveals a novel lineage close to the genus Parechovirus in the family Picornaviridae. J Gen Virol. 2006; 87: 3307-16. doi: 10.1099/vir.0.81804-0.

35. Bukh J. Animal models for the study of hepatitis C virus infection and related liver disease. Gastroenterology. 2012; 142: 1279-87.e3. doi: 10.1053/j.gastro.2012.02.016.

36. Costa-Mattioli M, Monpoeho S, Nicand E, Aleman MH, Billaudel S, Ferre V. Quantification and duration of viraemia during hepatitis A infection as determined by real-time RT-PCR. J Viral Hepat. 2002; 9: 101-6. doi: 10.1046/j.1365-2893.2002.00336.x.

37. Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nature immunology. 2010; 11: 373-84. doi: 10.1038/ni.1863.

38. Kawai T, Akira S. TLR signaling. Cell Death Differ. 2006; 13: 816-25. doi: 10.1038/sj.cdd.4401850.

39. Akira S, Takeda K. Toll-like receptor signalling. Nat Rev Immunol. 2004; 4: 499-511. doi: 10.1038/nri1391.

40. Qu L, Feng Z, Yamane D, Liang Y, Lanford RE, Li K, Lemon SM. Disruption of TLR3 signaling due to cleavage of TRIF by the hepatitis A virus protease-polymerase processing intermediate, 3CD. PLoS Pathog. 2011; 7: e1002169. doi: 10.1371/journal.ppat.1002169.

41. Yang Y, Liang Y, Qu L, Chen Z, Yi M, Li K, Lemon SM. Disruption of innate immunity due to mitochondrial targeting of a picornaviral protease precursor. Proc Natl Acad Sci U S A. 2007; 104: 7253-8. doi: 10.1073/pnas.0611506104.

42. Luther SA, Bidgol A, Hargreaves DC, Schmidt A, Xu Y, Paniyadi J, Matloubian M, Cyster JG. Differing activities of homeostatic chemokines CCL19, CCL21, and CXCL12 in lymphocyte and dendritic cell recruitment and lymphoid neogenesis. The Journal of Immunology. 2002; 169: 424-33. doi: 10.4049/jimmunol.169.1.424.

43. Bonacchi A, Petrai I, Defranco RMS, Lazzeri E, Annunziato F, Efsen E, Cosmi L, Romagnani P, Milani S, Failli P, Batignani G, Liotta F, Laffi G, et al. The chemokine CCL21 modulates lymphocyte recruitment and fibrosis in chronic hepatitis C. Gastroenterology. 125: 1060-76. doi: 10.1016/S0016-5085(03)01194-6.

44. Gerlach JT, Diepolder HM, Jung MC, Gruener NH, Schraut WW, Zachoval R, Hoffmann R, Schirren CA, Santantonio T, Pape GR. Recurrence of hepatitis C virus after loss of virus-specific CD4(+) T-cell response in acute hepatitis C. Gastroenterology. 1999; 117: 933. doi: https://doi.org/10.1016/S0016-5085(99)70353-7.

45. Thimme R, Wieland S, Steiger C, Ghrayeb J, Reimann KA, Purcell RH, Chisari FV. CD8+ T Cells Mediate Viral Clearance and Disease Pathogenesis during Acute Hepatitis B Virus Infection. Journal of Virology. 2003; 77: 68-76. doi: 10.1128/jvi.77.1.68-76.2003.

46. Liang TJ, Ghany MG. Current and Future Therapies for Hepatitis C Virus Infection. The New England journal of medicine. 2013; 368: 1907-17. doi: 10.1056/NEJMra1213651.

47. Cao J, Ou X, Zhu D, Ma G, Cheng A, Wang M, Chen S, Jia R, Liu M, Sun K, Yang Q, Wu Y, Chen X. The 2A2 protein of Duck hepatitis A virus type 1 induces apoptosis in primary cell culture. Virus Genes. 2016: 1-9. doi: 10.1007/s11262-016-1364-4.

48. Yang M, Cheng A, Wang M, Xing H. Development and application of a one-step real-time Taqman RT-PCR assay for detection of Duck hepatitis virus type1. J Virol Methods. 2008; 153: 55-60. doi: 10.1016/j.jviromet.2008.06.012.

49. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013; 30: 2725-9. doi: 10.1093/molbev/mst197.

50. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCtMethod. methods. 2001; 25: 402-8. doi: 10.1006/meth.2001.1262.

Supplementary information
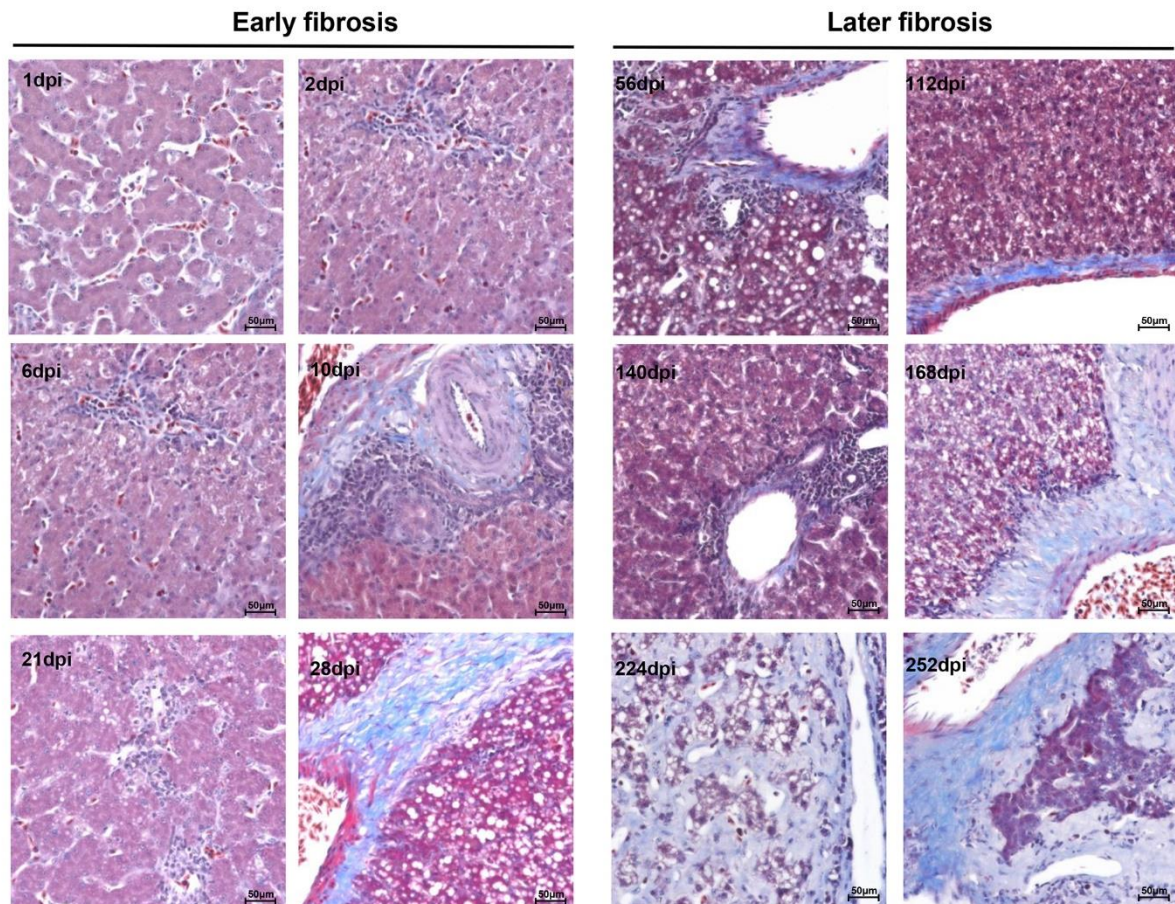
Figure S1: Liver fibrogenesis in mature ducks experimentally infected with DHAV-1 H strain. The representative liver Masson staining from 1d to 280d post-infection with DHAV are also displayed with two stages, early infection and later infection. Blue colour represents collagenous fibres. The brightness and contrast are slightly modified to create a uniform background.
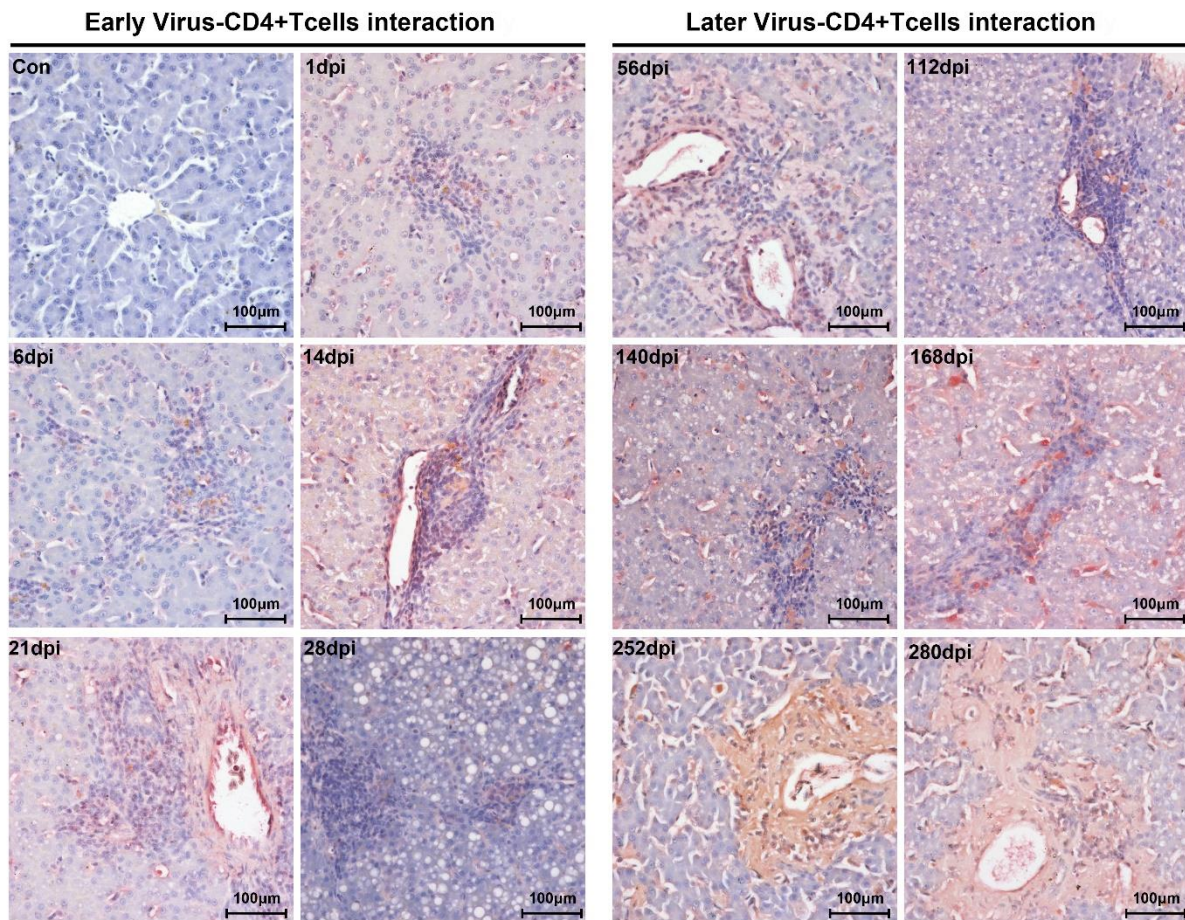
Figure S2: Double staining of viral capsid and CD4+ positive T-cells in liver. Viral capsid and CD4+ T-cells were double stained by rabbit anti DHAV capsid antigen polyclonal antibody and mouse anti duck CD4 monoclonal antibody. Red colour and brown colour represent positive capsid antigens and CD4+ T-cells, respectively. (Con) represents liver without double primary antibody. The brightness and contrast are slightly modified to create a uniform background.
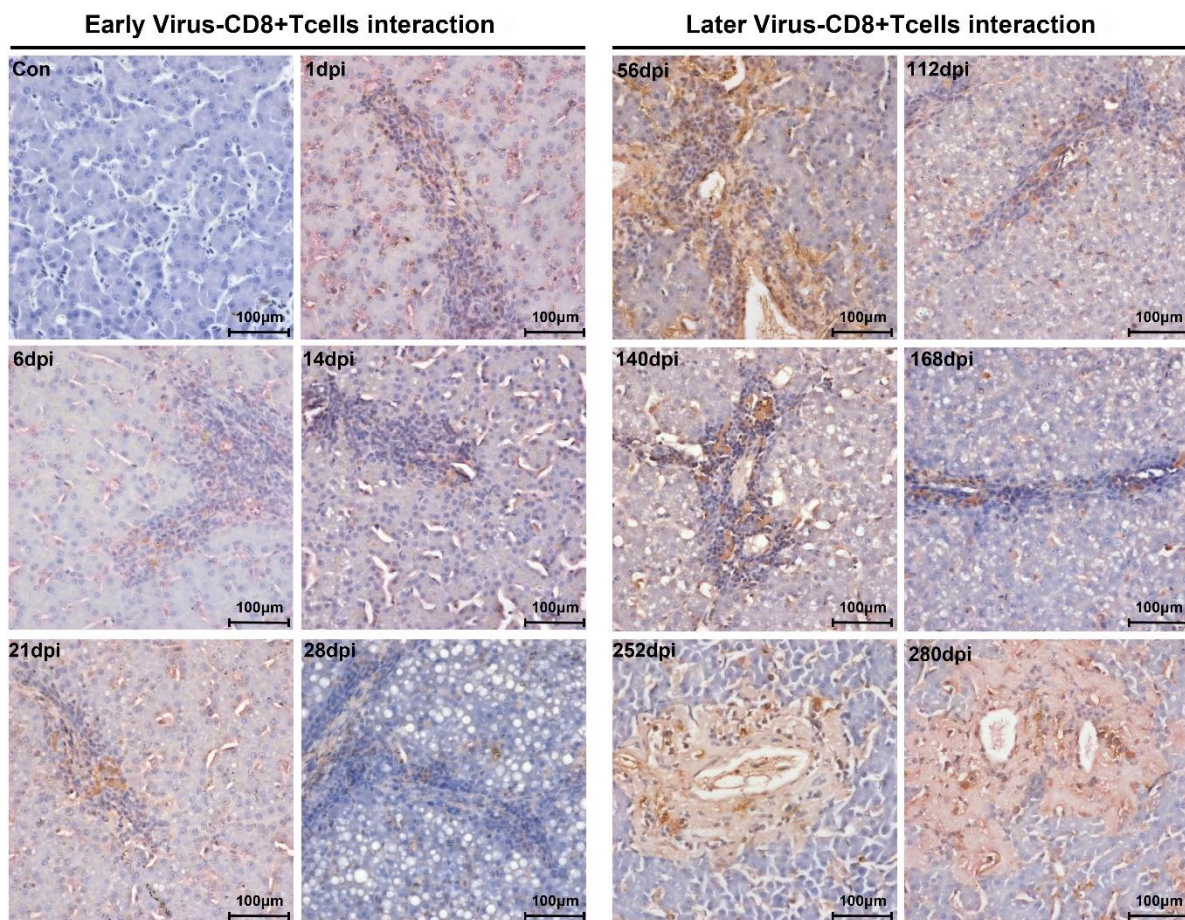
Figure S3: Double staining of viral capsid and CD8+ positive T-cells in liver. Viral capsid and CD8+ T-cells were double stained with rabbit and DHAV capsid antigen polyclonal antibody and mouse anti duck CD8α monoclonal antibody. Red colour and brown colour represent positive capsid antigens and CD8α+ T-cells, respectively. (Con) represents liver without double primary antibody. The brightness and contrast are slightly modified to create a uniform background.
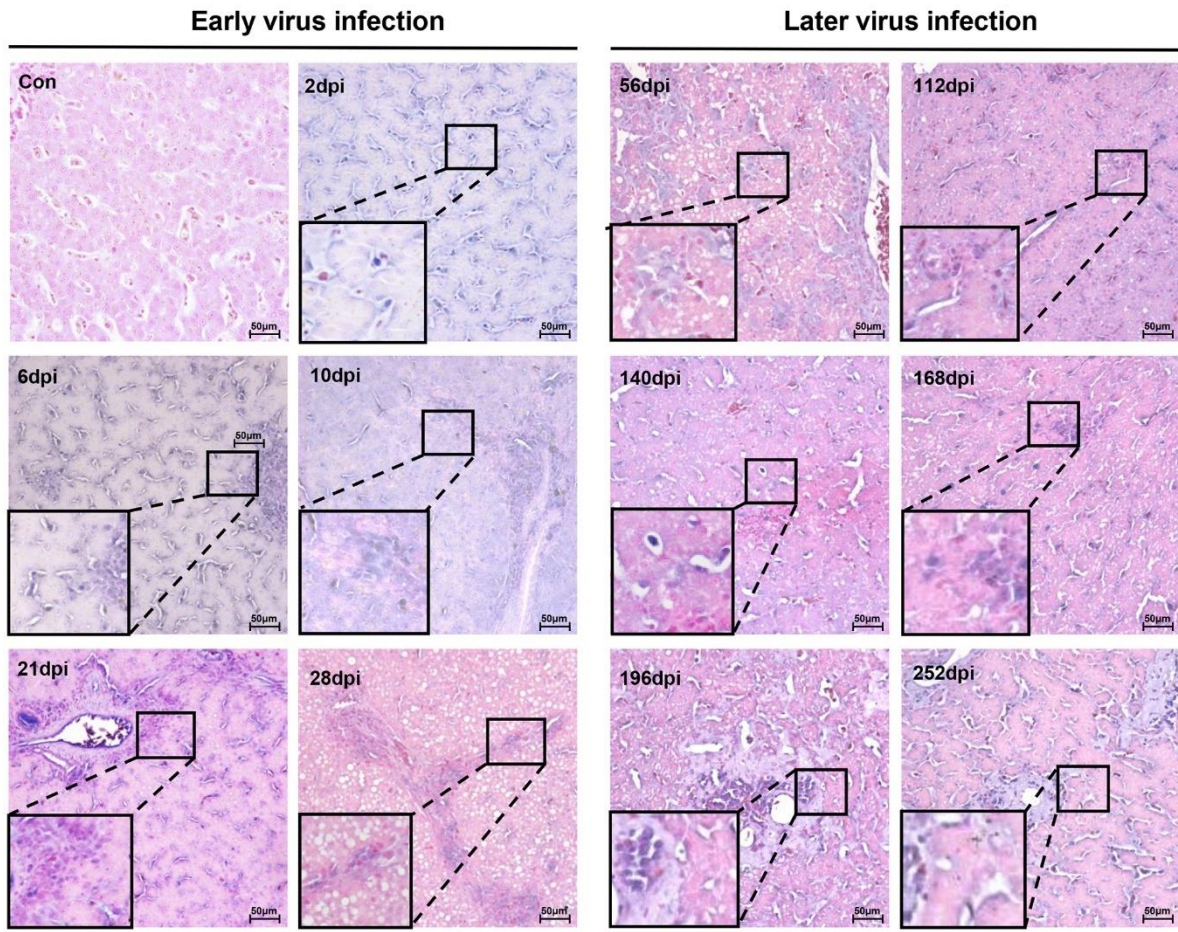
Figure S4: Dynamic expression levels of viral capsid antigens in liver. Viral capsid is shown by rabbit and DHAV capsid antigen polyclonal antibody with blueviolet. 400× magnification of photograph matrices according to the time post-infection. Typical virus distribution was exemplified at the corner of each photograph. (Con) represents liver without primary antibody. The brightness and contrast are slightly modified to create a uniform background.
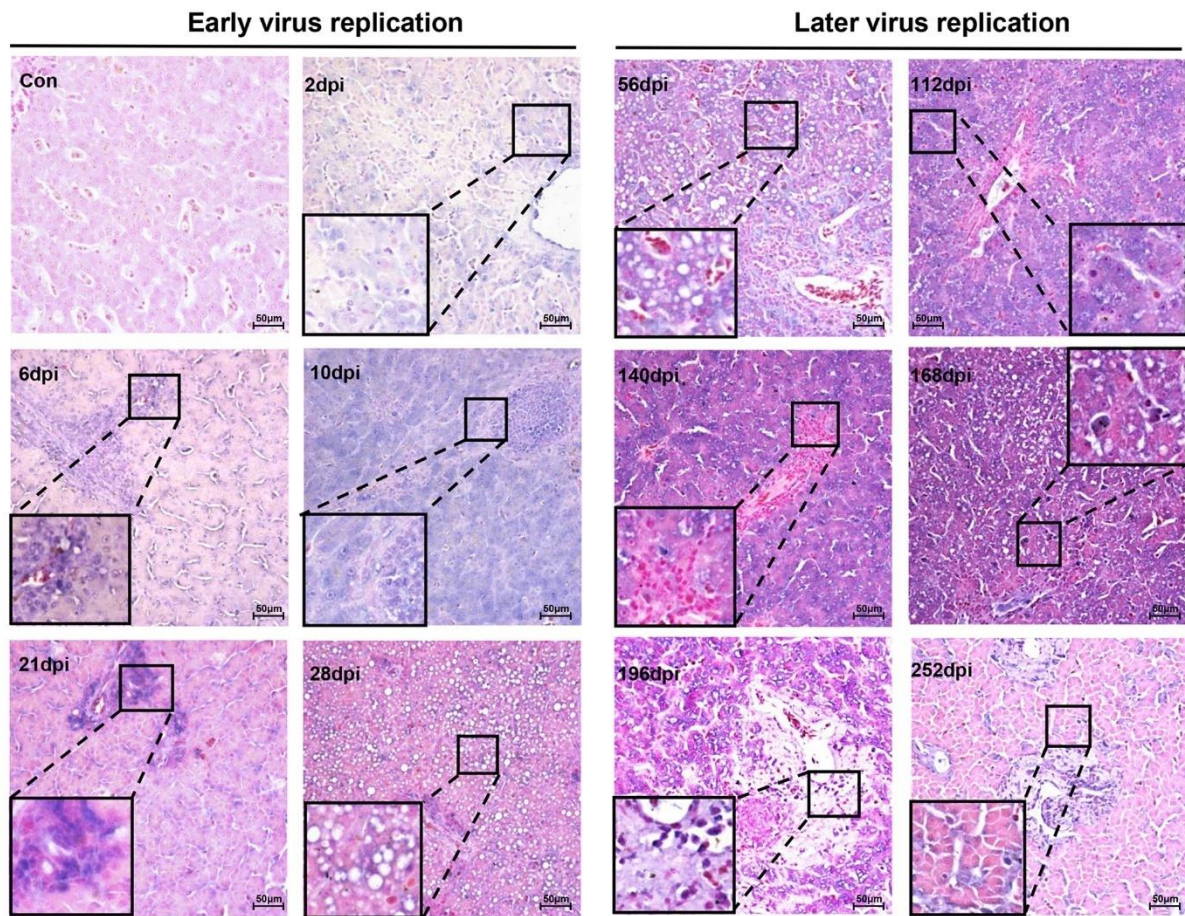
Figure S5: Dynamic expression level of viral RNA polymerases in liver. Viral replication was detected by rabbit anti DHAV RNA polymerase polyclonal antibody with blue-violet. (Con) represents liver without the primary antibody. The brightness and contrast are slightly modified to create a uniform background.
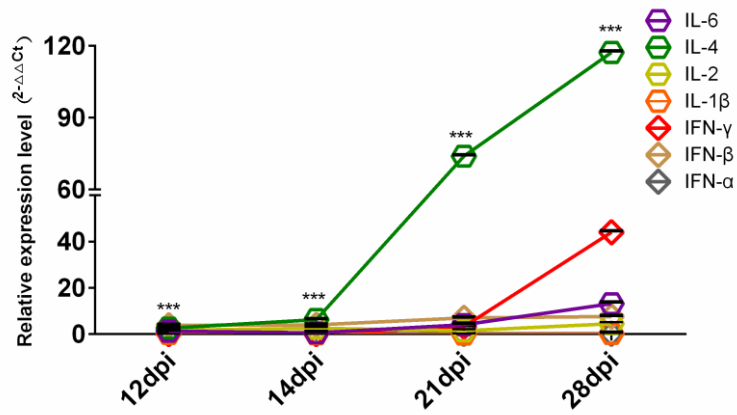
Figure S6: Comparative analysis of interferons and interleukins from 12-dpi to 28dpi. The relative values for each cytokine were calculated by $2^{-\Delta\Delta Ct}$ method by comparing with the non-infected group. The significant difference at different time points between IFN-γ and IL-4 were calculated by student's T-test. *, P<0.05; **, P<0.01; ***, P<0.001.

Chapter 7

Incompatible translation drives a convergent evolution and viral attenuation during the development of live attenuated vaccine

Xumin Ou, Mingshu Wang, Sai Mao, Jingyu Cao, Anchun Cheng, Dekang Zhu, Shun Chen, Renyong Jia, Mafeng Liu, Qiao Yang, Ying Wu, Xinxin Zhao, Shaqiu Zhang, Yunya Liu, Yanling Yu, Ling Zhang, Xiaoyue Chen, Maikel P Peppelenbosch, Qiuwei Pan

Abstract

Live attenuated vaccines are widely used to protect humans or animals from pathogen infections. We have previously developed a chicken embryo-attenuated Duck Hepatitis A Virus genotype 1 (DHAV-1) vaccine (CH60 strain). This study aims to understand the mechanisms that drive a virulent strain to an attenuated virus. Here, we systematically compared five DHAV-1 chicken embryo attenuated strains and sixty-eight virulent strains. Phylogenetic analysis indicated that duck virulent strains isolated from different geographic regions of China undergo a convergent evolution in the chicken embryos. Comparative analysis indicated that the codon usage bias of the attenuated strains was shaped by chicken codons usage bias, which essentially contributed to viral adaption in the unsuitable host driven by incompatible translation. Of note, the missense mutations in coding region and mutations in untranslated regions may also contribute to viral attenuation of DHAV-1 to some extent. Importantly, we have experimentally confirmed that the expression levels of four viral proteins (2A2pro, 2A3pro, 3Cpro and 3Dpro) in the liver and kidney of ducks infected with an attenuated strain are significantly lower than that infected with a virulent strain, despite with similar virus load. Thus, the key mechanisms of viral attenuation revealed by this study may lead to innovative and easy approaches in designing live attenuated vaccines.

Keywords: Duck hepatitis A virus, attenuation, RSCU, tRNA, translational selection, vaccine

Introduction

Attributing to the effective immunogenicity and protection, live attenuated vaccines are widely used to protect humans or animals from certain pathogen infections since the recent half century (Bhamarapravati and Sutee, 2000;Belshe et al., 2007;Hviid et al., 2008;Song et al., 2013;Minor, 2016). One of the successful examples is the development of attenuated strains of Duck Hepatitis A Virus genotype 1 (DHAV-1) through serial passaging in chick embryos, which can provide perfect protection in immune ducklings (Cheng et al., 1993;Ou et al., 2017a). DHAV-1, a member of the family Picornaviridae, was recently classified into the unique genus of Avihepatovirus (http://www.ictvonline.org/virusTaxInfo.asp). It is highly pathogenic to ducklings less than one week old (morbidity and mortality, 100% and 95%, respectively) (Salmon, 2013). We have previously established methods for virological and immunological detection of this virus and characterized the functions of several viral proteins (Cheng et al., 2009;Cao et al., 2012;Wen et al., 2014;Shen et al., 2015;Cao et al., 2016;Hu et al., 2016;Mao et al., 2016;Ou et al., 2016;Ou et al., 2017b;Yang et al., 2017;Zhang et al., 2017). However, the mechanisms of viral attenuation for the successful development of the DHAV vaccine remain largely elusive.

DHAV-1 is a single-stranded positive RNA virus with similar genomic organization of Picornavirus, including 5' untranslated regions (UTR) covalently linked by a VPg, a large open reading frame (ORF), 3'UTR polyadenyliced by the PolyA tail (Kim et al., 2006;Racaniello, 2013;Sun et al., 2016). Once the viral RNA released into the host cells, it serves as a template for both translation and replication to assemble a huge number of progeny viruses (Tuthill et al., 2010;Racaniello, 2013;Wen et al., 2015). However, the expression of viral proteins largely relies on the host translation system. During passaging in chicken embryos, DHAV-1 triggers diverse pathological changes from slight to serious upon serial passaging (Hwang and E. Dougherty, 1962;Salmon, 2013). During this process, synonymous and non-synonymous mutations will emerge to make up the pressures from translation or function of viral proteins (Ran et al., 2014). The subsequent selections in the viral genome are the consequences of multiple host factors, including codon autocorrelation, clustering of rare codons, mRNA secondary structure, ribosomal density, relative abundance of wobble base pairs and modified tRNA (Novoa and de Pouplana, 2012).

The missense mutations change certain amino acid and may lead to structural or functional alterations of the protein. These may contribute to fitness or defection of the virus (Nilsson et al., 2003;Appel et al., 2008;Voitenleitner et al., 2012). It has been reported that missense mutations in nonstructural protein of hepatitis A virus (2B protein and 2C protein) and hepatitis C virus (NS5A and NS5B) will increase viral adaptation (Emerson et al., 1992;Lohmann et al., 2001;Voitenleitner et al., 2012). In contrast, synonymous mutations do not change the protein sequence. But they are not necessary to be neutral, because prefect codon is translated more efficiently, resulting in high levels of protein expression. Thus, we hypothesize that synonymous and missense mutations occurred during the passage in chick embryo are

likely due to incompatibility of the host translation system. Because different types of hosts vary considerably in respect to the diversity of tRNA gene numbers, Relative Synonymous Codon Usage (RSCU) and tRNA modification enzymes (Novoa Eva Maria et al., 2012;Ran et al., 2014;Powell and Dion, 2015).

To address whether the incompatible translation system of the host causes viral attenuation, we first performed phylogenetic analysis of virulent and attenuated DHAV-1 strains to identify whether those attenuated strains undergo a similar evolution in chicken embryos. We then utilized the abundance of tRNA and RSCU in Gallus gallus and Anas Platyrhynchos genomes to understand the occurrence of those synonymous mutations upon serial passaging in chick embryos. Next, tertiary structural variations in structural and non-structural proteins of the virulent and attenuated strains were analyzed by homology modeling. Finally, we experimentally demonstrated the lower expression level of viral nonstructural proteins in the duck liver and kidney infected with the attenuated strains.

Materials and methods

Ethics Statement

The 9-day-old specific pathogen free (SPF) chicken embryos used in passage study were purchased at Merial Company (http://www.merial.com.cn), and this study performed in strict accordance with the recommendations in the ARRIVE guidelines (http://www.nc3rs.org.uk/arrive-guidelines). The animal experiment has been approved by the committee of experiment operational guideline and animal welfare of Sichuan Agricultural University, China (The approved permit number is XF2014-18). All ducks were handled in compliance with the animal welfare regulations and maintained according to standard protocols. All surgeries performed on animals were under sodium pentobarbital anesthesia, and all efforts were made to minimize suffering.

Sequences of virus and host

The DHAV-1 CH60 attenuated strain is a commercial vaccine approved by Ministry of Agriculture (PRC). The attenuated strain, which was derived from the DHAV-1 CH strain in the allantoic cavities of 9-day-old specific pathogen free (SPF) chicken embryos after 60 passages, is a commercial vaccine developed by our laboratory. Other genomes of 72 Duck hepatitis A virus species were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/genbank), including 4 strains of chicken embryo attenuated viruses, 68 strains of virulent viruses (**Supplementary Table 1**). The abundance of tRNA and RSCU of *Gallus gallus* and *Anas Platyrhynchos* were obtained from Genbank and Condon Usage Bias Databases, respectively (http://www.kazusa.or.jp/codon/). The ADAT$_2$ reference sequences of *Homo sapiens*(NP_872309), *Gallus gallus*(XP_419709) and *Anas platyrhynchos* (XP_012962288) were obtained from Protein database from NCBI.

Phylogenetic analysis

To elucidate the phylogenetic relationship between chick embryo attenuated strains and virulent strains, 68 virulent strains and 5 chick embryos attenuated strains were analyzed by Neighbor joining method using Mega 6.0 (Tamura et al., 2013).

Codon usage indices

Nucleotides frequencies at the third position of whole genome was calculated by bio-python software. The RSCU, Uracil at the 3rd codon position (U3s), A3s, G3s，C3s and the third synonymous codon position GC content (GC3s) were calculated by CodonW (http://www.molbiol.ox.ac.uk/cu, version 1.4.2) using Saccharomyces cerevisiae as reference. RSCU used by virulent and attenuated strains were visualized by Hemi software with hierarchical clustering analysis (http://hemi.biocuckoo.org/).

Correlation analysis of RSCU and tRNA copies in chicken and duck genomes

Those significant changed codons in chick embryo attenuated strains were correlated with RSCU and tRNA copies in chicken and duck, respectively. The corresponding codons in virulent strains were also analyzed by the above methods. Linear regression and significant correlations were calculated using IBM SPSS Statistic 20. The RSCU and tRNA copies used in chicken and duck genome were plotted by Graphpad Prim software. The NNA, NNT, NNG and NNC used in chicken and duck were compared by Wilcoxon matched-pairs signed-ranks test. $tRNA^{TNN}$, $tRNA^{ANN}$, $tRNA^{CNN}$ and $tRNA^{GNN}$ were also compared. $*P < 0.05$, $**P < 0.01$ or $***P < 0.001$.

Multiple sequences alignment

Conforming nucleotide substitutions were identified by Mega 6.0 using Cluster W method (Tamura et al., 2013). The details of synonymous and non-synonymous mutations were listed in supplementary Table 2. The frequencies of nucleotides in ORF (68 virulent strain and 5 attenuated strain) were calculated by biopython analysis. The nucleotide with the highest frequencies at each site was used to construct two conserved ORF sequence (virulent strain and attenuated strain).

Secondary structural prediction of 5'UTR and 3'UTR

The 5'UTR and 3'UTR of DHAV-1 virulent (ATCC, DRL-62 strain) and chick embryo attenuated virus (C80 strain) were imported into RNAfold web server to predict their secondary structures (Gruber et al., 2008). The sequences and CT files were visualized by Vienna VARNA package. In order to understand whether those fixed SNPs could significantly change the secondary structure of virulent strains, the 5'UTR and 3'UTR of artificial mutant 5'UTR or 3'UTR were also

imported into RNAfold web server to predict their secondary structure (Supplementary Table 2).

Experimental design and Immunohistochemistry (IHC)

15 ducks were randomly divided into three groups. Group 1 (CH60 strain) and group 2 (H strain) received 1 ml of virus (4.56×10^8 copies/ml) by intramuscular injection, while group 3 was injected with an equal volume of 0.85% physiological saline as a negative control. The tissues of the liver and kidney from the mature ducks (160 days of age) inoculated with virulent strains and chicken embryo attenuated strains at 4 days post-inoculation were fixed in 4% paraformaldehyde, dehydrated, embedded in paraffin, sectioned into 4-μm thick sections and followed with previous established IHC protocol (Ou et al., 2017a).

qRT-PCR

Virus loads in liver and kidney were detected by real-time PCR assay according to the previously established method (Yang et al., 2008). Briefly, one hundred milligrams of each tissue were used for RNA isolation, then the viral RNA copies was detected by one step Real-time PCR assay. The viral RNA copies in liver or kidney were translated by copies (Log_{10} /g).

Homologous modeling

Online homologous modeling software of SWISS-MODEL was used to analysis the structural variations resulted from missense mutations on Table S2 in the process of serial passages in chicken embryos (Biasini et al., 2014). It base sensitive Hidden Markov Models (HMM) to search against the SWISS-MODEL Template Library (SMTL). And those amino acid (aa) substitutes were demonstrated by the PyMOL molecular graphics system(DeLano, 2002)(**supplemental figure 2-4**). Templates for homologous modeling are as follows, Ljiugan viral capsids (PDB:3Jb4: A/B/C) for modeling DHAV-1 VP1/0/3(Zhu et al., 2015), GTPase IMAP family member 7 (PDB: 3zjc.3.A) for 2A2 protein (Schwefel et al., 2013), Adipose phospholipase A (PDB: 4fa0.1.A) for 2A3 protein (Pang et al., 2012), Actin-related protein 2/3 complex subunit 3(PDB: 4jd2.1.E) for 2B protein (Luan and Nolen, 2013), Minichromosome maintenance protein (MCM)(PDB: 4r7y.1.A) for 2C protein (Miller et al., 2014), AopB (PDB: 3wxx.2.B) for 3A protein (Nguyen et al., 2015), iron-regulated surface determinant protein H (PDB: 2lhr.1.A) for 3B protein (Spirig et al., 2013), Enterovirus 71 3C protein (PDB: 3qzq.4.A) for 3C protein (Wang et al., 2011), Poliovirus 3D polymerase (PDB: 4nlr.1.A)for 3D protein (Sholders and Peersen, 2014).

RESULTS

Convergent evolution of the virulent strains upon passaging in chicken embryos

In general, virulent strains of DHAV-1 isolated from ducklings can be attenuated through series of passaging in chicken embryos, accompanied by histopathological injury from mild to severe (**Figure 1a**)(Hwang and E. Dougherty, 1962). Phylogenetic analysis indicated a convergent evolution of the virulent strains isolated from different regions under the host selection of chicken embryos (**Figure 1b**). Genomic and protein sequence alignment between chick embryo attenuated strains and duck virulent strains indicated that a large number of identical mutations were selected in those five attenuated strains (Supplementary data 1-2). next, the nucleotide frequencies at the third position were compared between attenuated and virulent strains (**Figure 1c**). To estimate the percentage of synonymous and non-synonymous, the frequencies of nucleotides at each site were used to construct two conserved viral ORFs (virulent strain and attenuated strain). As calculated, the total mutations in the ORF is 208, when compared to the above constructed sequence. Specifically, 30 out of 208 mutations will lead to 27 non-synonymous mutations. Therefore, the percentage of synonymous mutation is 85.57% (178/208=0.8557). (**Supplementary data 3**). Those results indicated that codon usage bias was apparently changed after passaging in chick embryos. Specifically, Adenine at the 3rd codon position (A3s) and C3s in chicken embryo attenuated strains is much more abundant than that in virulent strains (P<0.01), but much less at T3s and G3s (**Figure 1c**). During the process of passaging, a number of SNPs were also acquired (**Supplementary Table 2**). These results indicated convergent evolution of genomic sequence and the pattern of codon usage of DHAV-1 during serial passage in chick embryo.

Incompatible host translation shapes viral codon usage bias

The virulent viruses insolated from duck are originally non-infectious to chicken. While after series of passages in chicken embryos, it indeed adapt the new host (Hwang and E. Dougherty, 1962). When we inoculate back this chicken attenuated virus to their original host, duck, virulence is attenuated. In all organism, protein translation is the last step to readout genetic codons. While different host translation machinery uses different decoding strategy, the major reason is diversity of codon usage bias used by different host, as chicken and duck in this study. Importantly, it had been proven that the expression of viral genes largely relies on the host translation system, and bias codons that is compatible to their host can improve translation efficiencies and protein fold (Hanson and Coller, 2017).To elucidate the potential mechanism of codon bias in chick embryo attenuated strains, the corresponding codon frequencies and tRNA copies in chickens and ducks were simultaneously analysed with viral codon usage bias. The correlation analysis indicates that the codon frequencies in attenuated strains are higher (24.45%) in correlation with chicken codon frequencies than that in duck

(Correlation index (0.570 vs 0.458)). But those codon frequencies are not correlated with tRNA copies in both chicken and duck genome (**Figure 2a**). Interestingly, the corresponding codon frequencies in virulent strains are also higher (30.70%) in correlation with chicken codon frequencies than that in duck (Correlation index (0.5163 vs 0.395)). Those codon frequencies in virulent strains are not correlated with tRNA copies in both chicken and duck genome (**Figure 2b**). In the genomes of both *Gallus gallus* and *Anas platyrhynchos*, the codon frequencies per thousand is positively correlated with the abundance of tRNA in its genome respectively (P<0.01) (**Figure 2c**). The corresponding codons, NNA, NNT, NNG and NNC, used in chicken are significantly higher than that in duck (P<0.05 at least) (**Figure 2d**). Since tRNA copies have significant impact on codon usage bias, the corresponding tRNA copies were also analyzed. tRNA^{ANN} and tRNA^{CNN}, but not tRNA^{TNN} and tRNA^{GNN}, are significantly higher than that in duck (P<0.05) (**Figure 2e**). Altogether, the higher correlation of codon frequencies in chicken embryo attenuated strains and their new host (chicken) indicated that viral codon usage bias was shaped by incompatible host translation that duck insolated virus must adapt in chicken embryos by using this incompatible host translation.

High variations in secondary structure of 5'UTR but not 3'UTR of the DHAV-1 genome

Because viral 5'UTR and 3'UTR are constructively needed for IRES-mediated translation. Their secondary structural variations could regulate the efficiency of viral translation. Therefore, the secondary structures of 5'UTR and 3'UTR of the DHAV-1 virulent strain (ATCC, DRL-62) and chick embryo attenuated strain (C80) were predicted by RNAfold web server according to the rule of minimum free energy (MFE) (Gruber et al., 2008). We next analyzed whether the confirmed mutations in supplementary Table 2 can lead to a secondary structural variations or not. We found that the secondary structure at 5'UTR was heavily changed during the process of serial passaging in chick embryos and this was consistent with DRL-62 mutation analysis (**Supplementary Figure 1a**). However, there was only one nucleotide substitution (145, A-G) occurred at 3'UTR which led to a slight extension of a central stem (**Supplementary Figure 1b**).
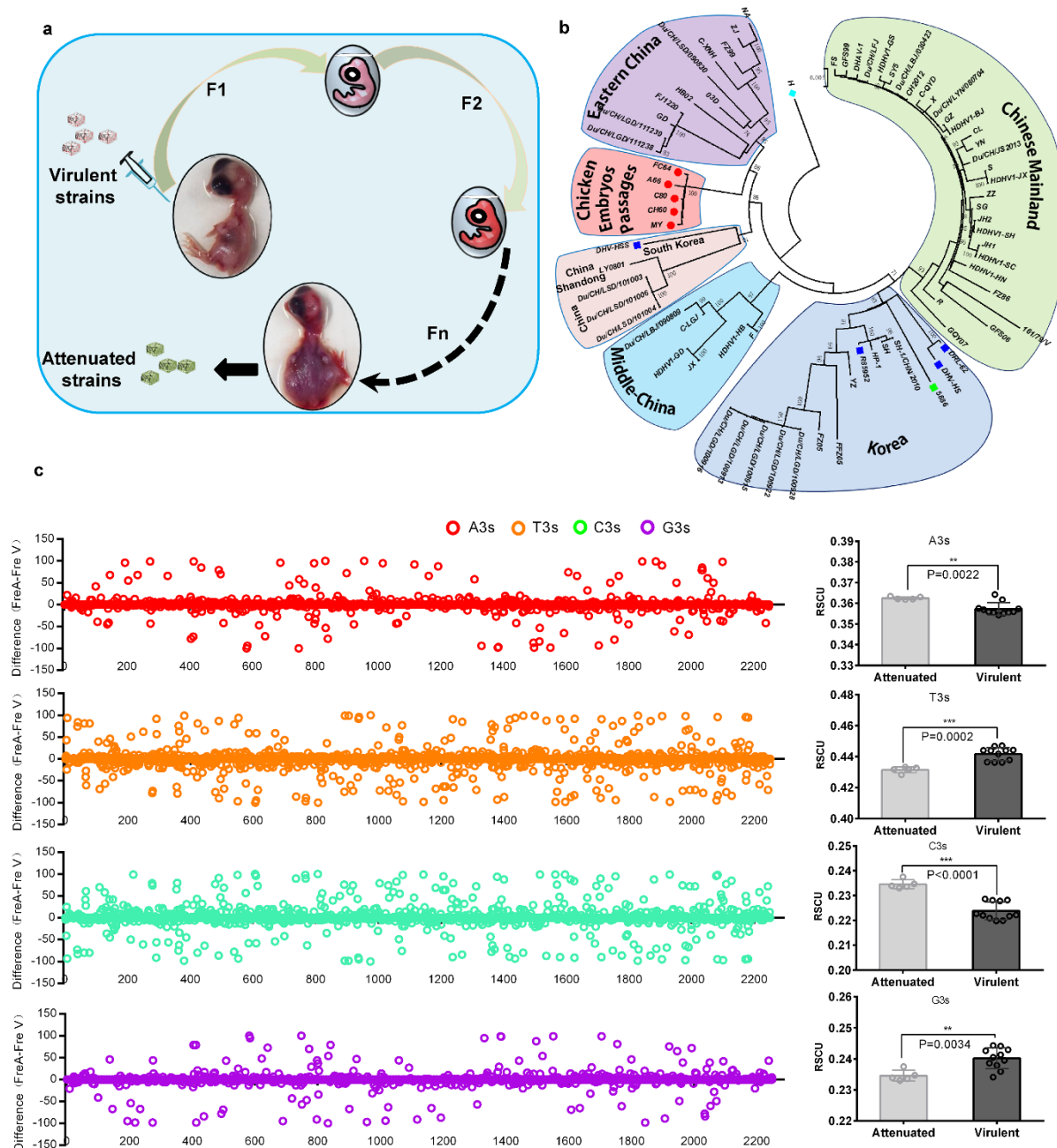
Figure 1 Overview of viral attenuation in chicken embryos and host selection on virus. a, Virulent strains isolated from ducklings can be attenuated through series passaging in chicken embryos, which accompanied by seriously histopathological injury after 60 passages(Hwang and E. Dougherty, 1962). b, Phylogenetic analysis indicated virulent strains isolated from different regions caused a convergent evolution under similar host selection-chicken embryos. In addition, genetic distance was much closer than virulent strains. The evolutionary tree was inferred using the Neighbor-Joining method with 1000 bootstrap test. c, Thymine at the 3rd codon position (T3s), A3s, G3s，C3s between the attenuated and virulent strains were compared in whole genome and sum (Left and right). The differences were calculated by N3s frequencies of attenuated strains minus N3s frequencies of virulent strains (Fre A-Fre V). The total N3s were also compared by Student T test. *P < 0.05, **P < 0.01 or ***P < 0.001.
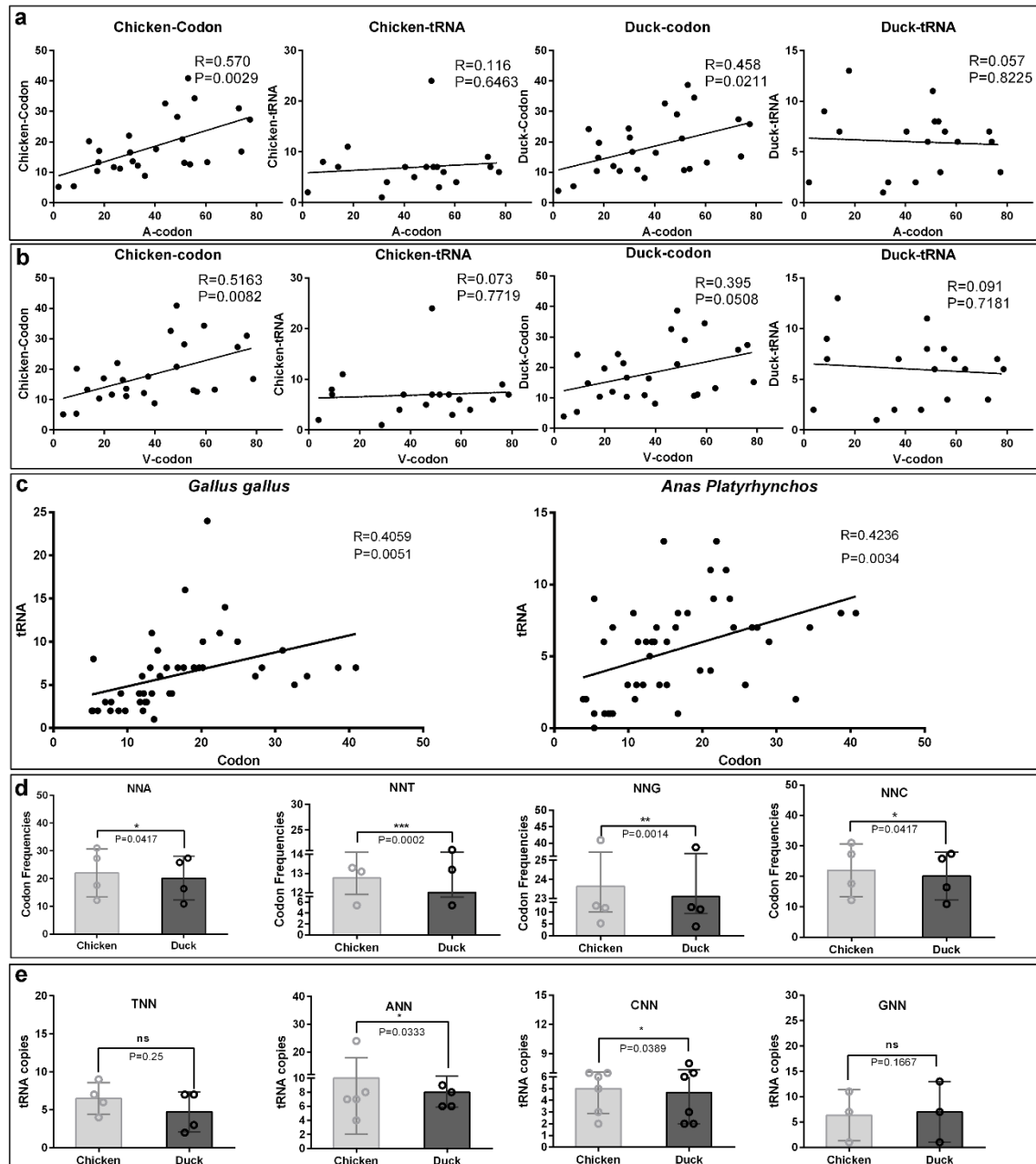
Figure 2 Comparative analysis of codon usage frequencies and tRNA copies in chickens and ducks. To elucidate the potential mechanism of codons bias in chick embryo attenuated strains, the corresponding RSCU and tRNA copies in chickens and ducks were analyzed. **a**, Those significant changed codons in the attenuated strains were correlated with corresponding codon usage frequencies and tRNA copies in chicken and ducks. Correlation index (R) and significant levels (P values) were also listed. **b**, The counterparts in the virulent strains were also analyzed by the same methods. **c**, Correlation between codon usage frequencies and tRNA copies in chicken and duck were both significant correlated (P < 0.05). **d**, NNA, NNT, NNG and NNC used in chicken and duck were compared. **e**, tRNA[TNN], tRNA[ANN], tRNA[CNN] and tRNA[GNN] used in chicken and ducks were compared. The difference of each groups were analysis by Wilcoxon matched-pairs signed-ranks test. *P < 0.05, **P < 0.01 or ***P < 0.001.

Tertiary structural variations in both virulent strains and chicken embryo attenuated strains

To further understand structural variations of viral proteome caused by those missense mutations, comparative homology modeling was used. In P1-coded viral capsid (P1), four substitutes (T3S, E205K, R217K, D234N) in the VP1 and three substitutes (P55L, T163A, A168T) in the VP0 were displayed in the **supplementary Figure 2**. These mutations are mainly located within the interface between VP1 and VP0. Except for the rest of P2 region, $2C^{pro}$ was the only one truncated into three fragments to build its tertiary structure, including $2C_{1-137}$ similar to Minichromosome Maintenance Protein (MCM) terminal, $2C_{138-264}$ to major MCM, $2C_{265-333}$ to central domain of MCM (**supplementary Figure 3d**). The mutation H142Y nearby G4 box in $2A_2^{pro}$, which changed from a basic aa to an aromatic aa, may attribute to the effect of maladaptive chick embryos. In the P3-region, only one substitute (E30G) in $3B^{pro}$ and four substitutes (G46E, C89S, D91E and L434F) in $3D^{pro}$ were identified. Those mutations focus on the surface loop of $3D^{pro}$ except for the central alpha-helix of $2B^{pro}$ (**Supplementary Figure 4d**).

Lower expression levels of viral nonstructural proteins in duck liver and kidney infected with chicken embryo attenuated DHAV-1

To experimentally verify the impaired expression of viral proteins of the chicken embryo attenuated DHAV-1, we performed infection assay in ducks and evaluated the expression levels of nonstructural proteins in the liver and kidney. We observed significantly lower expression levels of four viral proteins in chicken embryo attenuated strains infected liver and kidney compared to virulent strains infected tissues, despite with similar virus loads (**Figure 3 a-s**). The expression patterns of these viral nonstructural proteins in those two groups were also different, reflecting the stronger invasiveness of the virulent strains in the liver (**Figure 3 a-h**). Additionally, clear steatosis was observed in the virulent virus infected liver but not in chicken embryo attenuated virus infected liver (**Figure 3 a-h**).

DISCUSSION

If viruses survived in an unsuitable host, they likely "endure" and then "enjoy" the selections. The remarkable adaptation of some viruses to new hosts is heavily dependent on the generation of *de novo* mutations. The mutation rates vary tremendously among different types of viruses. RNA viruses in particular with a single-stranded genome mutate faster than DNA viruses. As a positive single-stranded RNA virus, the virulent DHAV-1 strains isolated from different geographic regions indeed undergo a convergent evolution driven by passaging in chicken embryos revealed by our phylogenetic analysis. During this process, different types of mutations, including synonymous and non-synonymous mutations in the coding region and mutations in the UTR, have emerged. Conceivably, these mutations co-ordinately drive the adaptation to the environment (**Figure 4**).

In respect to the successful development of the live attenuated DHAV-1 vaccine, we believe that the rapid emerging of many synonymous mutations (85.57%) is essential. Although these mutations do not change the amino acid sequence of the viral proteins, they highly regulate the efficiency of gene translation according to codon usage bias. We have previously identified that virulent and attenuated DHAV-1 strains have different codon usage bias (Ou et al., 2017a). In this study, we further revealed a higher correlation between viral codon usage of the attenuated strains with chicken codon usage, compared to the counterpart of virulent strains with duck. The universal and basic driving force is translation, because viral gene expression highly depends on the codon usage bias of the hosts (Ran et al., 2014;Powell and Dion, 2015). Previous studies have also demonstrated that preferred codon frequencies in highly expressed genes correlate with tRNA abundances (Novoa and de Pouplana, 2012;Powell and Dion, 2015). We found that significant more A3s and C3s in attenuated than virulent strains indeed couple with more NNA and NNC codon frequencies in chicken, but the corresponding tRNA copies (tRNA^{UNN} and tRNA^{GNN}) were not higher in chicken. A possible explanation could be that tRNA^{ANN} gene copies may have part or indirect impact on codon usage bias.

Importantly, we experimentally confirmed that viral protein expression in the liver and kidney is significantly lower with distinct expression patterns in ducks inoculated with the attenuated strain compared to the inoculation of virulent strain, despite with similar viral titers. Our findings are consistent with previous studies that viral virulence is decreased by the incompatible translation in the process of passages (Ou et al., 2017a) and preferred codon frequencies are correlated with high level of gene expression (Ran et al., 2014;Powell and Dion, 2015).

Besides, the missense mutations in coding region and mutations in UTR may also contribute to viral attenuation of DHAV-1. The previous study reported that that accumulated amino acid changes in capsid protein of Calicivirus resulted in disappearance of a helix structure, and thus a new phenotype (Nilsson et al., 2003). We observed that the capsid variations are mainly focus on the interface of VP1/0/3, which may lead to a variable spatial organization in viral morphologenisis (Nilsson et al., 2003;Wen et al., 2015). It has been reported that mutations in both 2B protein and 2C protein of hepatitis A virus are involved in adaptation in cell culture, especially in 3889 (Ala-Val), 4087 (Lys-Met), and 4222 (Phe-Ser) (Emerson et al., 1992). However, only one substitute Ala-Val (329, GCA-GTA) in 2A3 protein was identified in DHAV-1 virulent strains that is very close to 2B protein or 2C protein. While a single surface mutation (Arg-Gly) in NS5B of hepatitis C virus which is similar to 3D polymerase of Picoranvirus increases the replication in cell line (Lohmann et al., 2001). As identified in this study, all these mutations take place at the surface of 3D polymerase (one in palm, one in thumb, two in fingers). These viral-coded proteases interact with other viral proteins or host cell factors, which are vital for viral survival and shunting down host immune responses (Yang et al., 2007;Qu et al., 2011). The 5'UTR of *Picornaviridae* locates an Internal Ribosome Entry Site (IRES), which is vital for viral translation. We found that mutations in 5'UTR but not in 3'UTR

cause highly secondary structural variations, which may lead to a translation attenuation due to alterations of stem-loop structures (Ochs et al., 2003;Souii et al., 2013).
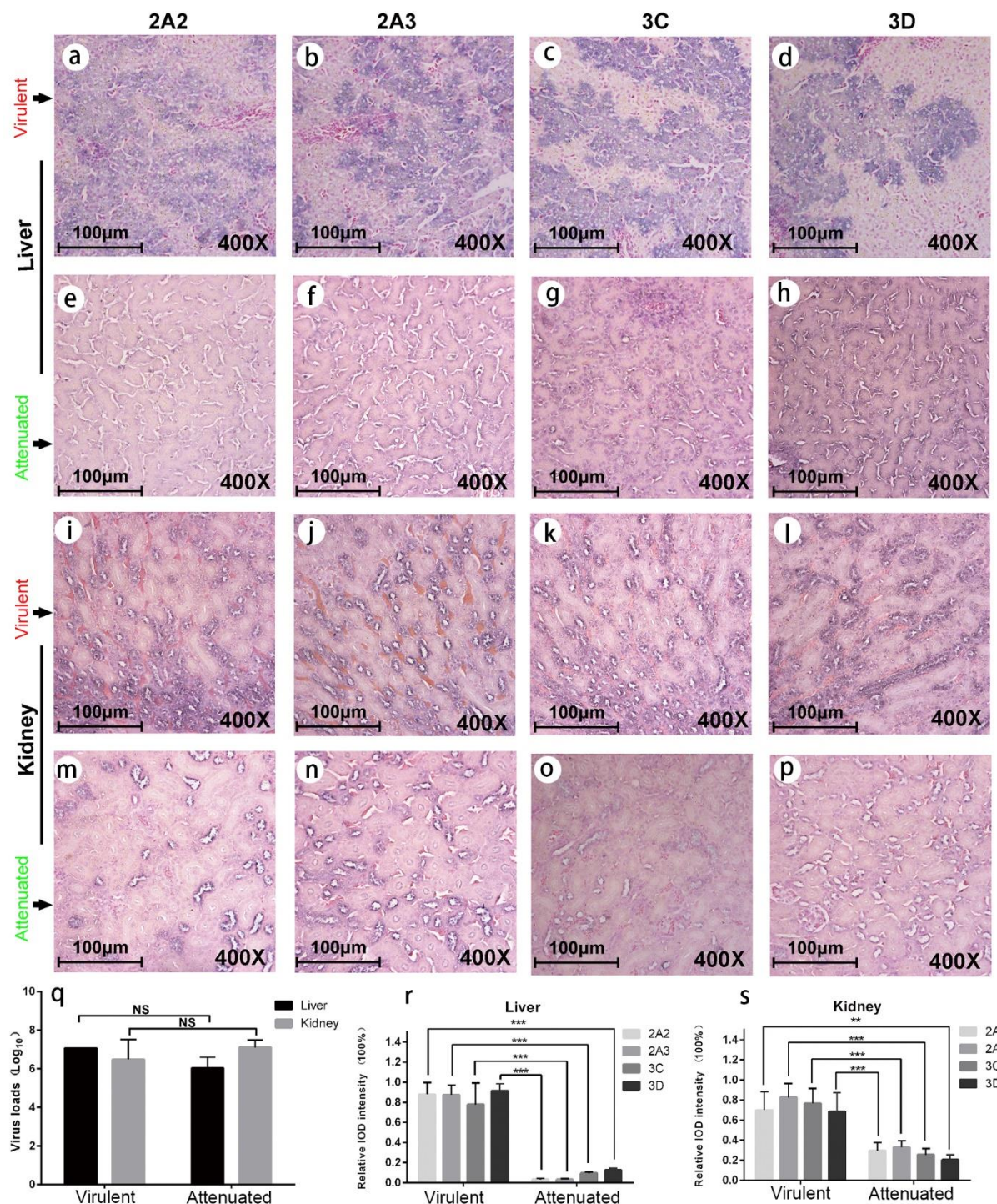

Figure 3 Viral protein expression in virulent and attenuated strains infected liver and kidney. IHC staining of $2A_2^{pro}$, $2A_3^{pro}$, $3C^{pro}$ and $3D^{pro}$ (left to right in each row, respectively) in liver and kidney infected with virulent virus and chicken embryo attenuated virus respectively. The first and second rows display the above four viral protein expressions in liver infected with virulent virus (a-d) and attenuated virus (e-h). The third and fourth rows display those four viral protein expressions in kidney infected with virulent virus(i-l) and attenuated virus(m-p). The virus loads in both liver and kidney were not significantly different between two groups (q).

However, the attenuated strain shows less viral protein expression in duck liver and kidney when compared to a virulent strain (r/s).
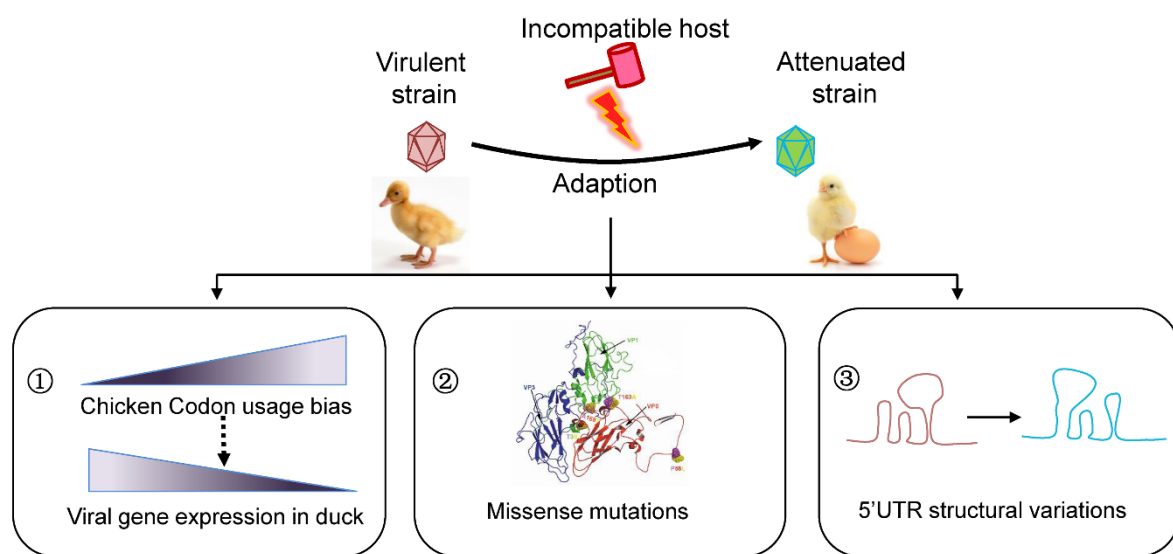
Figure 4 Schematic model for viral attenuation in chicken embryos.
Series of viral passaging in chicken embryo is very effective for development of an attenuated vaccine. During the passaging, different types of mutations will occur. Synonymous mutations do not change the amino acid sequences, but they significantly change the codon usage bias that highly regulate the efficiency of gene translation. In fact, codon usage bias of duck virulent strains skewed to the counterparts in chicken after series of passaging, which is essential to increasing viral gene expression in a given host. Of note, the missense mutations in coding region and mutations in untranslated regions may also contribute to viral attenuation of DHAV-1 to some extent. Consequently, the virulence will be attenuated when inoculate to duck, the origin host.

Vaccine development is essential in combating infections in human and animals. The live attenuated vaccines developed by series of passages in a given host are widely used since the recent half century (Bhamarapravati and Sutee, 2000;Belshe et al., 2007;Hviid et al., 2008;Song et al., 2013;Minor, 2016). However, developing such a vaccine is laborious and time-consuming. Thus, understanding the mechanisms of viral attenuation will hopefully provide a simple and universal way to vaccine development. In fact, codon deoptimization has been attempted in the development of live attenuated Influenza A and respiratory virus vaccines (Nogales et al., 2014;Cox et al., 2015;Fan et al., 2015;Stobart and Moore, 2015). However, these innovative approaches in vaccine development have been not widely explored and fundamental research is urgently required in parallel to further understand the mechanisms of viral attenuation.

In summary, we have demonstrated that virulent DHAV-1 strains undergo a convergent evolution in the chicken embryos during passaging. Consequently, their codon usage bias skewed to the counterparts in *Gallus gallus*, which is essential to viral attenuation drive by incompatible translation. These knowledges are important for the future development of innovative approaches in designing live attenuated vaccines.

References

Appel, N., Zayas, M., Miller, S., Krijnse-Locker, J., Schaller, T., Friebe, P., Kallis, S., Engel, U., and Bartenschlager, R. (2008). Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly. *PLoS Pathog* 4**,** e1000035.

Belshe, R.B., Edwards, K.M., Vesikari, T., Black, S.V., Walker, R.E., Hultquist, M., Kemble, G., and Connor, E.M. (2007). Live attenuated versus inactivated influenza vaccine in infants and young children. *New England Journal of Medicine* 356**,** 685-696.

Bhamarapravati, N., and Sutee, Y. (2000). Live attenuated tetravalent dengue vaccine. *Vaccine* 18**,** 44-47.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., and Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 42**,** W252-258.

Cao, J., Ou, X., Zhu, D., Ma, G., Cheng, A., Wang, M., Chen, S., Jia, R., Liu, M., Sun, K., Yang, Q., Wu, Y., and Chen, X. (2016). The 2A2 protein of Duck hepatitis A virus type 1 induces apoptosis in primary cell culture. *Virus Genes***,** 1-9.

Cao, Q., Cheng, A., and Wang, M. (2012). Characteristics and function of 3D gene and its encoding protein in picornavirus. *Reviews in Medical Microbiology* 23**,** 18-22.

Cheng, A., Liao, D., Xie, J., and Chen, X. (1993). A research on Duck viral hepatitis: pathogen isolation, identification and cultivation of attenuated strains. *China Veterinary Journal***,** 3-4.

Cheng, A., Wang, M., Xin, H., Zhu, D., Li, X., Chen, H., Jia, R., and Yang, M. (2009). Development and application of a reverse transcriptase polymerase chain reaction to detect Chinese isolates of duck hepatitis virus type 1. *Journal of Microbiological Methods* 76**,** 1-5.

Cox, A., Baker, S.F., Nogales, A., Martinez-Sobrido, L., and Dewhurst, S. (2015). Development of a mouse-adapted live attenuated influenza virus that permits in vivo analysis of enhancements to the safety of live attenuated influenza virus vaccine. *J Virol* 89**,** 3421-3426.

DeLano, W. L. (2002). The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA.

Emerson, S., Huang, Y.K., Mcrill, C., Lewis, M., and Purcell, R. (1992). Mutations in both the 2B and 2C genes of hepatitis A virus are involved in adaptation to growth in cell culture. *Journal of virology* 66**,** 650-654.

Fan, R.L., Valkenburg, S.A., Wong, C.K., Li, O.T., Nicholls, J.M., Rabadan, R., Peiris, J.S., and Poon, L.L. (2015). Generation of Live Attenuated Influenza Virus by Using Codon Usage Bias. *J Virol* 89**,** 10762-10773.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R., and Hofacker, I.L. (2008). The Vienna RNA Websuite. *Nucleic Acids Research* 36**,** W70-W74.

Hanson, G., and Coller, J. (2017). Codon optimality, bias and usage in translation and mRNA decay. *Nature Reviews Molecular Cell Biology* 19**,** 20.

Hu, Q., Zhu, D., Ma, G., Cheng, A., Wang, M., Chen, S., Jia, R., Liu, M., Sun, K., Yang, Q., Wu, Y., and Chen, X. (2016). A one-step duplex rRT-PCR assay for the simultaneous detection of duck hepatitis A virus genotypes 1 and 3. *Journal of Virological Methods* 236**,** 207-214.

Hviid, A., Rubin, S., and Muhlemann, K. (2008). Mumps. *Lancet* 371**,** 932-944.

Hwang, J., and E. Dougherty, R. (1962). Serial Passage of Duck Hepatitis Virus in Chicken Embryos. *Avian Diseases* 6**,** 435-440.

Kim, M.C., Kwon, Y.K., Joh, S.J., Lindberg, A.M., Kwon, J.H., Kim, J.H., and Kim, S.J. (2006). Molecular analysis of duck hepatitis virus type 1 reveals a novel lineage close to the genus Parechovirus in the family Picornaviridae. *J Gen Virol* 87**,** 3307-3316.

Lohmann, V., Körner, F., Dobierzewska, A., and Bartenschlager, R. (2001). Mutations in hepatitis C virus RNAs conferring cell culture adaptation. *Journal of virology* 75**,** 1437-1449.

Luan, Q., and Nolen, B.J. (2013). Structural basis for regulation of Arp2/3 complex by GMF. *Nat Struct Mol Biol* 20**,** 1062-1068.

Mao, S., Ou, X., Zhu, D., Chen, S., Ma, G., Wang, M., Jia, R., Liu, M., Sun, K., Yang, Q., Wu, Y., Chen, X., and Cheng, A. (2016). Development and evaluation of indirect ELISAs for the detection of IgG, IgM and IgA1 against duck hepatitis A virus 1. *J Virol Methods* 237**,** 79-85.

Miller, J.M., Arachea, B.T., Epling, L.B., and Enemark, E.J. (2014). Analysis of the crystal structure of an active MCM hexamer. *Elife* 3**,** e03433.

Minor, P.D. (2016). An Introduction to Poliovirus: Pathogenesis, Vaccination, and the Endgame for Global Eradication. *Poliovirus: Methods and Protocols***,** 1-10.

Nguyen, V.S., Jobichen, C., Tan, K.W., Tan, Y.W., Chan, S.L., Ramesh, K., Yuan, Y., Hong, Y., Seetharaman, J., Leung, K.Y., Sivaraman, J., and Mok, Y.K. (2015). Structure of AcrH-AopB Chaperone-Translocator Complex Reveals a Role for Membrane Hairpins in Type III Secretion System Translocon Assembly. *Structure* 23**,** 2022-2031.

Nilsson, M., Hedlund, K.O., Thorhagen, M., Larson, G., Johansen, K., Ekspong, A., and Svensson, L. (2003). Evolution of Human Calicivirus RNA In Vivo: Accumulation of Mutations in the Protruding P2 Domain of the Capsid Leads to Structural Changes and Possibly a New Phenotype. *Journal of Virology* 77**,** 13117-13124.

Nogales, A., Baker, S.F., Ortiz-Riaño, E., Dewhurst, S., Topham, D.J., and Martínez-Sobrido, L. (2014). Influenza A virus attenuation by codon deoptimization of the NS gene for vaccine development. *Journal of virology* 88**,** 10525-10540.

Novoa, E.M., and De Pouplana, L.R. (2012). Speeding with control: codon usage, tRNAs, and ribosomes. *Trends in Genetics* 28**,** 574-581.

Novoa Eva Maria, Pavon-Eternod Mariana, Pan Tao, and Lluís, R.D.P. (2012). A Role for tRNA Modifications in Genome Structure and Codon Usage. *Cell* 149**,** 202-213.

Ochs, K., Zeller, A., Saleh, L., Bassili, G., Song, Y., Sonntag, A., and Niepmann, M. (2003). Impaired binding of standard initiation factors mediates poliovirus translation attenuation. *Journal of virology* 77**,** 115-122.

Ou, X., Mao, S., Cao, J., Cheng, A., Wang, M., Zhu, D., Chen, S., Jia, R., Liu, M., Sun, K., Yang, Q., Wu, Y., and Chen, X. (2017a). Comparative analysis of virus-host interactions caused by a virulent and an attenuated duck hepatitis A virus genotype 1. *PLOS ONE* 12**,** e0178993.

Ou, X., Mao, S., Cao, J., Ma, Y., Ma, G., Cheng, A., Wang, M., Zhu, D., Chen, S., Jia, R., Liu, M., Sun, K., Yang, Q., Wu, Y., and Chen, X. (2017b). The neglected avian hepatotropic virus induces acute and chronic hepatitis in ducks: an alternative model for hepatology. *Oncotarget* 8**,** 81838-81851.

Ou, X., Mao, S., Jiang, Y., Zhang, S., Ke, C., Ma, G., Cheng, A., Wang, M., Zhu, D., Chen, S., Jia, R., Liu, M., Sun, K., Yang, Q., Wu, Y., and Chen, X. (2016). Viral-host interaction in kidney reveals strategies to escape host immunity and persistently shed virus to the urine. *Oncotarget* 8**,** 7336-7349.

Pang, X.Y., Cao, J., Addington, L., Lovell, S., Battaile, K.P., Zhang, N., Rao, J.L., Dennis, E.A., and Moise, A.R. (2012). Structure/function relationships of adipose phospholipase A2 containing a cys-his-his catalytic triad. *J Biol Chem* 287**,** 35260-35274.

Powell, J.R., and Dion, K. (2015). Effects of Codon Usage on Gene Expression: Empirical Studies on Drosophila. *Journal of molecular evolution* 80**,** 219-226.

Qu, L., Feng, Z., Yamane, D., Liang, Y., Lanford, R.E., Li, K., and Lemon, S.M. (2011). Disruption of TLR3 signaling due to cleavage of TRIF by the hepatitis A virus protease-polymerase processing intermediate, 3CD. *PLoS Pathog* 7**,** e1002169.

Racaniello, V.R. (2013). Picornaviridae: the viruses and their replication. *Fields virology* 1**,** 453-489.

Ran, W., Kristensen, D.M., and Koonin, E.V. (2014). Coupling Between Protein Level Selection and Codon Usage Optimization in the Evolution of Bacteria and Archaea. *mBio* 5**,** e00956-00914.

Salmon, D.E. (2013). "Diseases of Poultry", in: *Duck Hepatitis.* (ed.) P.R. Woolcock. 13 ed.: Wiley).

Schwefel, D., Arasu, B.S., Marino, S.F., Lamprecht, B., Kochert, K., Rosenbaum, E., Eichhorst, J., Wiesner, B., Behlke, J., Rocks, O., Mathas, S., and Daumke, O. (2013). Structural insights into the mechanism of GTPase activation in the GIMAP family. *Structure* 21**,** 550-559.

Shen, Y., Cheng, A., Wang, M., Chen, S., Jia, R., Zhu, D., Liu, M., Sun, K., Yang, Q., and Chen, X. (2015). Development of an indirect ELISA method based on the VP3 protein of duck hepatitis A virus type 1 (DHAV-1) for dual detection of DHAV-1 and DHAV-3 antibodies. *Journal of virological methods* 225**,** 30-34.

Sholders, A.J., and Peersen, O.B. (2014). Distinct conformations of a putative translocation element in poliovirus polymerase. *J Mol Biol* 426**,** 1407-1419.

Song, C., Yu, S., Duan, Y., Hu, Y., Qiu, X., Tan, L., Sun, Y., Wang, M., Cheng, A., and Ding, C. (2013). Effect of age on the pathogenesis of DHV-1 in Pekin ducks and on the innate immune responses of ducks to infection. *Arch Virol* 159**,** 905-914.

Souii, A., Gharbi, J., and M'hadheb-Gharbi, M.B. (2013). Impaired binding of standard initiation factors eIF3b, eIF4G and eIF4B to domain V of the live-attenuated coxsackievirus B3 Sabin3-like IRES-alternatives for 5' UTR-related cardiovirulence mechanisms. *Diagnostic pathology* 8**,** 161.

Spirig, T., Malmirchegini, G.R., Zhang, J., Robson, S.A., Sjodt, M., Liu, M., Krishna Kumar, K., Dickson, C.F., Gell, D.A., Lei, B., Loo, J.A., and Clubb, R.T. (2013). Staphylococcus aureus uses a novel multidomain receptor to break apart human hemoglobin and steal its heme. *J Biol Chem* 288**,** 1065-1078.

Stobart, C.C., and Moore, M.L. (2015). Development of next-generation respiratory virus vaccines through targeted modifications to viral immunomodulatory genes. *Expert Rev Vaccines* 14**,** 1563-1572.

Sun, D., Chen, S., Cheng, A., and Wang, M. (2016). Roles of the Picornaviral 3C Proteinase in the Viral Life Cycle and Host Cells. *Viruses* 8**,** 82.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30**,** 2725-2729.

Tuthill, T.J., Groppelli, E., Hogle, J.M., and Rowlands, D.J. (2010). Picornaviruses. *Curr Top Microbiol Immunol* 343**,** 43-89.

Voitenleitner, C., Bechtel, J., Arfsten, A., and Hamatake, R. (2012). Hepatitis C genotype 1a replicon improved through introduction of fitness mutations. *Biotechniques* 52**,** 273-275.

Wang, J., Fan, T., Yao, X., Wu, Z., Guo, L., Lei, X., Wang, J., Wang, M., Jin, Q., and Cui, S. (2011). Crystal structures of enterovirus 71 3C protease complexed with rupintrivir reveal the roles of catalytically important residues. *J Virol* 85**,** 10021-10030.

Wen, X., Cheng, A., Wang, M., Jia, R., Zhu, D., Chen, S., Liu, M., Liu, F., and Chen, X. (2014). Detection, differentiation, and VP1 sequencing of duck hepatitis A virus type 1 and type 3 by a 1-step duplex reverse-transcription PCR assay. *Poultry science* 93**,** 2184-2192.

Wen, X., Cheng, A., Wang, M., Jia, R., Zhu, D., Chen, S., Liu, M., Sun, K., Yang, Q., and Wu, Y. (2015). Recent advances from studies on the role of structural proteins in enterovirus infection. *Future Microbiology* 10**,** 1529.

Yang, M., Cheng, A., Wang, M., and Xing, H. (2008). Development and application of a one-step real-time Taqman RT-PCR assay for detection of Duck hepatitis virus type1. *J Virol Methods* 153**,** 55-60.

Yang, X., Cheng, A., Wang, M., Jia, R., Sun, K., Pan, K., Yang, Q., Wu, Y., Zhu, D., and Chen, S. (2017). Structures and Corresponding Functions of Five Types of Picornaviral 2A Proteins. *Frontiers in microbiology* 8.

Yang, Y., Liang, Y., Qu, L., Chen, Z., Yi, M., Li, K., and Lemon, S.M. (2007). Disruption of innate immunity due to mitochondrial targeting of a picornaviral protease precursor. *Proc Natl Acad Sci U S A* 104**,** 7253-7258.

Zhang, Y., Cao, Q., Wang, M., Jia, R., Chen, S., Zhu, D., Liu, M., Sun, K., Yang, Q., and Wu, Y. (2017). The 3D protein of duck hepatitis A virus type 1 binds to a viral genomic 3' UTR and shows RNA-dependent RNA polymerase activity. *Virus Genes***,** 1-9.

Zhu, L., Wang, X., Ren, J., Porta, C., Wenham, H., Ekstrom, J.O., Panjwani, A., Knowles, N.J., Kotecha, A., Siebert, C.A., Lindberg, A.M., Fry, E.E., Rao, Z., Tuthill, T.J., and Stuart, D.I. (2015). Structure of Ljungan virus provides insight into genome packaging of this picornavirus. *Nat Commun* 6**,** 8316.

Supplementary information
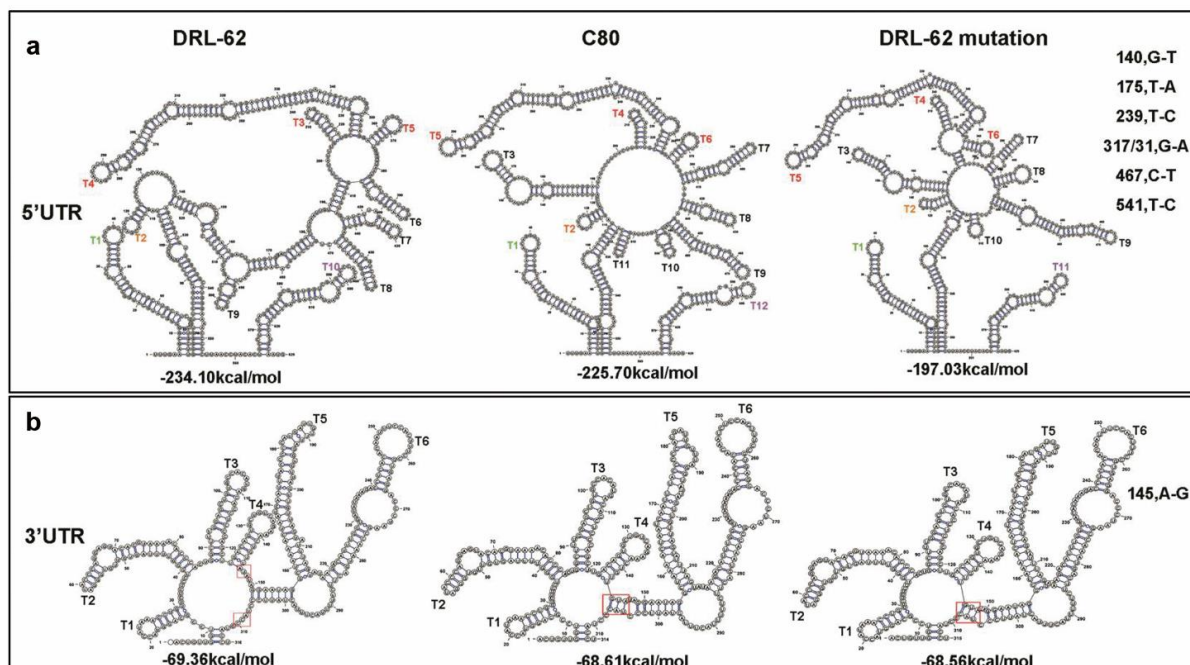
Supplementary Figure 1 Secondary structural variations of 5'UTR and 3'UTR in the process of serial passaging in chick embryos. (a) Online RNAfold web server was used to characteristic the secondary structural variation in chick embryo attenuated strains. In order to understand whether those fixed SNPs could significantly change the secondary structure of virulent strains, the 5'UTR of artificial mutant involved in the Table S2 were also imported into RNAfold web server to predict their secondary structure. (b) The same methods were also used to predict secondary structure of 3'UTR and their mutant. Apparently structural changes in 5'UTR but not in 3'UTR were identified in the DHAV-1 genome.
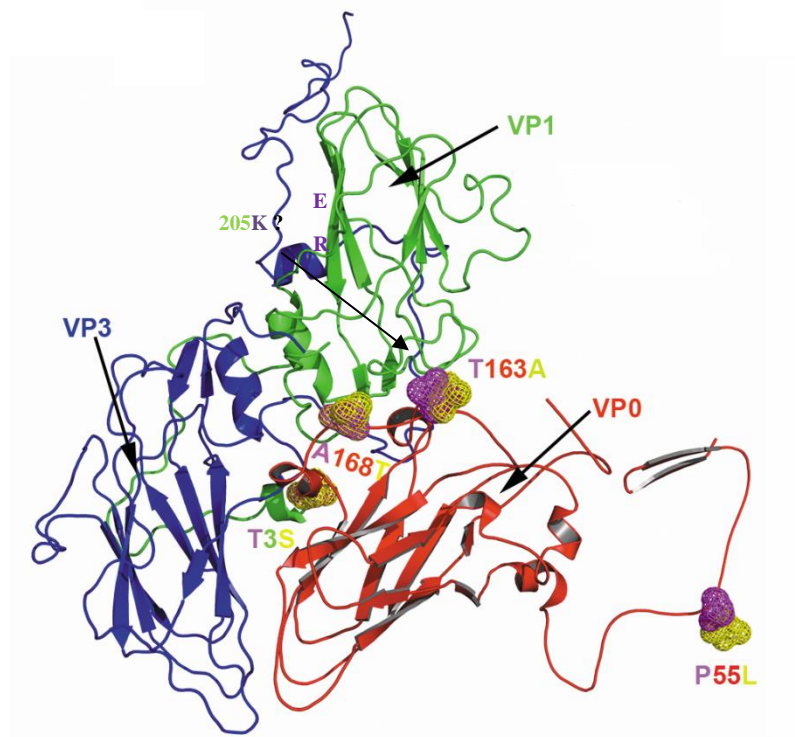
Supplementary Figure 2 Quarternary structural variations in capsid. Non-synonymous substitutes, such as T3S, were displayed using mesh model (purple to yellow) to indicated their spatial position. The color of number indicates where the non-synonymous substitutes take place, VP1(green), VP3(blue) or VP0(red). Due to the sequence difference with the modeling template, except T3S mutation of VP1, the other three C-terminal mutations (E205K, R217K and D234N) in spatial position were not available in this analysis. Three substitutes (P55L, T163A, A168T) of VP0 were also labeled. Those aa substitutes were mainly identified in the interface of VP1/0/3. All those modeling are predicted by SWISS-MODEL and visualized by Pymol software.
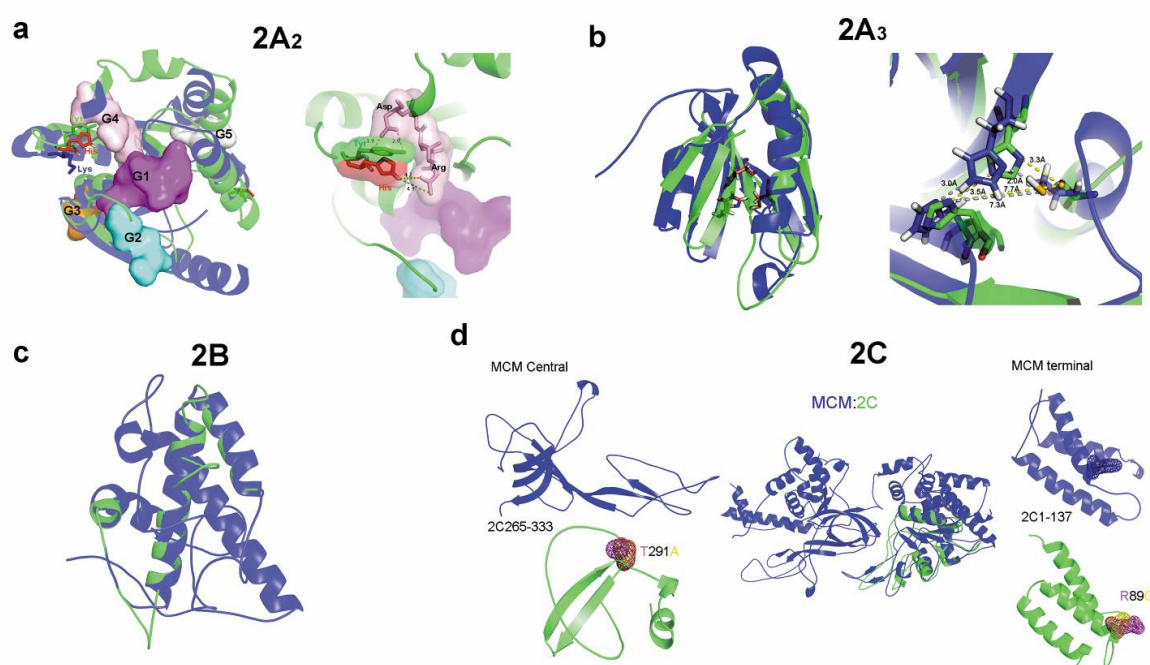
Supplementary Figure 3 Tertiary structural variations in P2-nonstructural proteins. Those nonstructural proteins (2A$_2^{pro}$, 2A$_3^{pro}$, 2B^{pro} and 2C^{pro}) coded in this region were well modeled with its template, except for 2C protein. Blue color indicates templates. **a**. Superposition of 2A2 protein with its template (GTPase, PDB: 3zjc.3.A). G1-G5 motif were labeled by different color. The mutation H142Y nearby G4 box changed from a basic aa to an aromatic aa, may attribute to the effect of maladaptive chick embryos. **b**. Superposition of 2A3 protein with Adipose phospholipase A (PDB: 4fa0.1.A). **c**. Superposition of 2B protein with Actin-related protein (PDB: 4jd2.1.E). **d**. Superposition of 2C protein with Minichromosome maintenance protein (MCM)(PDB: 4r7y.1.A). Because 2C protein were not predicted in full length, the DHAV-1 2C protein was truncated as three fragments to build its tertiary structure. Non-synonymous substitutes were also displayed using mesh model (purple to yellow) to indicated their spatial position. All those modeling are predicted by SWISS-MODEL and visualized by Pymol software.
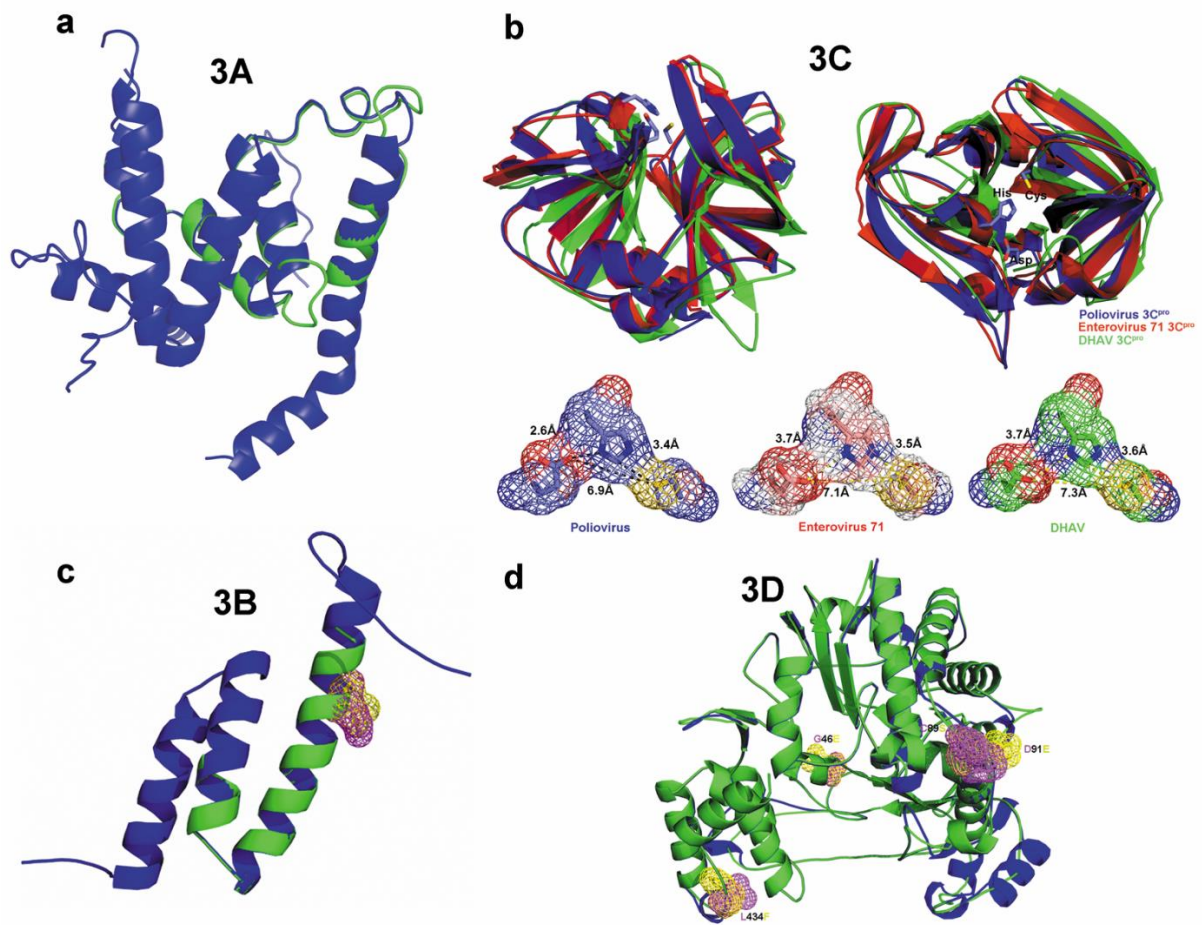
Supplementary Figure 4 Tertiary structural variations in P3 non-structure proteins. Blue color indicates templates. **a**. Superposition of 3A protein with its template (AopB, PDB: 3wxx.2.B). There were no substitutes in 3A protein. **b**. Superposition of 3C protein with Enterovirus 71 3C protein (PDB: 3qzq.4.A). The catalytic resides and their distance were also labeled. There were no substitutes in 3C protein. **c**. Superposition of 3B protein with iron-regulated surface determinant protein H (PDB: 2lhr.1.A). Non-synonymous substitutes were also displayed using mesh model (purple to yellow) to indicated their spatial position. The mutation of E30G at α-helix was labeled (purple to yellow). **d**. Superposition of 3D protein with Poliovirus 3D polymerase (PDB: 4nlr.1.A). Four substitutes (G46E, C89S, D91E and L434F) were identified on the surface loop of 3D^{pro}. All those modeling are predicted by SWISS-MODEL and visualized by Pymol software.

Supplementary Table 1 Representative viral strains in this study.

Groups	Strains	Accession No.	Comments
Chicken embryo	CH60	KU923754.1	60 passages (china, Sichuan)
attenuated strains	C80	DQ864514	80 passages (china, ?)
	FC64	HQ232302	64 passages (china Fujian)
	A66	DQ886445	66 passages (china, Anhui)
	MY	GU944671.1	86 passages (china, Xichang)
	DRL-62	DQ219396.1	ATCC(Korea)
Virulent strains	R85952	DQ226541	ATCC(Korea)
	03D	DQ249299	China Taiwan
	JX	EU371557	China
	F	EU264072	China Beijing
	5886	DQ249301	USA
	H	DQ249300	UK
	HP-1	EF151312	China Ha'erbin
	ZJ	EU841005	China Zhanjiang
	S	EF417871.1	China Beijing
	CL	EF427899.1	China Shandong
	YZ	EF427900.1	China Yangzhou
	DHV-HS	DQ812094.2	South Korea
	DHV-HSS	DQ812092.1	South Korea
	JH2	EU395435.1	China Shandong
	JH1	EU395436.1	China
	YN	EU395437	China
	FS	EU395438	China Foshan
	ZZ	EU395439.1	China Shandong
	DHAV-1	EU395440.1	China
	GZ	EU888310	China Guangzhou
	HDHV1-BJ	FJ157172.1	China
	HDHV1-JX	FJ157173.1	China
	HDHV1-SC	FJ157174.1	China
	HDHV1-GS	FJ157175.1	China
	HDHV1-HN	FJ157177.1	China
	HDHV1-SH	FJ157178.1	China
	HDHV1-GD	FJ157179.1	China
	HDHV1-HB	FJ157180.1	China
	LY0801	FJ436047	China Shandong
	FFZ05	FJ496340	China Fujian
	GFS06	FJ496341.1	China Fujian
	GQY07	FJ496342.1	China Guangdong
	X	FJ496343.1	China Sichuan
	GFS99	FJ496344.1	China Guangdong
	SG	FJ971623	China Shandong
	NA	GQ130377	China Fuzhou
	C-LGJ	GU066819	China
	C-XNH	GU066820	China
	C-QYD	GU066825	China
	SH-1	HQ232303.1	China
	SH	HQ265433.1	China Shanghai
	Du/CH/LBJ	JF828982	China
	Du/CH/LFJ	JF828983	China
	duck/CH/LGD/100913	JF828984	China

duck/CH/LGD/100915	JF828985	China
duck/CH/LGD/100916	JF828986	China
duck/CH/LGD/100922	JF828987	China
duck/CH/LGD/100928	JF828988	China
duck/CH/LSD/090830	JF828989	China
duck/CH/LSD/101003	JF828990	China
duck/CH/LSD/101004	JF828991	China
duck/CH/LSD/101006	JF828992	China
duck/CH/LYN/080704	JF828993	China
duck/CH/LBJ/090809	JF828997	China
HB02	JQ031262	China Hebei
Duck/LGD/111238	JQ804521	China
Duck/LGD/111239	JQ804522	China
FZ86	JX390982	China Fuzhou
FZ05	JX390983	China Fuzhou
FZ99	JX390984	China Fuzhou
FJ1220	KC904272	China Fujian (Pigeon)
161/79/V	EU753359	China Jiangsu
R	EF585200	China Guangzhou
SY5	JQ808453	China Zhejiang
Du/CH/JS2013	KP721458	China Jiangshu
GD	KM017068	China Guangdong
CH2012	KF953535	China

Supplementary Table 2 Fixed SNPs in the genome of attenuated and virulent strains.

Gene	5'UTR	VP0	VP3	VP1	2A1	2A2	2A3
Synonymous mutation	137,G-T 172,T-A 236,T-C 314/28,G-A 464,C-T 538,T-C	582,GTG-GTA 645,GTG-GTC	54,CAG-CAA 105,AAC-AAT 318,TTC-TTT 372,TTC-TTT 471,GTG-GTA 705,AAT-AAC	16/8,TTG-CTA 174,GTT-GTC 180,GAC-GAT 267,GCA-GCG 273,GTA-GTG 342,ATT-ATC 348,GCT-GCC 588,GGG-GGA	51,CCA-CCG 57,CCG-CCA	48,TGT-TGC 138,GAG-GAA 168,CCT-CCC 246,AAG-AAA 462,GCC-GCT	72,GTC-GTT 78,CAC-CAT 135,CAG-CAA 189,GCC-GCT 312,GAG-GAA
Missense mutation	None	164,CCA-CTA (55,Pro-Leu) 487,ACT-GC/AT (163,Thr-Ala/Asp) 502,GCA-ACA (168,Ala-Thr)	None	7,TCC-ACC(3,Ser-Thr) 613,GAA-AAA (205,Glu-Lys) 650,AGA-AAA (217,Arg-Lys) 700,AAT-GAT (234,Asn-Asp)	None	280,GGC-AGC (94,Gly-Ser) 424/6,CAC-TAT (142,His-Tyr)	329,GCA-GTA(110,Ala-Val)
Nonsyn/Syn		3/2=1.5	0/6=0	4/9=0.44	0/2=0	3/5=0.6	1/5=0.2

Gene	3'UTR	2B	2C	3A	3B	3C	3D
Synonymous mutation	144A-G	126,CTT-CTC 183,TTA-CTA 243,CTG-CTA 324,ACT-ACC	159,CAT-CAC 246,AAC-AAT 528,GGA-GGG 672,GCT-GCC 684,CAA-CAG 690,AGA-AGG 738,TGT-TGC 792,ATC-ATT 843,GCC-GCT	33,ACA-ACT 82,84,TTA-CTC 189,CGA-CGG 231,ATT-ATC	45,GAT-GAC 79,CTA-TTA	33,AAC-AAT 270,ACA-ACG 291,GGC-GGT 486,CAC-CAT	96,CCC-CCT 144,AAG-AAA 252,TCT-TCC 327,GAC-GAT 402,GTT-GTC 552,GAC-GAT 849,TCT-TCC 858,ACC-ACT 861,ACT-ACC 912,GTC-GTA 1128,ACC-ACT
Missense mutation	None	None	112/4,GTG-ATA (38,Val-Ile) 265,CGT-TGT (89,Arg-Cys) 871,ACT-GCT (291,Thr-Ala)	None	89,GAA-GGA (30,Glu-Gly)	None	137,GGG-GAG(46,Gly-Glu) 26265,TGC-AGC(89,Cys-Ser) 27273,GAT/C-GAA(91,Asp-Glu) 131300,CTT-TTT(434,Leu-Phe)
Nonsyn/Syn		0/4=0	3/9=0.33	0/4=0	1/2=0.5	0/4=0	4/11=0.36

Chapter 8

Tracing genetic signatures of bat-to-human coronaviruses and early transmission of North American SARS-CoV-2

Xumin Ou, Zhishuang Yang, Dekang Zhu, Sai Mao, Mingshu Wang, Renyong Jia, Shun Chen, Mafeng Liu, Qiao Yang, Ying Wu, Xinxin Zhao, Shaqiu Zhang, Juan Huang, Qun Gao, Yunya Liu, Ling Zhang, Maikel P Peppelenbosch, Qiuwei Pan, Anchun Cheng

Abstract

Highly pathogenic coronaviruses, including SARS-CoV-2, SARS-CoV, and MERS-CoV, are thought to be transmitted from bats to humans, but the viral genetic signatures that contribute to bat-to-human transmission remain largely obscure. In this study, we identified an identical ribosomal frameshift motif among the three bat-human pairs of viruses and strong purifying selection after jumping from bats to humans. This represents genetic signatures of coronaviruses that are related to bat-to-human transmission. To further trace the early human-to-human transmission of SARS-CoV-2 in North America, a geographically stratified genome-wide association study (North American isolates and the remaining isolates) and a retrospective study were conducted. We determined that the single nucleotide polymorphisms (SNPs) 1059.C>T and 25563.G>T were significantly associated with approximately half of the North American SARS-CoV-2 isolates that accumulated largely during March 2020. Retrospectively tracing isolates with these two SNPs was used to reconstruct the early, reliable transmission history of North American SARS-CoV-2, and European isolates (February 26, 2020) showed transmission 3 days earlier than North American isolates and 17 days earlier than Asian isolates. Collectively, we identified the genetic signatures of the three pairs of coronaviruses and reconstructed an early transmission history of North American SARS-CoV-2. We envision that these genetic signatures are possibly diagnosable and predictable markers for public health surveillance.

Keywords: genetic signatures; SARS-CoV-2; geographic transmission; GWAS

Introduction

Since 2003, coronaviruses (CoVs), specifically, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), severe acute respiratory syndrome-related coronavirus (SARS-CoV, 2003) and Middle East respiratory syndrome-related coronavirus (MERS-CoV, 2012), have caused three epidemics in human populations worldwide, including the ongoing COVID-19 pandemic caused by SARS-CoV-2 (Perlman, 2020; P. Zhou et al., 2020). Interestingly, these CoVs are thought to be associated with bat CoVs; for instance, human SARS-CoV-2 shares 96% nucleotide identity with bat CoVs (W. Li et al., 2005; P. Zhou et al., 2020). Human-to-human transmission of SARS-CoV-2 was confirmed on 14 Jan 2020 (Wu, Leung, & Leung, 2020). On 11 March 2020, the World Health Organization (WHO) officially declared that human COVID-19 caused by SARS-CoV-2 was a global pandemic. As of May 31, over 5.93 million human cases have been confirmed globally, of which over 2.74 million cases were from America, particularly North America (Wu et al., 2020). Thus, it is very important to understand the cross-species transmission of SARS-CoV-2 and its human-to-human transmission in America to inform public health measures (Ji, Ma, Peppelenbosch, & Pan, 2020).

It is well known that viral genetic variations are associated with many aspects of virology, most notably viral infectivity, and zoonotic transfer (Ou, Cao, Cheng, Peppelenbosch, & Pan, 2019; J. H. Zhou et al., 2019). Identifying distinctive genetic signatures of CoVs common to those viruses found in different host species (human and bat) as well as different geographic regions (North America and the rest) may provide differential markers to support epidemiological surveillance. However, these distinctive signatures are currently unknown. SARS-CoV-2 massively mutates, which hampers its transmission tracing (Tang et al., 2020). Herein, we aimed to identify the genetic signatures of three pairs of bat-to-human CoVs as well as those of the North American SARS-CoV-2 isolates associated with the early human-to human transmission history of COVID-19. The whole genomic sequences of three bat-human pairs of SARS-CoV-2 were analysed as well as bat-human pairs of SARS-CoV and MERS-CoV; the latter also included a MERS-CoV strain isolated from camel (Azhar et al., 2014). After a virus jumps to a human host, new mutations are fixed in the viral genome that may be geographically different. Identification of these fixed and common mutations remains a great challenge because of the complexity of these mutable viruses. In human population genetics studies, this type of complexity can be particularly addressed by genome-wide association studies (GWASs) (Power, Parkhill, & de Oliveira, 2017). Therefore, we aimed to use a geographically stratified GWAS to address the complexity of global SARS-CoV-2 isolates.

We primarily found that the genomic organization of the three human CoVs was similar to that of their paired bat CoVs within lineages and underwent strong purifying selection after jumping to the human host. For the early human-to-human transmission of North American SARS-CoV-2, we identified that two SNPs of complete linkage disequilibrium were exclusively present in more than half of the North American SARS-CoV-2-dominated lineage B.1. By

retrospectively tracing isolates with these two SNPs, an early transmission history of North American SARS-CoV-2 isolates was reconstructed.

Results

Genomic organization of bat-human CoV pairs

CoVs are positive single-stranded RNA viruses with a nonsegmented genome. The genome encodes a fixed array of necessary proteins (NPs), in the order ORF1ab, spike (S) protein, envelope (E) protein, membrane glycoprotein (M) protein, and nucleocapsid (N), as well as accessory proteins (APs) that differ by number and order among closely related CoVs (**Fig. 1**). We found that the genomic organization of each bat-human CoV pair was similar，as well as that of MERS-CoV between humans, bats and camels. For the NPs, the genomic organization among all three CoVs followed the same order (i.e., ORF1ab-S-E-M-N) (**Fig. 1**). The loci of the APs were largely different, which was likely caused by gene recombination (**Fig. 1**). Specifically, for SARS-CoV-2 and SARS-CoV, the ORF6, ORF7, and ORF8 genes are equally inserted between the M gene and the N gene. However, for MERS-CoV, this location has no gene insertion.

For the APs, the phylogenetic analysis suggested that the APs of human SARS-CoV-2 are evolutionally close to those of bat SARS-CoV-2 (i.e., ORF3a, ORF6, ORF7a, and ORF8), in which ORF6 and ORF7a are also close to those of bat or human SARS-CoV (**Fig. 1**, **Fig. S2** and **S3**). Strikingly, the human SARS-CoV-2 genome contains a novel ORF10 that is less related to any other human or bat CoV ORF.

Identical ribosomal frameshift motif between bat-human CoV pairs

For all CoVs, a programmed -1 ribosomal frameshift signal is essential, as it controls viral translation. The slippage signal is characterized by an X_3Y_3Z motif (X_3, any three identical nucleotides; Y_3, typically UUU or AAA; Z, A, C, or U). We found that the slippage signal, U_UUA_AAC, was identical among SARS-CoV-2, SARS-CoV, and MERS-CoV (Baranov et al., 2005) (**Fig. S4**). The slippage of ribosomal frameshifting from U_UUA_AAC to UUU_AAA_C does not change the growing peptides, as they both encode identical dipeptides because of the degeneracy of codon position 3. For the two flanking motifs, the 5' attenuator hairpin and 3' frameshift-stimulating three-stemmed pseudoknots are relatively conserved in both SARS-CoV-2 and SARS-CoV. However, inside the second flanking motif of MERS-CoV, an insertion of the AAT codon (encoding asparagine) was newly identified (**Fig. S4**). Prior research conducted by mutating this slippage signal to C_CUC_AAC shows thorough inhibition of the ribosomal frameshift (Kelly et al., 2020).
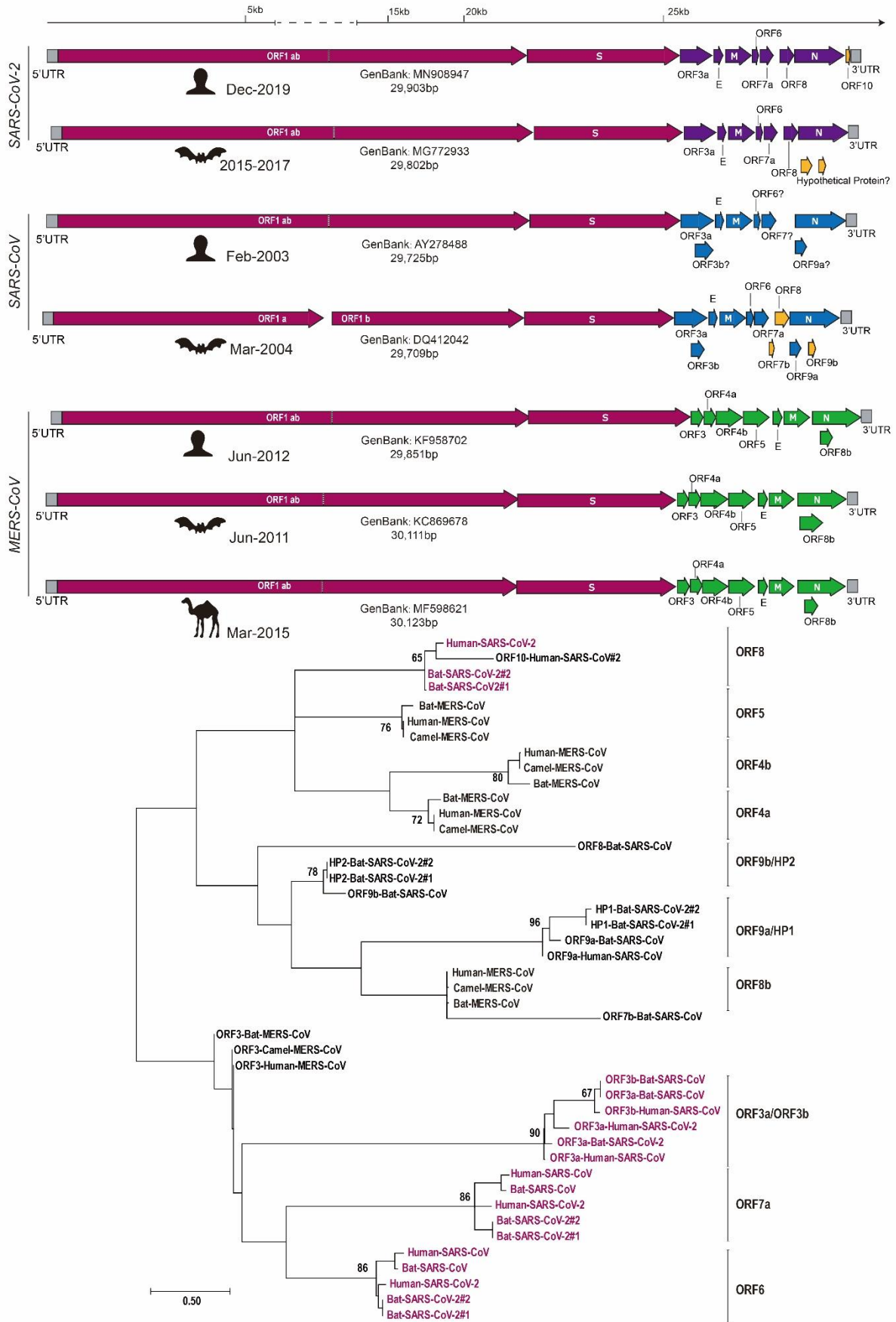
FIG 1 Genomic signatures of bat-human SARS-CoV-2, SARS-CoV, and MERS-CoV pairs. For the three CoV pairs, the necessary proteins are encoded in the same order in the genome (i.e., ORF1ab-S-E-M-N). The accessory protein-encoding genes vary by locus. SARS-CoV-2 and SARS-CoV have gene insertions between the M protein and N protein-coding genes, such as ORF6, ORF7a(b), and ORF8, while the same locus has no insertion in MERS-CoV. Of note, a novel ORF10 (yellow box) of human SARS-CoV-2 is less related to any human or CoV gene. Importantly, ORF6 and ORF7a of human SARS-CoV-2 are related to the equivalents of SARS-CoV isolated from both bats and humans (lower panel). Two new hypothetical proteins (i.e., HP1 and HP2) (yellow box) of bat SARS-CoV-2 are evolutionarily close to ORF9a and ORF9b of SARS-CoV. Bat-SARS-CoV-2 #1 and #2 represent two related bat CoVs. The phylogenetic tree was constructed by the maximum likelihood method. The evolutionary distances are calculated by base differences per site. Confidence probability was estimated using the bootstrap test (100 replicates).

Strong purifying selection of CoVs

Zoonotic transfer of CoVs involves mutagenesis and directional selection (Forni, Cagliani, Clerici, & Sironi, 2017). During the 2002–2004 epidemic, SARS-CoV mutated extensively, which enhanced its virulence (Consortium, 2004). To measure which type of mutation is linked to CoV virulence, synonymous differences (dSs) and nonsynonymous differences (dNs) were analysed between intro- and extra-branches (**Fig. 2**). The number of dSs between CoVs of intro-branches is much higher than that of dNs, which is similar among all NPs (i.e., ORF1ab-S-E-M-N) (Table S2-11, online). For instance, for ORF1ab, the number of dSs between bat and human SARS-CoV-2 (dS=1875.75) is approximately four times higher than the number of dNs (dN=441.25). The trend is the same for SARS-CoV and MERS-CoV. The number of dSs and dNs within intro-branches of CoVs is smaller than the number within extra-branches. This is consistent with the evolutionary distances of phylogenetic trees (**Fig. 2**).

Because massively synonymous mutations do not modify the protein sequence but change the overall codon usage bias, we performed linear regression of relative synonymous codon usage (RSCU) to see if bat CoVs show codon usage bias to human CoVs. However, only a slight RSCU shift from bat to human CoVs was observed, as the slope of the linear regression was slightly smaller than 1 (**Fig. S5**). Collectively, the large synonymous mutations were augmented, which may enhance SARS-CoV-2 virulence, similar to SARS-CoV, 2003.

The dN/dS ratio is a classic indicator of directional selection: a ratio above 1 implies positive selection (nature), a ratio less than 1 implies negative selection (purifying), and a ratio equal to 1 indicates no selection (neutral) (Kryazhimskiy & Plotkin, 2008). In contrast, purifying selection involves more synonymous mutations than nonsynonymous mutations. As discussed, this is true for the three bat-human CoV pairs (Table S2-11, online). Purifying selection primarily changes viral codon usage bias and thus can regulate viral virulence via optimization of a specific codon context (Coleman et al., 2008; Hanson & Coller, 2017). For ORF1ab, the dN/dS ratio between human and bat SARS-CoV-2 is 0.05, which is much lower than that of SARS-CoV-2 and MERS-CoV (Table S12-16, online). A similar trend was confirmed in the other

NPs. This is supported by a viral culture experiment, as SARS-CoV-2 grows better than SARS-CoV and MERS-CoV in human cells (Perlman, 2020). Using the codon-based Z-test of selection, the statistic shows that human SARS-CoV-2 undergoes significantly strong purifying selection (Table S12-16, online).
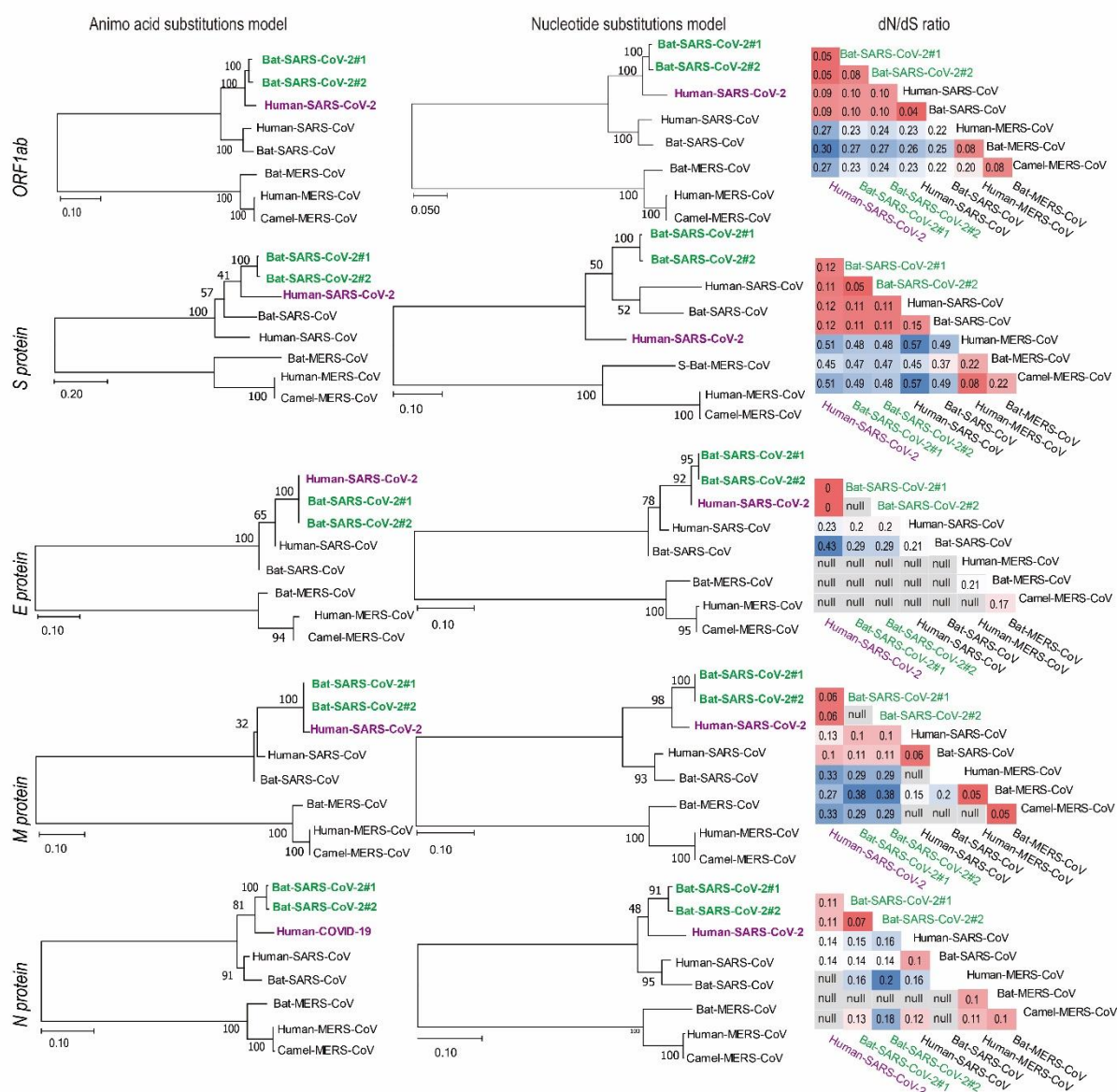

FIG 2 Evolutionary signatures of necessary protein-encoding genes of SARS-CoV-2. The differences between amino acid substitutions were used to construct phylogenetic trees (left panel). The same analysis was also performed by the differences in nucleotide substitutions (middle panel). dN/dS ratio matrixes are displayed (Table S12-16, online) (right panel). The sequential analysis of necessary proteins was performed from top to bottom (i.e., ORF1ab-S-E-M-N). The phylogenetic tree was constructed by the maximum likelihood method. Bat-SARS-CoV-2 #1 and #2 represent two related bat CoVs. The evolutionary distances are calculated by the differences between amino acid substitutions or nucleotide substitutions per site. Confidence probability was estimated using the bootstrap test (100 replicates).

Phylogenetics of global SARS-CoV-2 reveals that early North American isolates dominate lineage B.1

To understand the early human-to-human transmission of SARS-CoV-2 in North America, a phylogenetic analysis of the global SARS-CoV-2 population (2599 strains with high confidence) was conducted. We found that global SARS-CoV-2 was rooted in two lineages, lineages A (n=413) and B (n=2186), in which North American isolates dominated lineage B (n=818) and lineage B.1 (n=691) (**Fig. 3** and **Table 1**) (Rambaut et al., 2020). Importantly, the phylogenetic tree was inferred by producing mutations, and the identification of key mutations can provide clues for tracing the transmission route of SARS-CoV-2.

Geographic GWAS reveals SNPs associated with North American isolates

Calling key SNPs from massive mutations of the SARS-CoV-2 population requires a GWAS that has been learned from a human GWAS (Power et al., 2017). Because of the complexity of the phylogenetic tree, a phylogenetically stratified GWAS may not be feasible. Therefore, a geographically stratified GWAS was carried out, as the geographic location of individual isolates was reliable. The mutation features of SARS-CoV-2 between continents may reflect the incidence of emergence of a given viral population in different human hosts (Rambaut et al., 2020). By using a geographically stratified GWAS comparing North American isolates (n=1063) with the remaining isolates (n=1536), we found 21 significant SNPs or small insertion deletions (INDELs) out of 5312 (threshold p-value = 1.00×10^{-15}) (**Fig. 4A**). Specifically, the top two SNPs (i.e., 1059.C>T and 25563.G>T) were present in approximately half of North American SARS-CoV-2 isolates (479/1063 = 45% and 574/1063 = 54%), particularly North American lineage B.1 (479/691 = 69% and 573/691 = 83%) (**Table 1**). Interestingly, the two SNPs were in complete linkage disequilibrium, suggesting that the two SNPs concurrently occurred in the North American dominating lineage B.1 (479/691, 69%) (**Fig. 4B**). Importantly, the two SNPs resulted in two mutations (i.e., Thr265 Ile and Gln57 His) in ORF1ab and ORF3a, respectively.

Among these 21 SNPs, we also identified two previously reported SNPs, 8782.C>T and 28144.T>C (p-value = 4.03×10^{-28} and 9.73×10^{-33}), resulting in a synonymous mutation and a missense mutation (Leu 84 Ser) (Tang et al., 2020) (**Table 1**). Interestingly, three sequential SNP sites (28881-3.GGG>ACA) were fixed in 22% (207/951) of the European SARS-CoV-2 isolates, resulting in a synonymous mutation and two missense mutations (Arg 203 Lys and Gly 204 Arg) (**Table 1**). Tracing these three SNPs showed that the recent reemergence of COVID-19 in the Xinfadi market in Beijing, China, was associated with European isolates (Wenjie et al., 2020).

SNP tracing reconstructed an early transmission history of North American isolates

In the North American SARS-CoV-2 population, 45% of strains have these two SNPs, and 69% have these SNPs for North American lineage B.1 (**Table 1**). Because of the high occurrence of the two SNPs, tracing the two SNPs may provide a reliable transmission route of SARS-CoV-2 in the major North American human population. We thus performed a retrospective tracing study in our high confidential datasets (2599 flittered strains) to identify the time order of isolates occurring at the two SNPs on all continents and in lineage B.1. We found that the first isolate started in Europe (Feb 26, 2020) 3 days earlier than the occurrence date of the North American isolates (Feb 29, 2020) and 17 days earlier than the Asian isolates (Taiwan China dominated) (Mar 13, 2020) (**Fig. 5A**). By further tracing the accumulating frequencies per day of the two SNPs during mid to late March, we found that North American lineage B.1 highly accumulated these two SNPs (**Fig. 5B**). In addition, the mean number of all accumulating SNPs during mid to late March was significantly lower than that before or after the same period (**Fig. S6**). This evidence indicated that the two SNPs were strongly selected in the North American SARS-CoV-2 isolates, in particular lineage B.1 from mid to late March. The accumulation of the two SNPs may explain the sharp increase in confirmed cases in North America before early April (WHO reported) (WHO, 2020).

SARS-CoV-2 is thought to be transmitted from wildlife to humans before human-to-human transmission occurs (Andersen, Rambaut, Lipkin, Holmes, & Garry, 2020; Shi et al., 2020; Tang et al., 2020). We carefully checked whether bat or pangolin CoVs evolved with the two mutations before they jumped to the human species. However, we did not find either of these SNPs in the bat or pangolin-related CoVs (P. Zhou et al., 2020) (**Fig. S7**). Alternatively, in bat or pangolin CoVs, the 1059 site has no or a C>A variant, and the 25563 site has a G>A variant instead (**Fig. S7**). (P. Zhou et al., 2020)
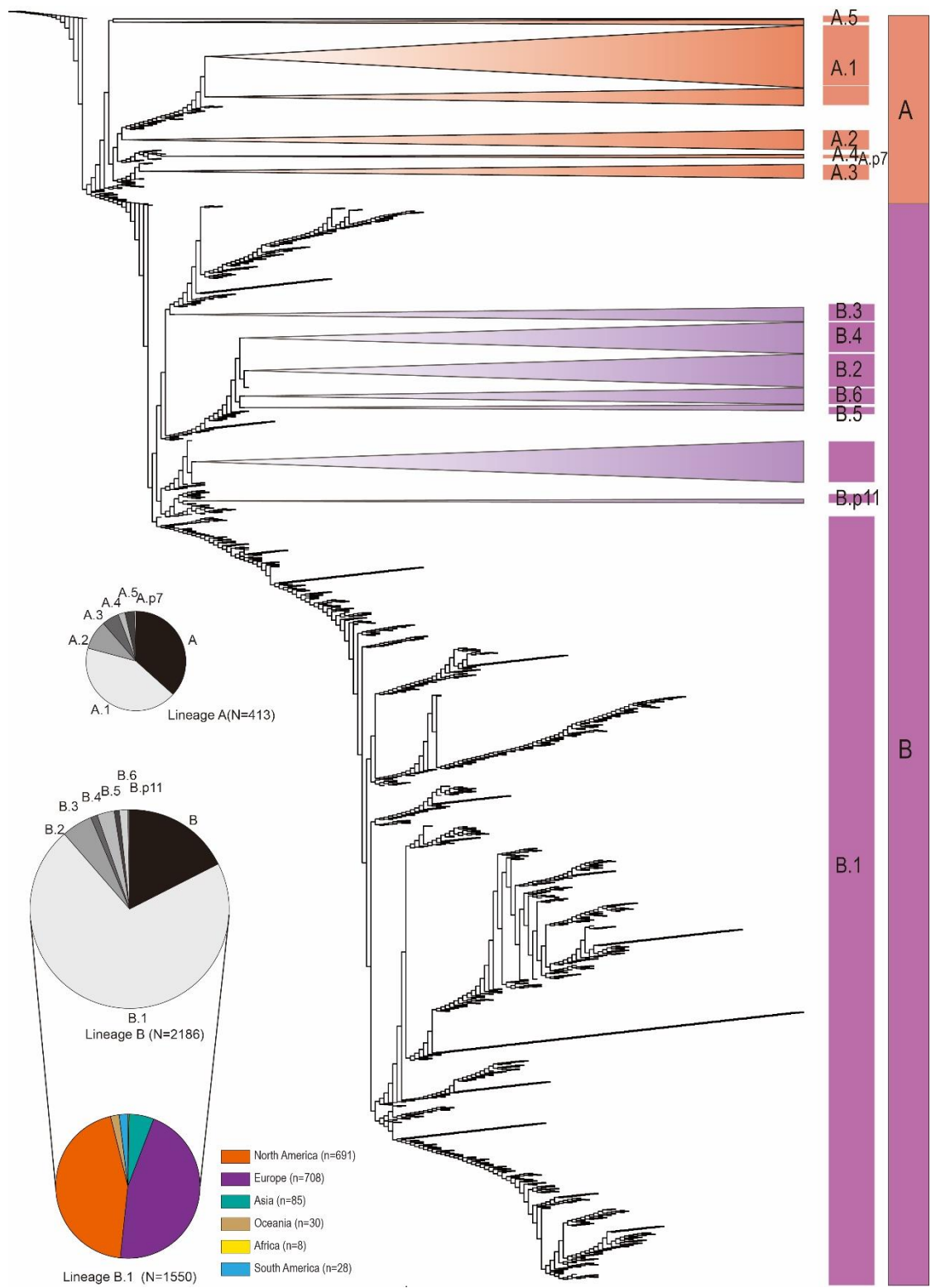
FIG 3 Phylogenetic tree of early global SARS-CoV-2. The 2599 full-genome sequences were used to construct the phylogenetic tree via the maximum likelihood (ML) method using IQ-TREE 2 software (version 2.1.2, model: GTR+Γ). Accordingly, the early SARS-CoV-2 isolates were rooted in two lineages, lineages A (n=413) and B (n=2186), in which North American isolates dominated lineage B (n=818) and sub-lineage B.1 (n=691). The constituents of the main lineages A and B as well as lineage B.1 are displayed by three pie charts. Specifically, North American and European isolates dominate lineage B.1.

FIG 4 GWAS and linkage disequilibrium (LD) analysis. (A) Manhattan plot comparing the North American SARS-CoV-2 isolates (n=1063) to the isolates from the remaining continents (n=1536). Genomic coordinates are displayed along the X-axis, and -log10 of the association p-value for each SNP is displayed on the Y-axis (threshold p-value = 1.00×10^{-15}). Different blocks indicate the different protein-encoding regions. (B) Linkage disequilibrium between SNPs in SARS-CoV-2. LD plot of any two SNP pairs among the 21 sites. The number near slashes shows the genomic coordinates. The colour in the square is given by the standard (D'/LOD), and the number in the square is the r2 value.

TABLE 1 List of top 21 hits of causative SNPs.

#	SNP	Gene	Effects	P-value	SNP Frequencies (%)				
					North America	Lineage B*	Lineage B.1*	Europe	Asia
1	25563.G>T	ORF3a	Missense(Gln57His)	2.98E-261	574/1063(54%)	574/818(70%)	573/691(83%)	119/951(13%)	19/448(4%)
2	1059.C>T	ORF1ab	Missense(Thr265Ile)	2.44E-212	479/1063(45%)	479/818(59%)	479/691(69%)	94/951(10%)	0/448(0%)
3	17858.A>G	ORF1ab	Missense(Met5865Val)	3.57E-117	194/1063(18%)	0/818(0%)	0/691(0%)	0/951(0%)	0/448(0%)
4	17747.C>T	ORF1ab	Synonymous	1.43E-116	193/1063(18%)	0/818(0%)	0/691(0%)	0/951(0%)	4/448(1%)
5	18060.C>T	ORF1ab	Missense(Ser5932Phe)	2.42E-113	196/1063(18%)	1/818(0%)	0/691(0%)	0/951(0%)	0/448(0%)
6	29553.G>A	ORF10	Upstream	7.26E-51	80/1063(8%)	80/818(10%)	80/691(12%)	0/951(0%)	30/448(7%)
7	28882.G>A	N	Synonymous	6.15E-39	53/1063(5%)	53/818(6%)	53/691(8%)	207/951(22%)	30/448(7%)
8	28883.G>C	N	Missense(Gly204Arg)	6.15E-39	53/1063(5%)	53/818(6%)	53/691(8%)	207/951(22%)	30/448(7%)
9	28881.G>A	N	Missense(Arg203Lys)	2.77E-38	54/1063(5%)	54/818(7%)	54/691(8%)	207/951(22%)	87/448(19%)
10	28144.T>C	ORF8	Missense(Leu84Ser)	9.73E-33	245/1063(23%)	0/818(0%)	0/691(0%)	49/951(5%)	0/448(0%)
11	27964.C>T	ORF8	Missense(Ser24Leu)	2.74E-29	47/1063(4%)	47/818(6%)	47/691(7%)	0/951(0%)	2/448(0%)
12	11916.C>T	ORF1ab	Missense(Ser3884Leu)	4.87E-29	51/1063(5%)	51/818(6%)	51/691(7%)	0/951(0%)	96/448(21%)
13	8782.C>T	ORF1ab	Synonymous	4.03E-28	245/1063(23%)	0/818(0%)	0/691(0%)	53/951(6%)	9/448(2%)
14	15324.C>T	ORF1ab	Missense(Thr5020Ile)	5.83E-27	2/1063(0.2%)	2/818(0%)	2/691(0%)	80/951(8%)	130/448(29%)
15	11083.G>T	ORF1ab	Missense(Leu3606Phe)	3.99E-25	71/1063(7%)	65/818(8%)	5/691(1%)	92/951(10%)	2/448(0%)
16	18998.C>T	ORF1ab	Missense(His6245Tyr)	1.10E-22	41/1063(4%)	41/818(5%)	41/691(6%)	0/951(0%)	2/448(0%)
17	29540.G>A	ORF10	Upstream	1.10E-22	41/1063(4%)	41/818(5%)	41/691(6%)	0/951(0%)	8/448(2%)
18	18877.C>T	ORF1ab	Synonymous(His6245Tyr)	2.31E-19	65/1063(6%)	63/818(8%)	62/691(9%)	10/951(1%)	0/448(0%)
19	29711.G>T	5'UTR	Downstream	4.30E-19	31/1063(3%)	31/818(4%)	1/691(0%)	0/951(0%)	1/448(0%)
20	1604.AATG>A	ORF1ab	Deletion(delTGA)	2.87E-17	4/1063(0.4%)	4/818(0%)	0/691(0%)	69/951(7%)	2/448(0%)
21	27046.C>T	M	Missense(Thr175Met)	3.48E-17	2/1063(0.2%)	2/818(0%)	2/691(0%)	60/951(6%)	60/951(13%)

*, North America lineage B or B.1.
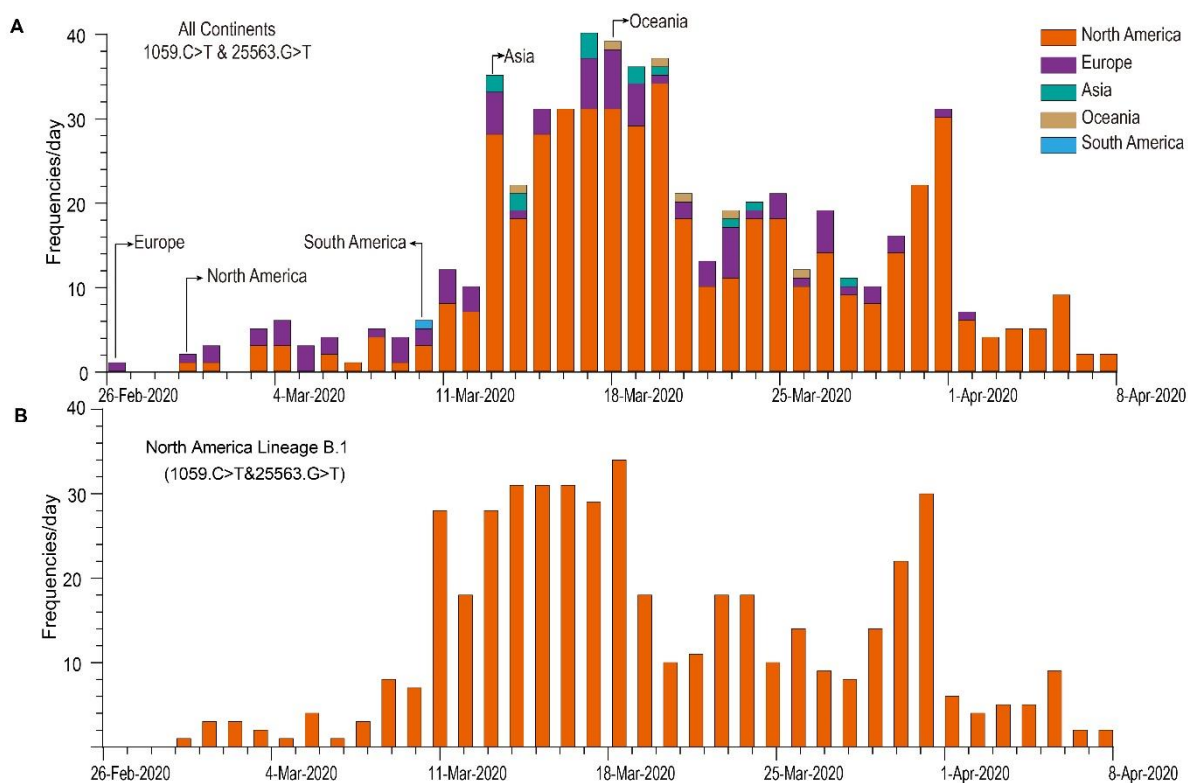
FIG 5 Retrospectively tracing the early SARS-CoV-2 isolates with SNPs (1059.C>T & 25563.G>T) **of all continents and North American lineage B.1.** (A) The time-dependent accumulating plot for frequencies of the two SNPs (1059.C>T & 25563.G>T) between continents. The continents are labelled by different colours, and the continents of the first occurrence of the two SNPs are indicated. (B) The time-dependent accumulation plot of North American lineage B.1. Of note, the two SNPs largely and concurrently accumulated during mid to late March and occurred most frequently in isolates of North American lineage B.1 (479/691, 69.31%) (**Table 1**).

Discussion

Herein, we identified the genetic signatures of bat-to-human CoVs and specified an early transmission history of North American SARS-CoV-2. Although human CoVs are highly similar to bat CoVs by sequence and genome organization (Perlman, 2020; P. Zhou et al., 2020), several specific genetic signatures were newly identified in this study, such as a unique ORF10 in human SARS-CoV-2, an identical ribosomal frameshift motif, and strong purifying selection after zoonotic transfer. In addition, we also found that the two causative SNPs that were present in approximately half of the North American SARS-CoV-2 isolates represented 69% of the isolates of North American lineage B.1 (Rambaut et al., 2020). The early transmission history of the major North American SARS-CoV-2 isolates was reconstructed by tracing the occurrence date of isolates with these two SNPs, and transmission started in Europe, North America, South America, and later Asia and Oceania.

The genetic signature and its extent determine the bat-to-human cross-species transmission of CoVs, which is still largely undocumented. However, distinctive genetic signatures can

possibly predict and estimate the risk of zoonotic transmission. The unique ORF10 of human SARS-CoV-2 and the insertion of the AAT codon in the slippage signal of MERS-CoV could serve as novel targets for differential diagnosis. Importantly, human SARS-CoV-2 as well as SARS-CoV and MERS-CoV undergo strong purifying selection. Strong purifying selection involving large synonymous mutations may promote fitness in the human system by regulating viral translation efficiency (Ou et al., 2018). Monitoring the mutation rate in particular synonymous mutations and its possible impact would help to predict the risk of zoonotic transfer of bat CoVs.

Following zoonotic transfer, understanding the trend of SARS-CoV-2 geographic transmission is important for public measures (CDC-USA, 2020). This study was published as a BioRvix preprint and gained the attention of certain medical communities, such as the Centers for Disease Control and Prevention of the USA and Washington State Department of Health (CDC-USA, 2020; health, 2020; Ou et al., 2020). Regardless of the early transmission history, the genetic signatures identified may help methodology development for precisely tracing viral transmission in real time, such as via the two causative SNPs. These two SNPs pose a great possibility for epidemiological surveillance in the North American population due to their high prevalence in the same population. In clinical diagnosis, these SNPs may improve methodology development to specifically detect North American isolates, such as via SNP-based allele-specific polymerase chain reaction (ASPCR) (Corman et al., 2020; Ugozzoli & Wallace, 1991). It has been reported that SARS-CoV-2 with the D614G mutation in the S protein increases infectivity in human lung cells (Yurkovetskiy et al., 2020). The two SNPs we identified are responsible for two missense mutations that probably change the protein structure and function to some extent. More mechanical investigations of the functional impact caused by these two SNPs would be enhanced in this aspect, as they are possibly druggable targets.

The hard lesson of the ongoing SARS-CoV-2 pandemic is its strain of the global public health system (Ji et al., 2020). Before the pandemic, researchers detected the proximal origin of SARS-CoV-2 in bats from 2015-2017 (Hu et al., 2018). However, its risks to public health were largely ignored. A platform for early surveillance and risk estimation of bat CoVs is very much needed. In the future, it is hoped that these exclusively genetic signatures may help public health surveillance and measures.

Materials and methods

Data acquisition

Human SARS-CoV-2 isolated from Wuhan, China, was obtained from the NCBI database (GenBank No.: MN908947.3). To identify the phylogenetically closed bat CoVs associated with human SARS-CoV-2, the BLAST searching tool of the NCBI viral database was used. Based on the constructed phylogenetic tree, we identified two bat SARS-like CoVs (referred to here as

bat SARS-CoV-2) (GenBank No: MG772933.1 and MG772934.1) that were evolutionarily close to human SARS-CoV-2 (**Table S1,**) (**Fig. S1**). For human SARS-CoVs and MERS-CoVs, similar approaches were used to identify their bat CoV pairs; the camel MERS-CoV-related literature was also reviewed (Madani, Azhar, & Hashem, 2014).

For the GWAS, full-genome sequences of global SARS-CoV-2 collected from 12 Dec 2019 to 24 Apr 2020 (8:00 GMT + 8) were archived from the database of the GISAID Initiative EpiCoV platform (GISAID; https://www.epicov.org). A total of 8480 sequences were archived and filtered by criteria, including high coverage only (> 29,000 bp, 1X coverage of genome), exclusion of low coverage, and sequences with unconfident bases (N) inside. The PANGOLIN isolate (EPI_ISL_410539) and bat CoVRaTG13 isolate were also used for phylogenetic analysis. The identical sequences were further removed by CD-HIT software (version 4.8.1, parameters: -aL 1 -aS 1 -c 1 -s 1) (Huang, Niu, Gao, Fu, & Li, 2010). A final 2599 sequences were used in this study (online data file S1).

Sequence alignment

A codon-based Cluster W method was used for the multiple sequence alignment and identification of the ribosomal frameshift motifs among the three CoV bat-human pairs. The genomic organization of annotated CoVs was visualized by SnapGene (Version 4.2.4).

Phylogenetic analysis

For the accessory and necessary protein-coding genes, phylogenetic trees were constructed by the maximum likelihood method. The evolutionary distances were calculated by the differences between amino acid substitutions or nucleotide substitutions per site. Confidence probability was estimated by the bootstrap test (100 replicates).

The 2599 full-genome sequences were aligned by MAFFT software (version 7.407, algorithm: FFT-NS-2) (Nakamura, Yamada, Tomii, & Katoh, 2018). The phylogenetic tree was constructed by IQ-TREE 2 (version 2.1.2, parameter: -nt -gtr -gamma) using the GTR+Γ model of nucleotide substitution (Minh et al., 2020). The phylogenetic tree was visualized by FigTree (version 1.4.4, http://tree.bio.ed.ac.uk/software/figtree/). Each descendant lineage was annotated according to criteria from recent publications (Rambaut et al., 2020).

Mutation analysis

Synonymous and nonsynonymous differences per sequence between human and bat CoVs were estimated using the Nei-Gojobori model by MEGA-X software (Table S2-11, online). The dN/dS ratio is an indicator of directional selection: a ratio above 1 implies positive selection (nature), a ratio less than 1 implies negative selection (purifying), and a ratio equal to 1

indicates no selection (neutral). The dN/dS ratio (Table S12-16, online) is calculated by the following equations: $d_N = -\frac{3}{4}\ln\left(1 - 4\frac{p_N}{3}\right)$; $d_S = -\frac{3}{4}ln(1 - 4\frac{p_S}{3})$; $d_N/d_S = \frac{d_N}{d_S}$.

Codon usage bias analysis

Relative synonymous codon usage (RSCU) of CoV necessary protein-coding genes (i.e., ORF1ab-S-E-M-N) was analysed by CODONW software (http://www.molbiol.ox.ac.uk/cu, version 1.4.2) using standard genetic codes. The linear regression of RSCU between bat-human CoV pairs was analysed by GraphPad Prism 8.0.

SNP calling

SNPs and INDEL polymorphisms were detected by MUMmer software (version 3.0, nucmer, show-snps) (Kurtz et al., 2004) using the Wuhan-Hu-1 strain (GISAID: EPI_ISL_402125, GenBank: NC_045512.2) as a reference genome. To validate the identity of the resulting polymorphisms, raw reads (40 out of 2599 strains, NCBI SRA database) were analysed by the bwa program (version 0.7.16a) (H. Li & Durbin, 2010) and the mpileup program of the SAMtools software (version 1.10) (H. Li, 2011). The validation was consistent with the polymorphisms detected by MUMmer software.

GWAS and linkage disequilibrium (LD) analysis

To identify causative SNPs in the population of North American SARS-CoV-2 (cases=1063, controls=1536), a geographically stratified genome-wide association study of 5312 mutations was performed using PLINK software (version 1.90) (Purcell et al., 2007). The empirical threshold of the p-value was suggested to be 9.41×10^{-6} ($0.05/5312=9.41 \times 10^{-6}$) calculated by the Benjamini & Hochberg method (1995) (Benjamini & Hochberg, 1995), but we further increased the threshold of the p-value to 1.00×10^{-15} to detect the most causative SNPs. The top 21 significant SNPs were listed (**Table 1**), and the LD of pairing SNPs was estimated and visualized by Haploview software (version 4.1) (Barrett, Fry, Maller, & Daly, 2005).

SNP accumulating analysis

To analyse the trend of SNP accumulation during March 2020, the frequencies of average SNP accumulation per day were counted. This same trend in 1059.C>T and 25563.G>T in North American SARS-CoV-2 and that of the other continents was traced by the date of occurrence of the two SNPs. The same analysis of these two SNPs in North American lineage B.1 was also conducted. These analyses were performed by Microsoft® Excel 2016 (data file S2, online).

Statistical analysis

The probability of rejecting the null hypothesis of strict neutrality (dN = dS) in favour of the alternative hypothesis (dN < dS) was calculated by the codon-based Z-test of purifying

selection. Data from the SNP accumulation analysis were plotted by GraphPad (Version 8.2.1). The mean differences of all types of SNP accumulation per day per strain were determined by the Mann-Whitney U test (interval = 10 days) (R version 3.6.2). P values less than 0.05 were considered significant.

References

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. Nature Medicine, 26(4), 450– 452.

Azhar, E. I., El-Kafrawy, S. A., Farraj, S. A., Hassan, A. M., Al-Saeed, M. S., Hashem, A. M., & Madani, T. A. (2014). Evidence for camel-to-human transmission of MERS coronavirus. New England Journal of Medicine, 370(26), 2499– 2505.

Baranov, P. V., Henderson, C. M., Anderson, C. B., Gesteland, R. F., Atkins, J. F., & Howard, M. T. (2005). Programmed ribosomal frameshifting in decoding the SARS-CoV genome. Virology, 332(2), 498– 510.

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics, 21(2), 263– 265.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B, 57(1), 289– 300.

Centers for Disease Control and Prevention-USA (2020). Hot topics of the day. Retrieved from https://phgkb.cdc.gov/PHGKB/phgHome.action?action=archive&date=05/14/2020

Coleman, J. R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E., & Mueller, S. (2008). Virus attenuation by genome-scale changes in codon pair bias. Science, 320(5884), 1784– 1787.

Consortium, & C. S. M. E (2004). Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. Science, 303(5664), 1666– 1669.

Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K. W., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M. L., Mulders, D. G. J. C., Haagmans, B. L., van der Veer, B., van den Brink, S., Wijsman, L., Goderski, G., Romette, J.-L., Ellis, J., Zambon, M., … Drosten, C. (2020). Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance Weekly, 25(3), 2000045.

Forni, D., Cagliani, R., Clerici, M., & Sironi, M. (2017). Molecular evolution of human coronavirus genomes. Trends in Microbiology, 25(1), 35– 48. https://doi.org/10.1016/j.tim.2016.09.001

Hanson, G., & Coller, J. (2017). Codon optimality, bias and usage in translation and mRNA decay. Nature Reviews Molecular Cell Biology, 19, 20.

Hu, D., Zhu, C., Ai, L., He, T., Wang, Y. I., Ye, F., Yang, L. U., Ding, C., Zhu, X., Lv, R., Zhu, J., Hassan, B., Feng, Y., Tan, W., & Wang, C. (2018). Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. Emerging Microbes & Infections, 7(1), 154.

Huang, Y., Niu, B., Gao, Y., Fu, L., & Li, W. (2010). CD-HIT Suite: A web server for clustering and comparing biological sequences. Bioinformatics, 26(5), 680– 682.

Ji, Y., Ma, Z., Peppelenbosch, M. P., & Pan, Q. (2020). Potential association between COVID-19 mortality and health-care resource availability. The Lancet. Global Health, 8(4), e480.

Kelly, J. A., Olson, A. N., Neupane, K., Munshi, S., San Emeterio, J., Pollack, L., Woodside, M. T., & Dinman, J. D. (2020). Structural and functional conservation of the programmed −1 ribosomal frameshift signal of SARS coronavirus 2 (SARS-CoV-2). Journal of Biological Chemistry, https://doi.org/10.1074/jbc.AC120.013449

Kryazhimskiy, S., & Plotkin, J. B. (2008). The Population Genetics of dN/dS. Plos Genetics, 4(12), e1000304.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. Genome Biology, 5(2), R12.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics, 27(21), 2987–2993.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics, 26(5), 589– 595.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J. H., & Wang, L.-F. (2005). Bats Are Natural Reservoirs of SARS-Like Coronaviruses. Science, 310(5748), 676– 679.

Madani, T. A., Azhar, E. I., & Hashem, A. M. (2014). Evidence for camel-to-human transmission of MERS coronavirus. New England Journal of Medicine, 371(14), 1360.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution, 37(5), 1530– 1534.

Nakamura, T., Yamada, K. D., Tomii, K., & Katoh, K. (2018). Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics, 34(14), 2490– 2492.

Ou, X., Cao, J., Cheng, A., Peppelenbosch, M. P., & Pan, Q. (2019). Errors in translational decoding: tRNA wobbling or misincorporation? Plos Genetics, 15(3), e1008017.

Ou, X., Wang, M., Mao, S., Cao, J., Cheng, A., Zhu, D., Chen, S., Jia, R., Liu, M., Yang, Q., Wu, Y., Zhao, X., Zhang, S., Liu, Y., Yu, Y., Zhang, L., Chen, X., Peppelenbosch, M. P., & Pan, Q. (2018). Incompatible Translation Drives a Convergent Evolution and Viral Attenuation During the Development of Live Attenuated Vaccine. Frontiers in Cellular and Infection Microbiology, 8, https://doi.org/10.3389/fcimb.2018.00249

Ou, X., Yang, Z., Zhu, D., Mao, S., Wang, M., Jia, R., & Wu, Y. J., (2020). Tracing two causative SNPs reveals SARS-CoV-2 transmission in North America population. BioRvix, https://doi.org/10.1101/2020.05.12.092056

Perlman, S. (2020). Another decade, another coronavirus. New England Journal of Medicine, 382(8), 760– 762.

Power, R. A., Parkhill, J., & de Oliveira, T. (2017). Microbial genome-wide association studies: Lessons from human GWAS. Nature Reviews Genetics, 18(1), 41– 50.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics, 81(3), 559– 575.

Rambaut, A., Holmes, E. C., O'Toole, Á., Hill, V., McCrone, J. T., Ruis, C., du Plessis, L., & Pybus, O. G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nature Microbiology, 5(11), 1403– 1407.

Shi, J, Wen, Z, Zhong, G, Yang, H, Wang, C, Huang, B… Bu, Z. (2020). Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS–coronavirus 2. Science, 368(6494), 1016– 1020.

Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., Duan, Y., Zhang, H., Wang, Y., Qian, Z., Cui, J., & Lu, J. (2020). On the origin and continuing evolution of SARS-CoV-2. National Science Review, 26(4), 450– 452.

Ugozzoli, L., & Wallace, R. B. (1991). Allele-specific polymerase chain reaction. Methods, 2(1), 42– 48.

Washington State Department of Health (2020). 2019-nCoV Literature Situation Report (pp. 1– 8). Health. Retrieved from https://www.doh.wa.gov/Portals/1/Documents/1600/coronavirus/LitRep-20200514.pdf

Wenjie, T., Peihua, N., Xiang, Z., Yang, P., Yong, Z., Lijuan, C., & Guizhen, W. (2020). Notes from the Field: Reemergent Cases of COVID-19 — Xinfadi Wholesales Market, Beijing Municipality, China, June 11, 2020. China CDC Weekly, https://doi.org/10.46234/ccdcw2020.132

WHO (2020). WHO report Situation Report – 160. Retrieved from https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200513-covid-19-sitrep-114.pdf?sfvrsn=17ebbbe_4

Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. The Lancet, 395(10225), 689– 697.

Yurkovetskiy, L., Wang, X., Pascal, K. E., Tomkins-Tinch, C., Nyalile, T. P., Wang, Y., Baum, A., Diehl, W. E., Dauphin, A., Carbone, C., Veinotte, K., Egri, S. B., Schaffner, S. F., Lemieux, J. E., Munro, J. B., Rafique, A., Barve, A., Sabeti, P. C., Kyratsous, C. A., … Luban, J. (2020). Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. Cell, 183(3), 739– 751 e738.

Zhou, J. H., Li, X. R., Lan, X., Han, S. Y., Wang, Y. N., Hu, Y. H., & Pan, Q. W. (2019). The genetic divergences of codon usage shed new lights on transmission of hepatitis E virus from swine to human. Infection Genetics and Evolution, 68, 23– 29.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X. I., … Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature, 579(7798), 270– 273.

Supplementary information
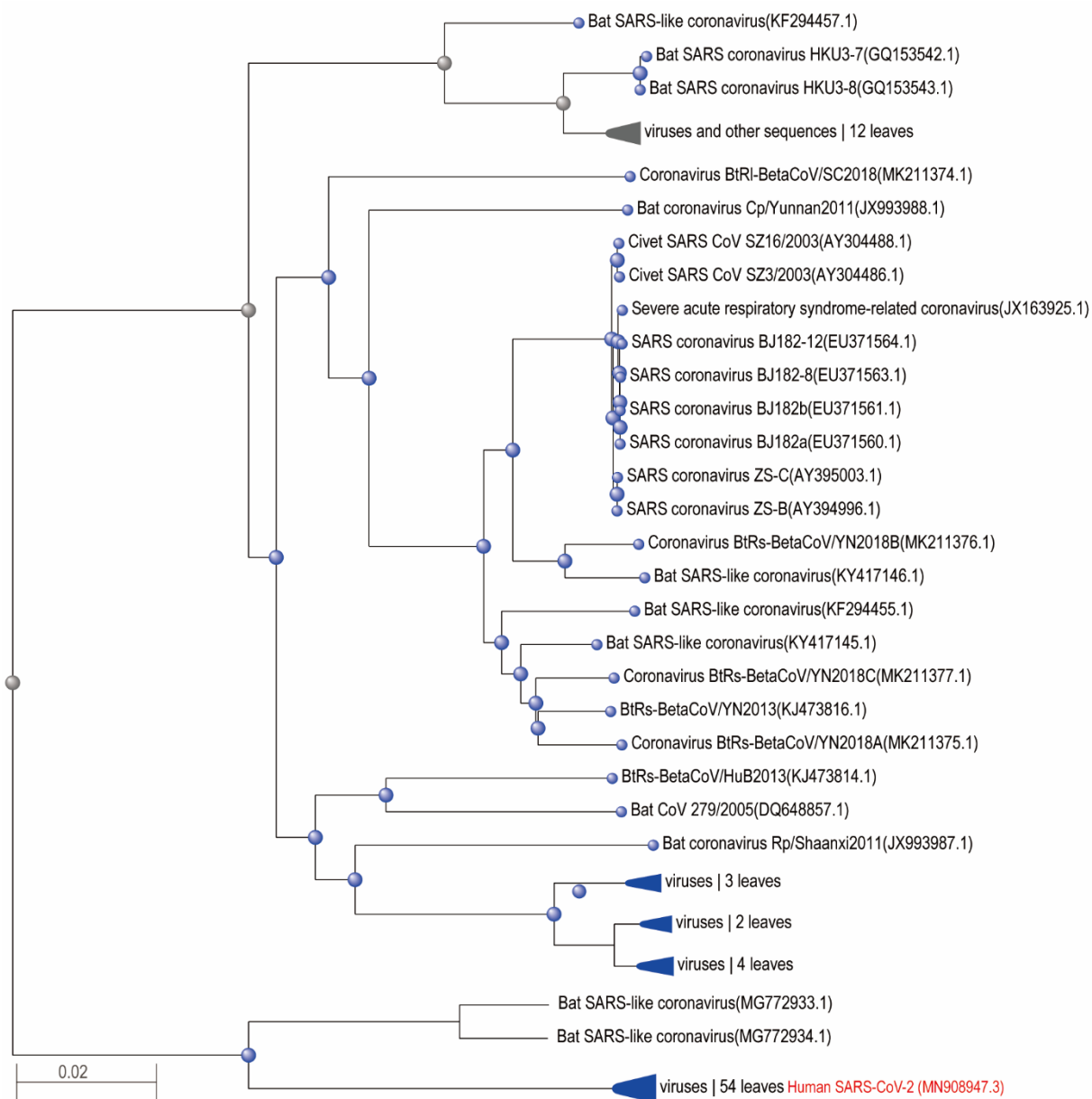
Fig. S1 The phylogenetic tree of complete genome sequences of SARS-COV-2
The human SARS-COV-2 and its related CoVs were selected based on the searching results using the BLAST tool. For human SARS-COV-2, only two bat SARS-like CoVs (Genbank: MG772933.1 and MG772934.1) are phylogenetically closed to the SARS-COV-2. Thus, the two bat SARS-like CoVs are tentatively referred by bat SARS-CoV-2 in the context for clearness.
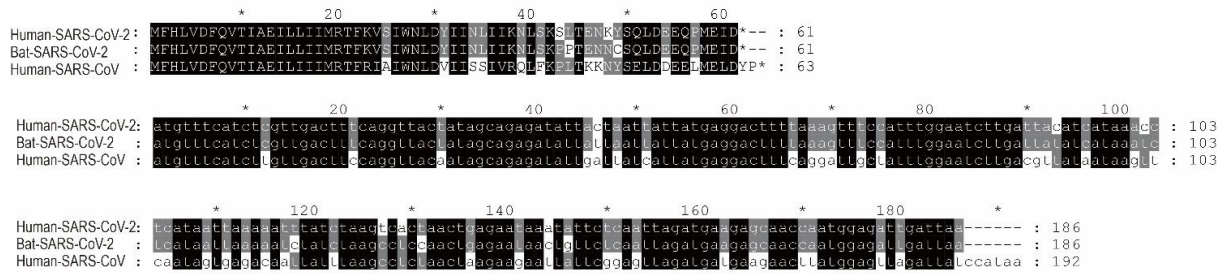
Fig. S2 Alignment of ORF6 sequence at protein and nucleotide level (upper and lower panel)
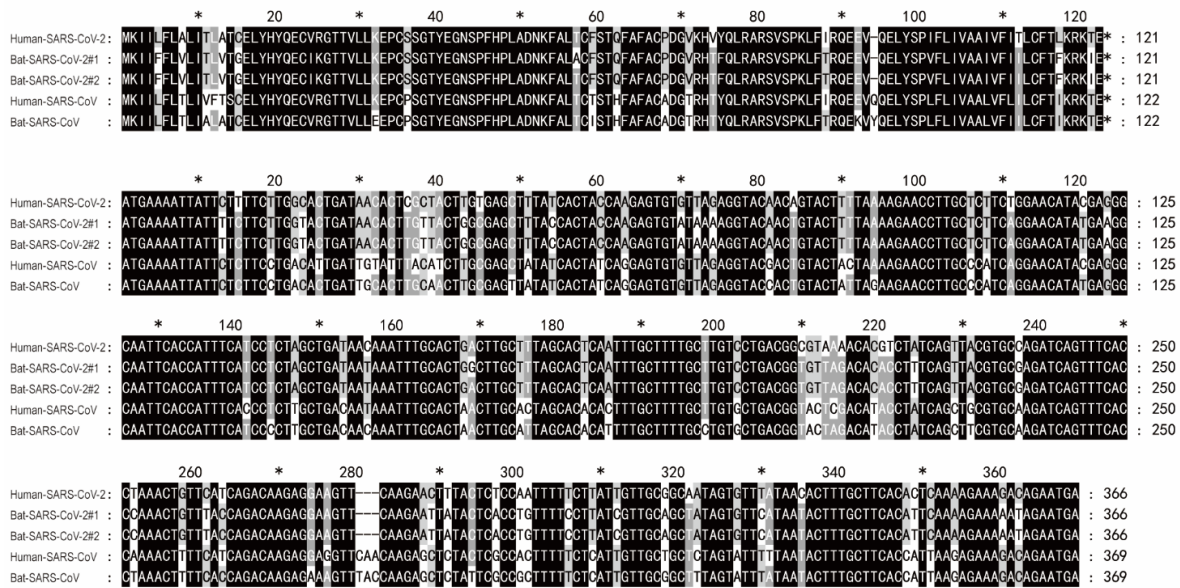
Fig. S3 Alignment of ORF7a sequence at protein and nucleotide level (upper and lower panel)
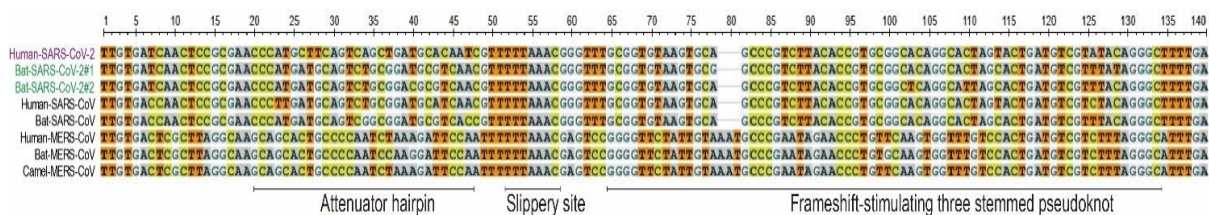
Fig. S4 The programmed -1 ribosomal frameshift signals among three CoVs

The codon-based alignment of ribosomal frameshifting signals among the three pairing CoVs are shown. The slippery sites are highly conserved and are identical U_UUA_AAC motif. Moreover, there is a conserved deletion of AAT codon (encoding asparagine) followed by the slippage site in both SARS-CoV-2 and SARS-CoV instead of MERS-CoV.
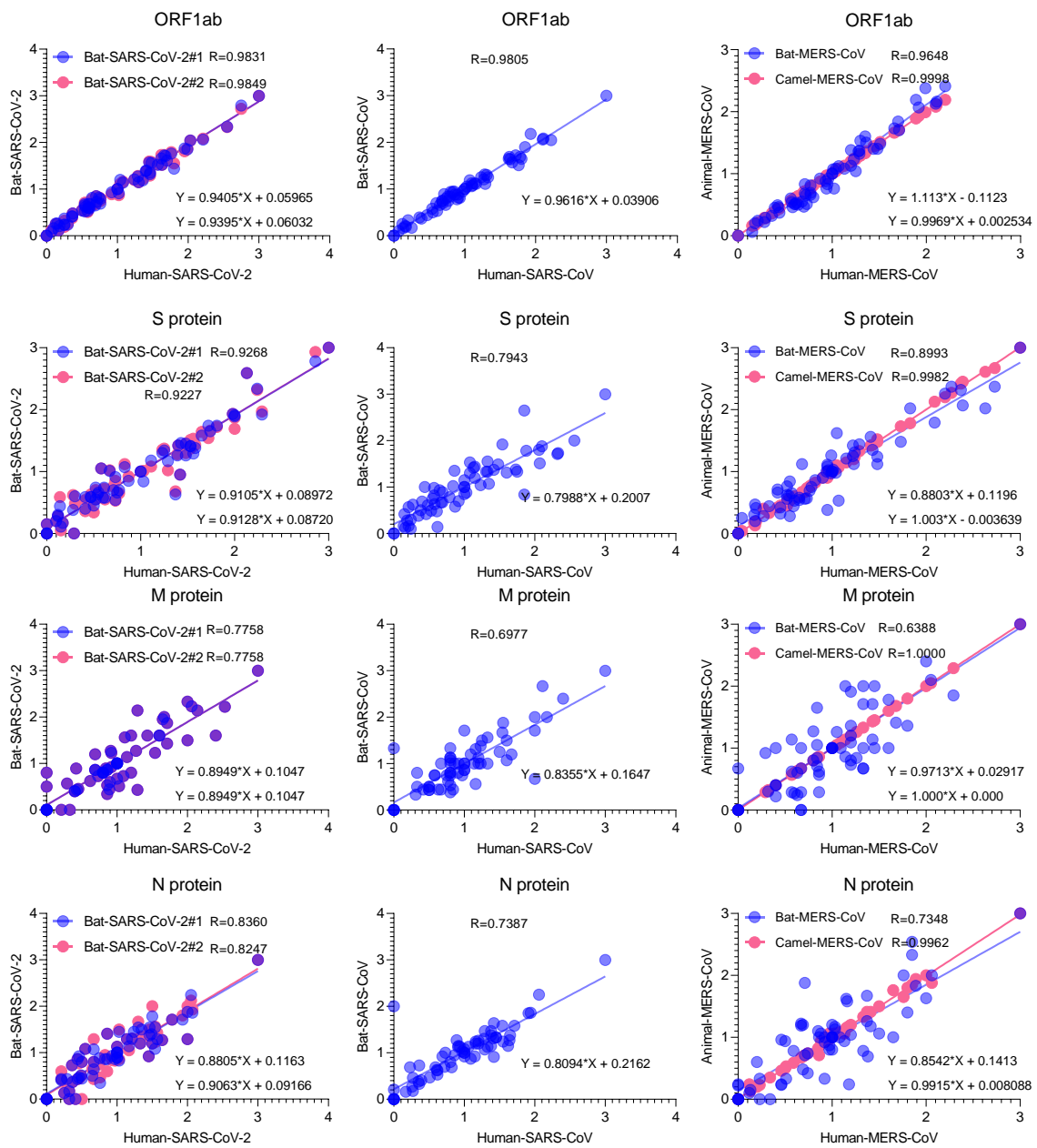
Fig. S5 RSCU based linear regression between bat CoVs and human CoVs
The RSCU data of bat CoVs and human CoVs are indicated by Y-axis and X-axis, respectively. For MERS-CoVs, the same linear regression of camel CoV and human CoV was also performed. The analysis of necessary protein-encoding genes is listed from top to bottom (i.e., ORF1ab-S -M-N). And the analysis of SARS-CoV-2, SARS-CoV, and MERS-CoV are displayed from left to right. The R coefficients and linear equations are shown correspondingly. Among all linear regressions, the slopes of linear equations are slightly lower than 1. This indicates that the RSCU of bat CoVs is slightly shifting to human CoVs to some extent. However, the slope of the linear equation for bat MERS-CoV and human MERS-CoV is a little higher than 1 (i.e., 1.113). RSCU, Relative Synonymous Codon Usage.
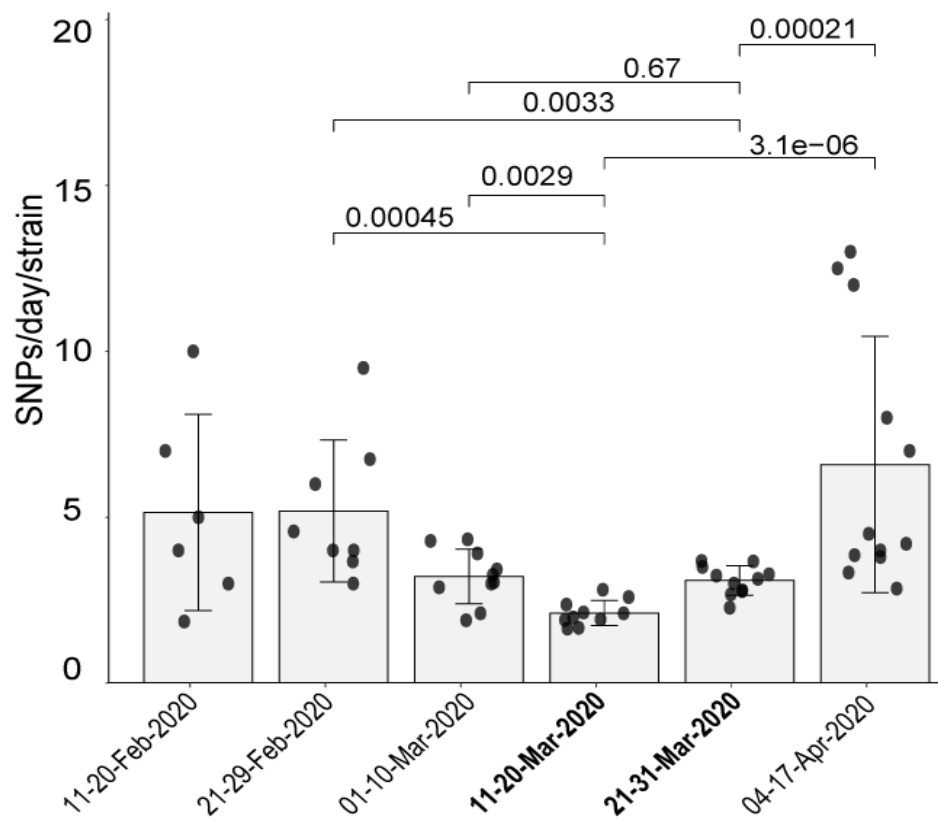
Fig. S6 All types of SNPs accumulation during mid to late March 2020
The mean differences of all types of SNPs accumulation per day per strain were analyzed by the Mann-Whitney U test (Interval = 10 days). A p-value of less than 0.05 is considered as statistically significant. The mean numbers before or after mid to late March are much higher than those during the same period.
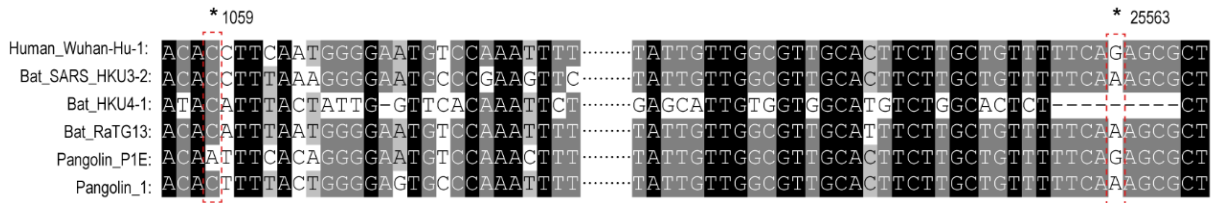
Fig. S7 SNPs calling in bat and pangolin related SARS-CoV-2 at 1059 and 25563 sites
Sequence alignment between bat and pangolin related SARS-CoV-2 was displayed. The Human Wuhan-hu-1 strain was used as a reference genome. Strains from bat and pangolin were named with its host name. Differently, at 1059 site, there is only one SNP for pangolin P1E strain (C>A) but not C>T. And the 25563 site has another type of SNP (G>A) but not G>T. Wuhan-Hu-1 reference strain is indicated.

Supplemental Tables

Table S1 Key source data of human and bat pairing coronaviruses

Virus	Host	Country (sampling time)	GenBank No.
SARS-CoV-2	Human	China (Dec-2019)	MN908947.3
	Bat	China (2015-2017)	MG772933.1
			MG772934.1
SARS-CoV	Human	China (Feb-2003)	AY278488.2
	Bat	China (Mar-2004)	DQ412042.1
MERS-CoV	Human	Saudi Arabia (Jun-2012)	KF958702
	Bat	South Africa (2011)	KC869678.4
	Camel	UAE (Mar-2015)	MF598621.1

Chapter 9

Summary and Discussion

Translational regulation bridges the cellular transcriptome and proteome, and this has become an overarching theme in modern biology and medicine [1]. From this thesis, it has emerged that among all layers of translational regulations, regulation at the level of translational decoding may be the most vital. Because the process of translational decoding builds a direct association between cellular mRNAs and proteins; this shapes almost every aspect of life science. This connection relies on the physical recognition of tRNA anticodon and mRNA codon, which constitutes the center of translational decoding [2,3]. As all domains of life science are subject to this basic principle, I feel it might be in practice the most important level of gene expression control, at least in eukaryotes. Thus driven, I investigated the key factors of translational decoding, tRNA molecules, and mRNA codon, and in **Chapter 1** of this thesis, I performed a first attempted to decipher how these factors contribute to gene expression in the setting of certain viral infections and cancer development [4]. From **Chapter 2** to **Chapter 5**, I have tried to establish a comprehensive understanding of tRNA biology starting from the genomic identification of tRNA genes (see **Chapter 2**) to determining the levels of mature tRNAs (see **Chapter 3**), and I investigated remodeling of the tRNA landscape in cells infected with HEV, as well as tRNA dysregulation in the setting of liver cancer (see **Chapter 5**). The cell has impressive mechanisms to prevent mistranslation from happening, nevertheless, aberrant decoding of codons does occur. In **Chapter 4**, I discussed the potential mechanism underlying erroneous translation and established that the main causes are tRNA misdecoding and misacylation. I also highlighted the consequences of mistranslation for viral evolution and cancer development. Understanding the crucial and unique tRNA-codon interaction inside the hepatitis viruses infected liver cells may yield novel avenues for developing antiviral medications necessary to improve treatment of the associated diseases [5]. Similarly, by targeting the key tRNA species that mediate the relevant oncogene expression during cancer development, novel medication for cancer therapy may be developed, and proof for this notion was provided in this thesis for HCC, the most important primary hepatocyte cancer.

The instrumental role that animal models fulfill in our efforts for understanding the pathogenesis is exemplified by many mammal models, such as mice, rats, chimpanzees, rabbits, pigs, and horses. Nevertheless, all these models suffer from substantial limitations and additional models are necessary to advance the field. Thus, I propose in this thesis an alternative avian hepatotropic virus (DHAV) infection model. This virus can cause obvious liver disease characterized by swelling livers mottled with hemorrhages in ducklings. In **Chapter 6**, the course of hepatitis and host immune responses in mature ducks induced by DHAV was investigated by assessing liver cell damage, fibrosis, and hepatocyte regeneration. I could link symptoms of the disease provoked and also its severity, at least theoretically, to the effectiveness of translational decoding of viral genes and pathogenesis-contributing genes, as well as immune regulator genes, collectively highlighting the importance of translational decoding for understanding and modifying pathophysiology.

By investigating another side of the translational decoding center, in **Chapters 7 and 8**, I have investigated the role of codon variations in artificial and natural viral adaption from one host

to another. The former artificial process has been widely applied for the development of attenuated vaccines, such as the DHAV live attenuated vaccine developed by passaging the duck isolated strains into chicken embryos (see **Chapter 7**). While the later naturally viral adaption involves the viral cross-species infection from one host to another, such as the bat-to-human coronaviruses (see **Chapter 8**). Together my work has advanced and enhanced the awareness of the importance of translational decoding in medical biology.

Role of tRNA biology in viral infection and cancer development

In **Chapter 1**, I reviewed the current knowledge of tRNA decoding with viral gene expression and oncogenic gene expression, as well as corresponding implications for health and disease. Translation of the involved genes highly relies on the host tRNAs which mediate mRNA decoding and this observation, in turn, suggests alternative approaches to treat viral infection and cancer, especially targeting specific tRNA species that are uniquely used. In this chapter I discuss two important questions 1) how host tRNA is specifically regulated by viral infection? [6] 2) how dysregulation of tRNA biology is implicated in different cancer types and what are the potential underlying mechanisms? [7,8]

Host tRNAs are conserved genes that are indispensable molecules to decode all 64 genetic codons present in mRNA. For a virus that can infect multiple host species (for instance a zoonotic virus), the viral protein-encoding gene must be decoded by the host tRNA pool and failure of this step will result in the absent production of viral proteins and consequently failure of infection. If a virus breaks the host species barrier, it is theoretically true that the virus may still be properly decoded by different hosts, its tRNA repertoire may be similar. To understand the repertoire of tRNA genes in the genome of humans and two important avian species (in casu chickens and ducks), I performed a comparative analysis of the tRNA sequences among these three species. In **Chapter 2**, I established that certain tRNA species (based on the presence of specific anticodons) are concurrently missing in human and avian species, including 8 tRNA-A_{34}NN and 6 tRNA-G_{34}NN. Also, the specifically intragenic promoter of tRNA genes (i.e., Box A and Box B) is conserved between humans and the two birds, in which box A and B are identical when the chicken and duck genome are compared. These similarities of tRNA characteristics may support viral cross-species infection because similar tRNA decoding strategies in two distinct hosts are used to translate the same viral genes.

At the transcriptional level of tRNA, currently, there is no easy-to-do approach for tRNA detection and quantification. Such analyses are largely impeded by the extremely short sequence of tRNA transcripts and the presence of redundant tRNA genes. Especially the levels of the mature (functional) type of tRNAs are difficult to determine. In **Chapter 3**, I developed a simple tRNA detection approach by the quantitative PCR method. Using this novel methodology, I observed that HEV infection dramatically reprograms the hepatic tRNAome that is likely to facilitate translation of the viral capsid-encoding gene instead of host antiviral genes, such as interferon stimulating genes.

Despite the existence of an impressive array of mechanisms to avoid erroneous translation of mRNA molecules, such mRNA translation is not always perfect and errors in the amino acid composition may occur. In **Chapter 4**, I discussed the key causes of mistranslation in the tRNA decoding process of the codons in mRNA. Such errors mainly derive from tRNA misdecoding (leading to misincorporation or stop codon readthrough) and misacylation (leading to wrong tRNA-amino acid coupling), especially when certain codon-paired tRNA species are missing (see **Chapter 2**). I also extended the discussion by adding potential implications on viral evolution and cancer development from the standpoint of protein synthesis instead of DNA and/ or RNA replication. Testing my ideas directly, however, was difficult in the absence of technological approaches that allow easy determination of the tRNA landscape in experimental samples.

Encouraged by the development of the novel tRNA methodology in **Chapter 3**, in **Chapter 5**, I mapped the tRNAome landscape in liver cancer patients and explore potential therapeutic targets at the interface of charging amino acid with tRNA. Interestingly, high expression of tRNA-Lys-CUU in HCC tumors was associated with more tumor recurrence (HR 1.1; P = .022) and worse patient survival (HR 1.1; P = .0037). The expression of Lysyl-tRNA Synthetase (KARS), was significantly upregulated in HCC tumor tissues compared to tumor-free liver tissues. In HCC cell lines, lysine deprivation, KARS knockdown, or treatment with the KARS inhibitor cladosporin effectively inhibited overall cell growth, single cell-based colony formation, and cell migration. This was mechanistically mediated by cell cycling arrest and induction of apoptosis. Finally, these inhibitory effects were confirmed in 3D cultured patient-derived CC organoids. Overall, the tRNA-Lys-CUU charging process can be therapeutically targeted and represents unexplored territory for developing novel treatment strategies against liver cancer. However, this experimental study was performed on an in vitro cell model though the potential therapeutical effect on the tumor organoids is promising. In the future, it is interesting to test the effect on a mouse model.

The implications of translational decoding in artificially and naturally viral cross-species infection

In humans, the pathogenesis of acute and chronic hepatitis is not well understood because of the limitations of models in experimental animals. In **Chapter 6,** I investigated the progress of viral hepatitis in mature ducks that is caused by DHAV over a prolonged period, while simultaneously I measured the expression of viral proteins in the liver in situ. Following experimental infection, the duck model displayed hepatocellular lesions (vacuolation, acidophilic degeneration, and steatosis), lymphocyte recruitment (neutrophil granulocytes, heterophilic granulocytes, and T cells or plasm cells), and repair (activation of hepatic stellate cells, fibrosis, and regeneration). These pathological features are accompanied by increases in liver injury biomarkers, such as AST and ALT. Considering the similarities in pathological and virological phenotypes with human HAV infection, I envisioned that ducks might be an alternatively small animal model that would provide insight into the pathogenesis of viral

hepatitis, fibrosis, and liver regeneration. The success of this animal model is linked to high viral gene expression and dysregulation of liver function-related gene expression, as well as abnormal expression of immune regulators during disease progression.

In its natural duck host, DHAV can replicate well and cause diverse symptoms, in particular in liver tissue. I observed that susceptible ducklings can be prevented by vaccination with a live attenuated vaccine, which was developed by serial passaging of the Duck Hepatitis A Virus in a chicken embryo. The artificial adaption of a virus in an unsuitable host has been widely used for vaccine development, including DHAV living but attenuated vaccines. To understand the role of codon variations in this type of artificially cross-species infection, in **Chapter 7**, I characterized the codon variations of DHAV duck isolates after passaging them in the chicken embryo. Phylogenetic analysis indicated that duck isolates from different geographic regions of China undergo a convergent evolution in the chicken embryos. This indicated that the same host translation system may contribute to the convergent evolution of duck-originated viruses in chickens. In support of this notion, my comparative analysis indicated that the codon usage bias of the attenuated strains was shaped by chicken codons usage bias, which appeared to be the driving force in viral adaption as a consequence of the incompatible translation present in the unsuitable host. I confirmed that the expression levels of four nonstructural proteins in the liver and kidney of ducks infected with an attenuated strain are significantly lower than those infected with a virulent strain, despite a similar virus load. Thus, the key mechanisms of viral attenuation revealed by this study may lead to innovative and easy approaches for designing live attenuated vaccines.

Also inspired by the global COVID pandemic, in **Chapter 8**, my research focus moved to the codon variations of certain famous coronaviruses, specifically, SARS-CoV-2, SARS-CoV, and MERS-CoV. In this chapter, I first identified the genetic signatures including the codon variations of three pairs of bat-to-human coronavirus, and the strong selection bias for codon use which appeared after jumping from bats to humans. By focusing on the early SARS-CoV-2 transmission in the North American human population, employing data obtained in a geographically stratified genome-wide association study (North American isolates and the remaining isolates), I was capable of performing a retrospective study. I established that the single nucleotide polymorphisms (SNPs) 1059.C>T and 25563.G>T were present in approximately half of the North American SARS-CoV-2 isolates that were obtained in March 2020. Retrospectively tracing isolates with these two SNPs was used to reconstruct the early, reliable transmission history of North American SARS-CoV-2. My analysis showed European isolates (February 26, 2020) transmission occurs 3 days earlier than North American isolates and 17 days earlier than Asian isolates. Collectively, I identified the genetic signatures of the three pairs of coronaviruses and reconstructed an early transmission history of North American SARS-CoV-2. These genetic signatures are possibly diagnosable and predictable markers for public health surveillance.

Thus overall, my thesis largely enhances our understanding of translation decoding, especially the tRNA-mRNA codon interactions, in virology and cancer biology and holds great promise in developing novel antiviral strategies, anticancer therapy, and new attenuated vaccine development via harnessing the tRNA decoding system.

References

1. Tahmasebi S, Khoutorsky A, Mathews MB, Sonenberg N. Translation deregulation in human disease. *Nature reviews Molecular cell biology* 2018; 19(12):791-807..
2. Torres AG, Batlle E, Ribas de Pouplana L. Role of tRNA modifications in human diseases. *Trends in molecular medicine* 2014; **20**(6): 306-14.
3. Koonin EV, Novozhilov AS. Origin and Evolution of the Universal Genetic Code. *Annu Rev Genet* 2017; **51**: 45-62.
4. Crick FH. Codon--anticodon pairing: the wobble hypothesis. *Journal of molecular biology* 1966; **19**(2): 548-55.
5. Sulima SO, Hofman IJF, De Keersmaecker K, Dinman JD. How Ribosomes Translate Cancer. *Cancer Discov* 2017; **7**(10): 1069-87.
6. Nunes A, Ribeiro DR, Marques M, Santos MAS, Ribeiro D, Soares AR. Emerging Roles of tRNAs in RNA Virus Infections. *Trends in biochemical sciences* 2020; **45**(9): 794-805.
7. Goodarzi H, Nguyen HCB, Zhang S, Dill BD, Molina H, Tavazoie SF. Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. *Cell* 2016; **165**(6): 1416-27.
8. Rapino F, Close P. Wobble uridine tRNA modification: a new vulnerability of refractory melanoma. *Mol Cell Oncol* 2018; **5**(6): e1513725.

Chapter 10

Nederlandse Samenvatting

Dutch Summary

Het centrale dogma van de moleculaire biologie stelt dat genetisch informatie ligt in het DNA, dat die informatie wordt afgeschreven als messenger-RNA (mRNA) en dat dit mRNA weer wordt vertaald in eiwitten, die dan de uiteindelijke effectoren zijn van de genetisch informatie. Dit dogma ligt aan de basis van ons denken over gezondheid en ziekte. Het lijkt evident dat de vertaling van mRNA naar eiwit hierbij van eminent belang is. Toch is onze kennis over hoe deze vertaling tot stand komt relatief beperkt. Met dit proefschrift probeer ik hier verandering in aan te brengen en ook probeer ik de inzichten die ik daarbij verkrijg te gebruiken om tot betere behandeling te komen van virale hepatitis (leverontsteking) en van leverkanker.

Het vertalen van mRNA in een eiwitstreng van aminozuren (translatie is een term die veel gebruikt wordt) gebeurt in kleine fabriekjes die we ribosomen noemen. Hier worden drie basen (ook wel de letters van de genetische code genoemd) van het mRNA vertaald in één aminozuur, waarna het ribosoom drie basen opschuift op het mRNA molecuul en het volgende groepje van drie basen gaat vertalen in één aminozuur. Zo'n groepje van drie basen/letters in het mRNA noemen we een codon. Het uitlezen van welk aminozuur bij welk codon hoort gebeurt door de transfer-RNA (tRNA) moleculen. Hoewel er 64 verschillende mRNA-codons zijn, zijn er geen 64 verschillende tRNA-moleculen. De verschillende tRNA moleculen die er wel zijn worden niet in gelijke mate aangemaakt door de cel. mRNA moleculen die codons bevatten waarvoor maar weinig tRNAs zijn, worden slecht afgeschreven en er zal dan maar weinig eiwit gevormd worden. De aanwezige kennis hierover vat ik samen in **Hoofdstuk 1**, waar ik ook drie vragen formuleer die ik in dit proefschrift wil beantwoorden: 1) hoe kunnen we het tRNA landschap (de relatieve hoeveelheden van de verschillende soorten tRNA moleculen) bepalen? 2) Wat betekenen de onregelmatigheden in het tRNA landschap voor ons denken over virale infectie en kanker? 3) Kunnen we deze onregelmatigheden gebruiken om betere therapie te ontwikkelen tegen virale hepatitis en leverkanker?

Net als mRNA moleculen, ligt de genetisch informatie die nodig is om tRNA moleculen te maken ook verscholen in het DNA. Een eerste inzicht over hoe tRNA landschappen er uit kunnen zien is dus te verkrijgen door het DNA in ons genoom te analyseren op de aanwezigheid van tRNA moleculen. Dit doe ik in **Hoofdstuk 2**, waar ik tRNA genen in mensen, kippen en eenden bekijk. Op genomisch niveau trof ik weinig verschillen aan wat betreft tRNA genen tussen deze soorten. Het tRNA landschap, echter, is wel heel anders in deze soorten. Dit resultaat geeft aan dat de verschillen in tRNA landschap niet hard verankerd zijn in het DNA maar het gevolg zijn van dynamische regulatie. Ik heb dit resultaat wereldkundig gemaakt middels publicatie in het vaktijdschrift *Infection Genetics and Evolution*.

Nu ik vastgesteld had dat het tRNA landschap dynamisch geregeld wordt, wilde ik natuurlijk weten hoe dit landschap eruit zou zien onder verschillende omstandigheden. Er was echter geen technologie aanwezig om dit makkelijk te bepalen. In **Hoofdstuk 3** beschrijf ik mijn inspanningen en resultaten die hebben geleid tot een bepalingsmethode die dat wel mogelijk maakt. Ik gebruik de methode om te laten zien dat infectie van levercellen met het Hepatis E-virus ook dit landschap verandert. Het virus zorgt ervoor dat er tRNA moleculen verschijnen

in de levercel die helpen de virale mRNA moleculen af te schrijven. Een de voor de hand liggende implicatie van deze bevinding is dat als we dit ombouwen van het tRNA landschap door het virus kunnen remmen we ook de ernst van de Hepatitis E-virus infectie kunnen remmen. Deze resultaten zijn verspreid in de wetenschappelijke wereld middels een publicatie in het vaktijdschrif *FEBS letters*.

Een interessant aspect van de tRNA biologie is dat het systeem ook wel eens fouten maakt: het verkeerde codon wordt gekoppeld aan het verkeerde aminozuur. Waarom dit gebeurt en hoe cel en virus met de fenomeen omgaan heb ik onderzocht in **Hoofdstuk 4** van dit proefschrift, waarin ik aannemelijk maak dat juist dergelijke fouten van belang zijn bij het ontwikkelen van virale leverontsteking en bij het kankerproces. Ik kon deze resultaten publiceren in het vooraanstaande vaktijdschrift *PLOS Genetics*.

In **Hoofstuk 5** probeer ik vervolgens een praktische toepassing te vinden voor de kennis die ik in dit promotieonderzoek heb vergaard. Om hun werking in het vertalen van mRNA moleculen naar eiwitstrengen uit te kunnen voeren moeten tRNA moleculen binden aan de jusite aminozuren. In dit hoofdstuk ga ik kijken wat er gebeurt als één bepaald aminozuur niet langer aan haar tRNA molecuul kan binden. Ik laat zien dat leverkankercellen hier slecht tegen kunnen terwijl ook bewijs presenteer dat normale niet-kankercellen hier waarschijnlijk wel tegen kunnen. Deze bevinding zou de basis kunnen zijn van een geheel nieuwe vorm van therapie voor leverkanker. Ik kon deze bevindingen wereldkundig maken middels een publicatie in het vooraanstaande vaktijdschrift *Liver International*.

Zoals gezegd zijn niet alle tRNA moleculen in gelijke mate in het tRNA landschap aanwezig. Het gevolg is dat codons in het mRNA die uitgelezen worden door zeldzame tRNA niet efficiënt vertaald worden en er ontstaatin zo'n geval dan weinig eiwit uit het betreffende mRNA. Een virus dat mRNA met veel zeldzame codons bevat zal dus verzwakt zijn en zich niet snel repliceren. Een dergelijk verzwakt virus is een makkelijke prooi voor het immuunsysteem, dat snel zal leren om dat virus te bestrijden. Heeft ons immuunsysteem eenmaal dit virus in de smiezen, dan zal het ook prima een variant van het virus met niet zeldzame codons aankunnen. In **Hoofdstuk 7** laat ik zien dat dit in praktijk ook echt werkt en deze aanpak een nieuwe waardevolle strategie kan zijn bij het ontwerpen van vaccins. Deze resultaten heb ik wereldkundig gemaakt middels een publicatie in *Frontiers in Cellular and Infection Microbiology*. In **Hoofdstuk 8** werk ik de onderliggende processen nog verder uit en laat zien dat ook de inmiddels beruchte Coronavirussen hun codons hebben aangepast aan het menselijk tRNA landschap, wat het belang van dit landschap in het begrijpen van menselijke gezondheid en ziekte benadrukt. Ook deze resultaten heb ik weten te publiceren, in dit geval in het vaktijdschrift Transboundary and Emerging Diseases .

Alles overziend heb ik het gevoel dat ik met proefschrift geslaagd ben in mijn ambitie het begrip te vergroten wat de vertaalslag van mRNA in eiwit kan betekenen voor ons denken over gezondheid en ziekte en ik werk deze gedacht verder uit in de discussie van mijn proefschrift

in **Hoofdstuk 9**. Hiermee hoop ik een verdere bijdrage te hebben geleverd aan de bewustwording van het concept dat translatie een belangrijk biologisch fenomeen is.

Appendix

Acknowledgements

Publications

PhD Portfolio

Curriculum Vitae

Acknowledgments

To Dr. Buyun Ma, Dr. Wenshi Wang, Dr. Wen Dang, Shihao Ding, Dr. Shan Li, Dr. Changbo Qu, Dr. Sunrui Chen, Dr. Jiaye Liu, Dr. Wanlu Cao, Dr. Guoying Zhou, Dr. Meng Li, Dr. Qin Yang, and Dr. Pengyu Liu, my respectful elder fellows, thanks for sharing me many skills on the experiment. I am so glad to work with you in Rotterdam and wish you all and your family healthy and happy.

To Zhijiang Miao, Peifa Yu, Yang Li, Yunlong Li, and Ling Wang, we started the PhD study in the same year. We spend most of our time in the biweekly group meeting, MDL meeting as well as Molmed class. Let's remember every Dutch festival and traditional Chinese festival that we celebrate them overseas. I sincerely APPRECIATED working with you at MDL. Have a nice future, my friends.

To Ruyi, you are a smart raising star in tRNA research. It was very nice to cooperate with you, a very hard-working and helpful partner. To Junhong, I was impressed by your deep thinking related to belief and we have many times of enjoyable discussions while we have lunch.

To my roommates in Ee830, Pauline, Suk yee Lam, Rachid, Jorker, Floris, Ruby, Michiel, Gulce, Kateryna, and other colleagues in this office. I am very happy for sharing the office with all of you. Every afternoon on Friday, that is beer and chatting "meeting", a very precious memory for my life.

To all MDL colleagues: Raymond, Leonie, Jan, Auke, Henk, Sonja, Marla, Andre, Hugo, Jaap, Lucia, Natasha, Kelly, Patrick, Anthonie, Marcel, Monique, Shanta, Petra, Lisanne, and everyone in our lab. We are a big family; I want to thank all of you for your help and support during my study in the MDL lab.

To Holsan, Jan Szubert, Ronald, Henk, Shun fong, Pieter, Chin-ho, Akash, Tim, Koen, thank you for your pick up, wonderful communication at Rotterdam, Leiden and Den Haag meeting. Rejoice Always.

To my supervisor and co-supervisors: 感谢四川农业大学预防兽医研究所程安春、汪铭书教授，黑龙江八一农垦大学动物科技学院郭东华教授、黑龙江大学化工与材料学院王彬教授对我研究学习期间的帮助，以及国外求学生活的关心与支持。

To 我的家人和亲人们：感谢亲人们对我们家庭的关心与帮助，祝我的家人亲人们工作顺利，幸福安康！

To my beloved wife, Jingyu, thank you for "across the ocean to see me". Thank you for everything you have done for me and our family. All the best wishes to our bright future. To my little Son, Yiqian (一谦), I hope you to be a modest person always.

Publication list

1. **Ou, X.**, Cao, J., Cheng, A., Peppelenbosch, M.P. and Pan, Q., Errors in translational decoding: tRNA wobbling or misincorporation?. **PLoS genetics**, 2019. 15(3), p.e1008017.

2. **Ou, X.**, Ma, B., Zhang, R., Miao, Z., Cheng, A., Peppelenbosch, M.P. and Pan, Q., A simplified qPCR method revealing tRNAome remodeling upon infection by genotype 3 hepatitis E virus. **FEBS letters**, 2020. 594(12), pp.2005-2015.

3. **Ou, X.**, Peng, W., Yang, Z., Cao, J., Wang, M., Peppelenbosch, M.P., Pan, Q. and Cheng, A., Evolutionarily missing and conserved tRNA genes in human and avian. **Infection, Genetics and Evolution**, 2020. 85, p.104460.

4. Zhang, R., Noordam, L., **Ou, X**., Ma, B., Li, Y., Das, P., Shi, S., Liu, J., Wang, L., Li, P. and Verstegen, M.M., The biological process of lysine‐tRNA charging is therapeutically targetable in liver cancer. **Liver International**, 2021. 41(1), pp.206-219.

5. **Ou, X.**, Wang, M., Mao, S., Cao, J., Cheng, A., Zhu, D., Chen, S., Jia, R., Liu, M., Yang, Q. and Wu, Y., Zhao, X. Zhang, S., Liu, Y., Yu, Y., Zhang, L., Chen, X., Peppelenbosch, M.P., Pan, Q. and Cheng, A., Incompatible translation drives a convergent evolution and viral attenuation during the development of live attenuated vaccine. **Frontiers in cellular and infection microbiology**, 2018. 8, p.249.

6. **Ou, X.**, Mao, S., Cao, J., Cheng, A., Wang, M., Zhu, D., Chen, S., Jia, R., Liu, M., Sun, K. and Yang, Q., Wu, Y., Chen, X., Comparative analysis of virus-host interactions caused by a virulent and an attenuated duck hepatitis A virus genotype 1. **PloS one**, 2017. 12(6), p.e0178993.

7. **Ou, X.**, Mao, S., Cao, J., Ma, Y., Ma, G., Cheng, A., Wang, M., Zhu, D., Chen, S., Jia, R. and Liu, M., The neglected avian hepatotropic virus induces acute and chronic hepatitis in ducks: an alternative model for hepatology. **Oncotarget**, 2017. 8(47), p.81838.

8. **Ou, X.**, Mao, S., Jiang, Y., Zhang, S., Ke, C., Ma, G., Cheng, A., Wang, M., Zhu, D., Chen, S., Jia, R., Liu, M., Sun, K., Yang, Q, wu., Y, Chen, X. Viral-host interaction in kidney reveals strategies to escape host immunity and persistently shed virus to the urine. **Oncotarget** 2016. 8:7336-7349.

9. **Ou, X.**, Yang, Z., Zhu, D., Mao, S., Wang, M., Jia, R., Chen, S., Liu, M., Yang, Q., Wu, Y., Zhao, X. Zhang, S., huang, J., Gao, Q., Liu, Y., Zhang, L., Peppelenbosch, M.P., Pan, Q. and Cheng, A., Tracing genetic signatures of bat-to-human coronaviruses and early transmission of North American SARS-CoV-2. **Transbound Emerg Dis** 2021,00:1-13. doi: 10.1111/tbed.14148

10. Miao, Z., Zhang, S., **Ou, X.**, Li, S., Ma, Z., Wang, W., Peppelenbosch, M.P., Liu, J. and Pan, Q., 2020. Estimating the global prevalence, disease progression, and clinical outcome of hepatitis delta virus infection. **The Journal of infectious diseases**, 221(10), pp.1677-1687.

11. Li, Y., Qu, C., Yu, P., **Ou, X.**, Pan, Q. and Wang, W., The interplay between host innate immunity and hepatitis E virus. **Viruses**, 2019. 11(6), p.541.

12. Mao, S., **Ou, X.**, Wang, M., Sun, D., Yang, Q., Wu Y., Jia R., Zhu D., Zhao X., Chen, S., Liu, M., Zhang, S., Huang, J., Gao, Q., Liu Y., Yu, Y., Zhang, L., Tian B., Pan, L., Chen, X., Cheng, A. Duck hepatitis A virus 1 has lymphoid tissue tropism altering the organic immune responses of mature ducks. **Transbound Emerg Dis**. 2020.doi:10.1111/tbed.13966.

13. Cao, J., **Ou, X.**, Zhu, D., Ma, G., Cheng, A., Wang, M., Chen, S., Jia, R., Liu, M., Sun, K. and Yang, Q., Wu, Y., Chen, X. The 2A2 protein of Duck hepatitis A virus type 1 induces apoptosis in primary cell culture. **Virus genes**, 2016. 52(6), pp.780-788.

14. Mao, S., Wang, M., **Ou, X.**, Sun, D., Cheng, A., Zhu, D., Chen, S., Jia, R., Liu, M., Sun, K. and Yang, Q., Wu, Y., Zhao, X., Chen, X., Virologic and immunologic characteristics in mature ducks with acute duck hepatitis A virus 1 infection. **Frontiers in immunolog**y, 2017.8, p.1574.

15. Xie, J., Zeng, Q., Wang, M., **Ou, X**., Ma, Y., Cheng, A., Zhao, X.X., Liu, M., Zhu, D., Chen, S. Jia, R., Yang, Q., Wu, Y., Zhang, S., Liu, Y., Yu, Y., Zhang, L. and Chen, X. Transcriptomic characterization of a chicken embryo model infected with duck hepatitis A virus Type 1. **Frontiers in immunology**, 2018.9, p.1845.

16. Zhu, D., Chen, H., **Ou, X.**, Liu, M., Wang, M., Zhao, X., Jia, R., Chen, S., Sun, K., Yang, Q. and Wu, Y., Chen, X. and Cheng, C., Comparison of immunohistochemistry and Ziehl‑Neelsen staining for detecting the distribution of Mycobacterium avium subsp avium in naturally infected domestic Pekin ducks (Anas platyrhynchos domestica). **Veterinary medicine and science**, 2020. 6(2), pp.242-247.

17. Zhu, D.K., Song, X.H., Wang, J.B., Zhou, W.S., **Ou, X.M.**, Chen, H.X., Liu, M.F., Wang, M.S., Jia, R.Y., Chen, S. and Sun, K.F., 2016. Outbreak of avian tuberculosis in commercial domestic Pekin ducks (Anas platyrhynchos domestica). **Avian Diseases**, 60(3), pp.677-680.

PhD Portfolio

Name of PhD student   Xumin Ou

Department   Gastroenterology and Hepatology, Erasmus MC-

University Medical Center, Rotterdam

PhD Period   September 2017 – September 2021

Promotor   Prof. dr. Maikel P. Peppelenbosch

Co-promotor   Dr. Qiuwei Pan

PhD training

Seminars

- 2017-2019, weekly MDL seminar program in experimental Gastroenterology and Hepatology (attending); (42 weeks/year; @1.5h) (ECTS, 4.5)
- 2017-2019, weekly MDL seminar program in experimental Gastroenterology and Hepatology (presenting); (preparation time 16h; 2 times/year) (ECTS, 2.3)
- 2017-2019, biweekly research group education(attending); (20 times/year; @1.5h) (ECTS, 2.1)
- 2017-2019, biweekly research group education(presenting); (preparation time 8h; 4 times/year) (ECTS, 2.3)

General Courses and Workshops

- 2019, Biomedical English Writing Course (ECTS, 1.6)
- 2019, The Data Analysis in Python (ECTS, 1.4)
- 2018, The English Social Skills Course (ECTS, 1.6)
- 2017, The Basic Genetics Course: Genetics for Dummies (ECTS, 0.6)
- 2017, The programming with Python (ECTS, 1.0)
- 2018, The SNP course XV: SNP and human Diseases (ECTS, 2.0)
- 2018, The Course Biomedical Research Techniques VXII (ECTS, 1.5)
- 2018, The Photoshop and illustrator CS6 Workshop for PhD students (ECTS, 0.3)
- 2018, The workshop on Microsoft Excel 2010: Advanced (ECTS, 0.4)

- 2018, The Course in Virology 2018 (ECTS, 1.4)

- 2017, The workshop Writing successful Grant Proposal (ECTS, 0.5)

- 2018, The Microscopic Image Analysis: From theory to Practices (ECTS, 0.8)

- 2018, The Basic and Translational Oncology 2018 (ECTS, 1.8)

- 2018, Research Integrity (ECTS, 0.8)

- 2018, The Basic Introduction Course on SPSS (ECTS, 1.0)

National and International Conferences

- 2017, July. 5th Youth BBS for Molecular Biotechnology of Avian Disease, China (Abstract)

- 2018, March. 22th Molecular medicine day, Erasmus MC, Netherlands (Poster)

- 2019, March. 23th Molecular medicine day, Erasmus MC, Netherlands (Poster)

- 2018, 33rd Erasmus Liver Day, Rotterdam, Netherlands

Academic Awards and Scientific grants

- 2010, October. National Encouragement scholarship (¥5,000), China

- 2012, October. National Encouragement scholarship (¥5,000), China

- 2017, China Scholarship Council (CSC) Scholarship (File No. 201706910003)

- 2018, National Scholarship for Graduate Students (¥30,000), China

- 2020, Science and Technology Program of Sichuan Province (¥100,000, 2020YJ0396)

- 2021, National Natural Science Foundation of China (¥240,000, 3210190289)

Curriculum Vitae

Xumin Ou was born on November 30th, 1988 in Qingshen, Sichuan province, China. He was raised and finished his primary, secondary, and high school in his hometown.
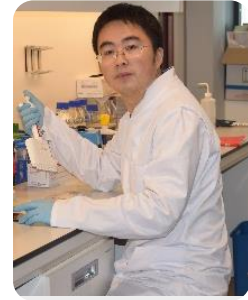
In 2007, he started his college study in veterinary medicine at the Heilongjiang University of Technology. In 2010, he initiated his bachelor study at Heilongjiang Bayi Agricultural University, China. Two years later, he formally started his Master's research training at the Institute of Preventive Veterinary Medicine, Sichuan Agricultural University. Under the supervision of Prof. Anchun Cheng and Prof. Mingshu Wang, he worked on the immune response induced by an avian attenuated hepatotropic virus, that is duck Hepatitis A virus chicken embryo-attenuated vaccine strain.

In 2017, with the support of the China Scholarship Council (CSC), he got an opportunity to start his PhD research at the Department of Gastroenterology and Hepatology, Erasmus University Medical Center, Rotterdam, the Netherlands. Under the supervision of Prof. Maikel P. Peppelenbosch and Dr. Qiuwei (Abdullah) Pan, he focused on the tRNA decoding biology in Hepatitis Viruses and Liver Cancer.