



# uORF-Mediated Translational Regulation of *ATF4* Serves as an Evolutionarily Conserved Mechanism Contributing to Non-Small-Cell Lung Cancer (NSCLC) and Stress Response

Wenjing Xiao<sup>1</sup> · Yang Sun<sup>1</sup> · Jinpeng Xu<sup>1</sup> · Na Zhang<sup>1</sup> · Lina Dong<sup>1</sup>

Received: 7 March 2022 / Accepted: 20 July 2022 / Published online: 13 August 2022  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Diseases and environmental stresses are two distinct challenges for virtually all living organisms. In light of evolution, cellular responses to diseases and stresses might share similar molecular mechanisms, but the detailed regulation pathway is not reported yet.

We obtained the transcriptomes and translomes from several NSCLC (non-small-cell lung cancer) patients as well as from different species under normal or stress conditions. We found that the translation level of gene *ATF4* is remarkably enhanced in NSCLC due to the reduced number of ribosomes binding to its upstream open reading frames (uORFs). We also showed the evolutionary conservation of this uORF-*ATF4* regulation in the stress response of other species. Molecular experiments showed that knockdown of *ATF4* reduced the cell growth rate while overexpression of *ATF4* enhanced cell growth, especially for the *ATF4* allele with mutated uORFs. Population genetics analyses in multiple species verified that the mutations that abolish uATGs (start codon of uORFs) are highly deleterious, suggesting the functional importance of uORFs.

Our study proposes an evolutionarily conserved pattern that enhances the *ATF4* translation by uORFs upon stress or disease. We generalized the concept of cellular response to diseases and stresses. These two biological processes may share similar molecular mechanisms.

**Keywords** Non-small-cell lung cancer (NSCLC) · *ATF4* · Upstream open reading frame (uORF) · Translation regulation · Evolutionarily conserved · Stress response

## Abbreviations

ATF4	Activating transcription factor 4.
DEG	Differentially expressed gene.
NSCLC	Non-small-cell lung cancer.
TF	Transcription factor.
mRNA	Messenger RNA.
TE	Translation efficiency.
UTR	Untranslated region.
CDS	Coding sequence.
uORF	Upstream open reading frame.
TPM	Transcripts per million mapped reads.
RPKM	Reads per kilobase per million mapped reads.
PCC	Pearson correlation coefficient.

MEF	Mouse embryonic fibroblast.
NC	Negative control.
UPR	Unfolded protein response.
IRES	Internal ribosome entry site

## Introduction

Non-small-cell lung cancer (NSCLC) is a major type of lung cancer that is still causing death every year. Studying the molecular mechanism underlying NSCLC oncogenesis is certainly important. Clarification of the cellular response to such diseases would help people understand the regulation of oncogenesis at the molecular level. From a more general and broader point of view, stress response, which shares several similarities with disease response, is also challenging for virtually all living organisms (Feder and Krebs, 1997; Kilian et al., 2007; Yang et al., 2019). Either the environmental changes or the diseases will force the organisms to reprogram their cellular system/component to adapt to the

Handling editor: Konstantinos Voskarides.

✉ Wenjing Xiao  
Wjxiaomed@163.com

<sup>1</sup> Department of Radiotherapy, The Affiliated Hospital of Qingdao University, Qingdao, Shandong, China

current condition (Arella et al., 2021; Chu and Wei, 2019; Li et al., 2020a; Zhang et al., 2021b). Therefore, environmental stress and diseases are usually interconnected at the molecular level and are simultaneously studied in various pieces of literatures (Du et al., 2015; Hoeijmakers et al., 2017; Kosuge et al., 2018). This raises the evolutionary relevance between diseases and stresses, which might be essentially the same for the cellular defense and immune systems. A typical example is virus infection (Li et al., 2020a; Zhang et al., 2021a, 2022). At the individual level, virus infection is regarded as disease, but at the cellular level, virus infection is essentially a kind of stress. This analogy further suggests that diseases and stresses have various common features. However, the detailed regulation mechanisms together with target genes are largely unknown.

Intriguingly, a gene named activating transcription factor 4 (*ATF4*), which plays a role in cancers, is also linked with multiple stress responses like nutrient deprivation (starvation) (Ye et al., 2010) and hypoxia (Rzymiski et al., 2010). Studies on either of the two aspects (disease or stress) will help understand the molecular mechanisms underlying how the cells and organisms take action to adapt to environmental fluctuation. The case of *ATF4* promotes us to find a unified model to link disease and stress. We looked for related literatures and found that traditional stress or disease studies are largely based on differentially expressed genes (DEG). For example, upon salt stress, the model plant *Arabidopsis thaliana* would up and down-regulate particular genes to alleviate the severity (Chu and Wei, 2020; Jin et al., 2021; Liu et al., 2020; Wu et al., 2019). The neuron-specifically expressed genes guide the mouse behavior during starvation stress (Hellsten et al., 2017). Upon the heat and cold stress, the fruitflies change the alternative splicing mode or modification patterns to activate the heat shock protein (Desrosiers and Tanguay, 1986; Fujikake et al., 2005). So far, the regulatory networks upon environmental changes or diseases are mainly studied at the transcription level (namely differential expression analysis) but less studied at the translation level (Lukoszczek et al., 2016). More importantly, there lacks a unified model (e.g., the role of a particular *cis* element) to explain the DEG or differentially translated genes.

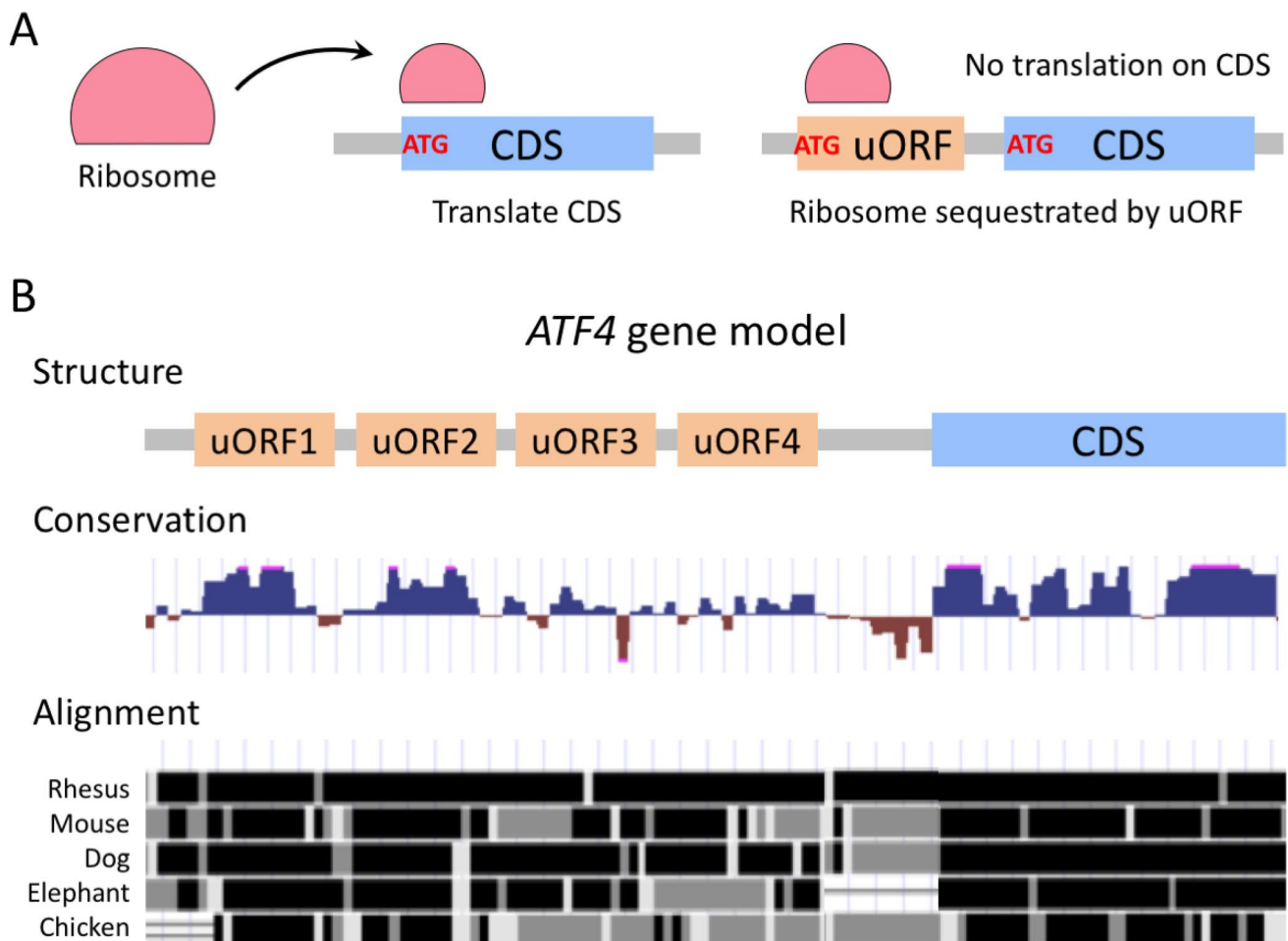
mRNA translation is the fundamental biological process that is as essential as the transcription process. Translation is highly regulated. The natural selection pressure acts on the translation initiation (Wang et al., 2021; Zhang et al., 2022) or elongation rates (Chu and Wei, 2021b; Li et al., 2021, 2020c; Yu et al., 2021) suggests the necessity for maintaining a normal translation rate. To date, the most powerful tool for translational studies is the ribosome profiling technique (Ingolia et al., 2009, 2011). This technique captures and sequences the mRNA fragments being translated by ribosomes (usually around 30 bp long), providing global and local maps of ribosome occupancy (translation rate).

Among the various *cis* and *trans* determinants of the translation efficiency (TE) of genes, the most influential element is the short upstream open reading frame (uORF) (Chew et al., 2016) in the 5'UTR regions that starts with an ATG triplet (Fig. 1A). uORFs are located upstream of the main CDS and serve as roadblocks to inhibit the translation of the main CDS. The more ribosomes blocked by uORFs, the less ribosomes would reach the main CDS. Thus, uORFs are strong inhibitors of CDS translation.

Interestingly, the regulation mediated by uORFs is associated with many stress responses and diseases in a wide range of species. In humans, hyperosmotic stress reduces the translation efficiency of *MDM2* and *eIF2D* mRNAs via one single uORF on each gene (Akulich et al., 2019). The mutation that abolishes uORFs directly causes human malignancies (Schulz et al., 2018). Engineering the plant genome with uORFs also creates higher disease resistance (Xu et al., 2017). These facts suggest that there is an evolutionarily conserved mechanism to regulate the translation of genes by uORFs during environmental changes and diseases, but a broader range of target genes regulated by uORFs is largely unidentified and the detailed mechanism is unknown.

There was an early report (Harding et al., 2000) on the mechanism of how mammalian PERK and GCN2, two eIF2 kinases, suppress the translation of most mRNAs but specifically increase *ATF4* mRNA translation. This provides us with an example of the uORF-regulated gene. The authors concluded that the evolutionarily conserved uORFs in *ATF4* are responsible for translational regulation (Harding et al., 2000). Another study from the same group (Lu et al., 2004) showed that an artificial eIF2- $\alpha$  uncoupled from the stress signaling pathway could single-handedly activate the expression of many stress-induced genes. The authors proposed that both the translational regulation and gene expression activation roles of eIF2- $\alpha$  contribute to cytoprotection (Lu et al., 2004). Meanwhile, there was a detailed report on the molecular mechanism underlying *ATF4* translational control (Vattem and Wek, 2004). Under stress, when the eIF2-GTP complex is scarce, most genes would be translationally inhibited. However, under low eIF2-GTP concentration, ribosome scanning has a higher chance of missing the uORFs and translating the main CDS of *ATF4*, leading to the increase in *ATF4* translation (Vattem and Wek, 2004). This paper nicely explains why *ATF4* behaves conversely to the normal genes under stress. Then, a paper (Chan et al., 2013) introduces a new mechanism that upon unfolded protein response (UPR), the translation of *ATF4* is not controlled by uORFs but mediated by internal ribosome entry site (IRES) in 5'UTR. This study used a human *ATF4* isoform with four uORFs (Chan et al., 2013).

A recent study (Vasudevan et al., 2020) used the *UAS-RNAi* system in *Drosophila melanogaster* to screen the known translation initiation factors required for *ATF4*



**Fig. 1** uORF and *ATF4* gene. **A** Normally, ribosomes will translate the CDS from start codon ATG. If there is uORF in the 5'UTR, then the ATG of uORF will sequester the ribosomes and lead to the translation on uORF, preventing the CDS from being translated. **B** Mammalian gene *ATF4* (activating transcription factor 4) has four uORFs. The conservation level (given by phyloP value) is obvi-

ously higher in the uORF regions compared to the rest of the 5'UTR regions. The sequence alignment also shows high similarity among vertebrates. The black regions mean identical nucleotides with human sequence. Gray regions mean non-identical nucleotides with human sequence. Gaps are shown as thin lines. The gene model is just an example because in reality the uORF could extend into the CDS

translation. The authors found that loss of *eIF2D* and *DENR* would make fruitflies more vulnerable to amino acid deprivation and show phenotypic defects similar to *ATF4* mutant fruitflies. The mechanistic connection between *eIF2D* and *ATF4* is achieved via the uORFs in *ATF4* 5'UTR, what the authors called the “5' leader sequence.” The uORFs mainly control the translation (but not transcription) of the *ATF4* gene (Vasudevan et al., 2020). However, apart from the functional importance of uORFs/*ATF4* verified in lab strains of *Drosophila*, it remains unclear whether these uORFs are well maintained in natural *Drosophila* populations as well as many other model organisms. Intuitively, if the uORFs are functional, then the mutations that abolish uORFs should be deleterious and be suppressed in natural populations. This question could be answered by investigating the population SNP data.

Based on the prevalence of uORFs in the genome, we believe that the uORF-mediated translation regulation should participate in a much broader range of diseases and stress conditions. Given the fact that the *ATF4* gene has 4 uORFs which are highly conserved in vertebrates (Fig. 1B) and even in invertebrates, we are eager to see whether the uORF-*ATF4*-disease/stress axis exists in NSCLC oncogenesis. We obtained the transcriptomes and translomes from seven NSCLC patients and normal tissue and found that the *ATF4* translation is enhanced in NSCLC due to the reduced number of ribosomes binding to uORFs. To show the conservation of this uORF-*ATF4*-disease/stress pathway, we further sought the transcriptome and translome data of mouse and *Drosophila* upon nutrient deprivation and found exactly the same pattern of enhanced *ATF4* translation and reduced uORF translation under stress. We experimentally

verified the biological function of *ATF4* in the human cell line. Knockdown of *ATF4* reduced the cell growth rate while overexpression of *ATF4* enhanced cell growth, especially for the *ATF4* allele with mutated uORFs. Population genetics analyses in multiple species verified that the mutations that abolish uATGs (start codon of uORFs) are highly deleterious, suggesting the functional importance of uORFs. Our study proposes an evolutionarily conserved mechanism that enhances the *ATF4* translation by uORFs upon stress or disease.

## Methods

### Availability of Data and Material

The transcriptome and translome data were obtained from NCBI with accession IDs ERP105150 (human NSCLC patients), SRP263114 (mouse embryonic fibroblasts), and SRP101682 (*Drosophila* S2 cell). There were seven anonymous NSCLC patients with transcriptome and translome. The reference genomes of humans (*Homo sapiens*), mice (*Mus musculus*), and flies (*Drosophila melanogaster*) were got from the UCSC genome browser website (Kent et al., 2002). The human NSCLC cell line is pursued from the cell bank of the Chinese Academy of Sciences. The human 1000 genome SNPs were downloaded from the link (<ftp://ftp.1000genomes.ebi.ac.uk/>). The latest version was used. The *Drosophila* SNPs were downloaded from the *Drosophila* genetic reference panel (<http://dgrp2.gnets.ncsu.edu/>) with the latest update. The SNPs in the population of *Arabidopsis thaliana* were retrieved from previous literatures (Alonso-Blanco et al., 2016; Chu and Wei, 2021a; Wei, 2020). The sequences and SNPs of the world-wide SARS-CoV-2 population were downloaded from GISAID (Shu and McCauley, 2017) as previously literatures instructed (Liu et al., 2022; Zhu et al., 2022).

### Mapping the Reads

We used tophat (Trapnell et al., 2009) and cufflinks (Ghosh and Chan, 2016) to align the reads to the reference genome. Single mappers were kept for further analysis. Gene expression was measured by RPKM (reads per kilobase per million mapped reads). Translation efficiency  $TE = RPKM_{\text{ribosome}} / RPKM_{\text{mRNA}}$ . mRNA means reads in mRNA-seq, ribosome means reads from ribosome profiling. The mRNA RPKM and TE values are then used to calculate the foldchange between NSCLC and normal white blood cells. The RPKM on CDS and uORFs are calculated with the same pipeline. Since uORFs are usually short, the reads count on all uORFs of the same gene (if the gene has multiple uORFs) is combined. Reads count on individual uORFs (like the multiple

uORFs in the *ATF4* gene) would be shown as special cases. The detailed algorithm defining uORF is introduced below.

### Sequencing Depth

At the genome-wide level, we extracted the sequencing depth on each position with samtools depth (Li et al., 2009). All depth values are first normalized by the sample size, that is, the number of total mapped reads of each library. In the genome, the depth value (namely coverage) of a region (such as the uORF) is the mean depth of each position within this region. The ribosome density of a region is the ratio of ribosome depth to mRNA depth.

### Defining uORFs in the Genome

uORFs are defined, to begin with an ATG in the 5'UTR. The ATG of a uORF is termed uATG. The uORF ends with an in-frame stop codon. The termini of uORFs are not necessarily located in 5'UTR. The uORFs could extend to the CDS. Importantly, the different isoforms derived from alternative splicing might have different CDS and 5'UTR regions. To exclude potential false-positive translation signal on uORFs, we removed the uORF regions that overlapped with any CDS regions (not remove the whole uORF but only remove the overlapped regions). In the SNP analyses, we classified the 5'UTR into three exclusive regions: uATGs, uORFs, and the remaining 5'UTR. The mutations were assigned to a region according to their locations.

### Population Genetic Analysis

The SNP (single-nucleotide polymorphism) files of target species came from the following resources: the human 1000 genome project from (Kuehn, 2008) (<ftp://ftp.1000genomes.ebi.ac.uk/>); the *Drosophila* genetic reference panel from (Mackay et al., 2012) (<http://dgrp2.gnets.ncsu.edu/>); the 1000 genome project of *Arabidopsis thaliana* from (Alonso-Blanco et al., 2016; Chu and Wei, 2021a; Wei, 2020); and the millions of world-wide SARS-CoV-2 sequences from (Liu et al., 2022; Shu and McCauley, 2017; Zhu et al., 2022).

The SNP data were all presented in variant calling format (VCF). Each row of the VCF file is a mutation site. The columns include information on these mutation sites. For the default format of VCF files, the first two columns are the chromosome and position information, telling users the genomic coordinate of the mutation (SNP). The next two columns are the reference nucleotide and alternative nucleotide. For example, if a SNP is an A-to-C mutation, then the reference nucleotide is A, and the alternative nucleotide is C. The fifth column of VCF could be the annotation. For example, whether a SNP is located in gene region or intergenic region, in coding sequence or non-coding sequence, a



missense mutation or a synonymous mutation. The following columns in the VCF file could be variable. The strand information is optionally available, which refers to whether the SNP belongs to a gene located in the positive strand or negative strand of the genome. Then, the allele frequency (AF) information is also provided in the VCF files. AF means the fraction of alleles containing the alternative alleles. Importantly, for a mutation (SNP), higher AF is regarded as more adaptive. Therefore, comparing the AF of different sets of SNPs would tell us which set of mutations are beneficial or deleterious. This kind of analysis belongs to the basic population genetic analysis.

In our analysis, for each of the four species (human, *Drosophila*, *Arabidopsis*, SARS-CoV-2), we extracted all the SNPs in 5'UTR and classified them into three exclusive categories: (1) mutations in uATGs (the ATG of uORFs), (2) mutations in uORF (but not including uATG), and (3) the remaining 5'UTR (the 5'UTR region minus the uORF region). Next, the AF of SNPs of these three regions

was compared to infer their relative adaptiveness and deleteriousness.

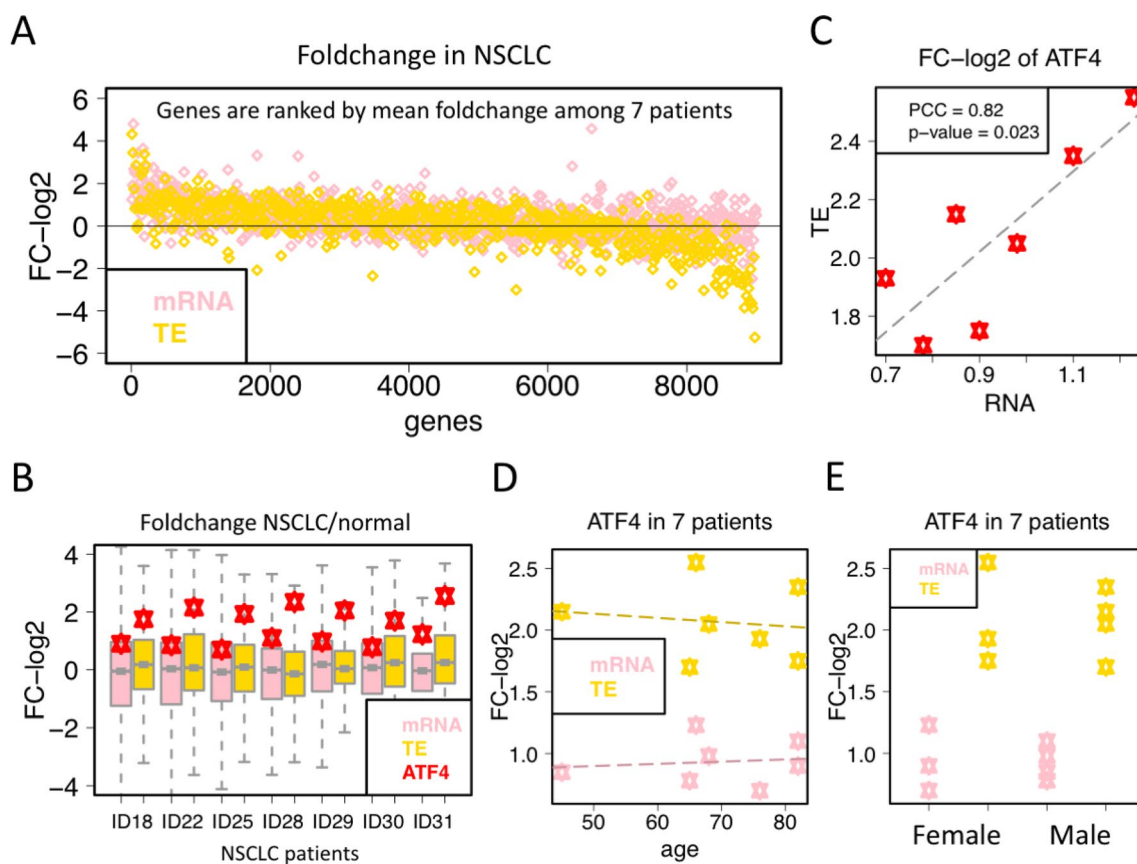
## Graphic Works

The graphic works were plotted in EXCEL or by R language.

## Results

### ATF4 Is Translationally Up-Regulated in NSCLC

In seven NSCLC patients, we quantified the gene expression (mRNA) and translation efficiency (TE) of each gene and calculated the foldchange in NSCLC compared to normal tissues. Foldchange > 0 means up-regulation in NSCLC, and vice versa. To screen for the most significant genes that are differentially expressed, we ranked the genes by the mean foldchange value among seven patients (Fig. 2A). For a



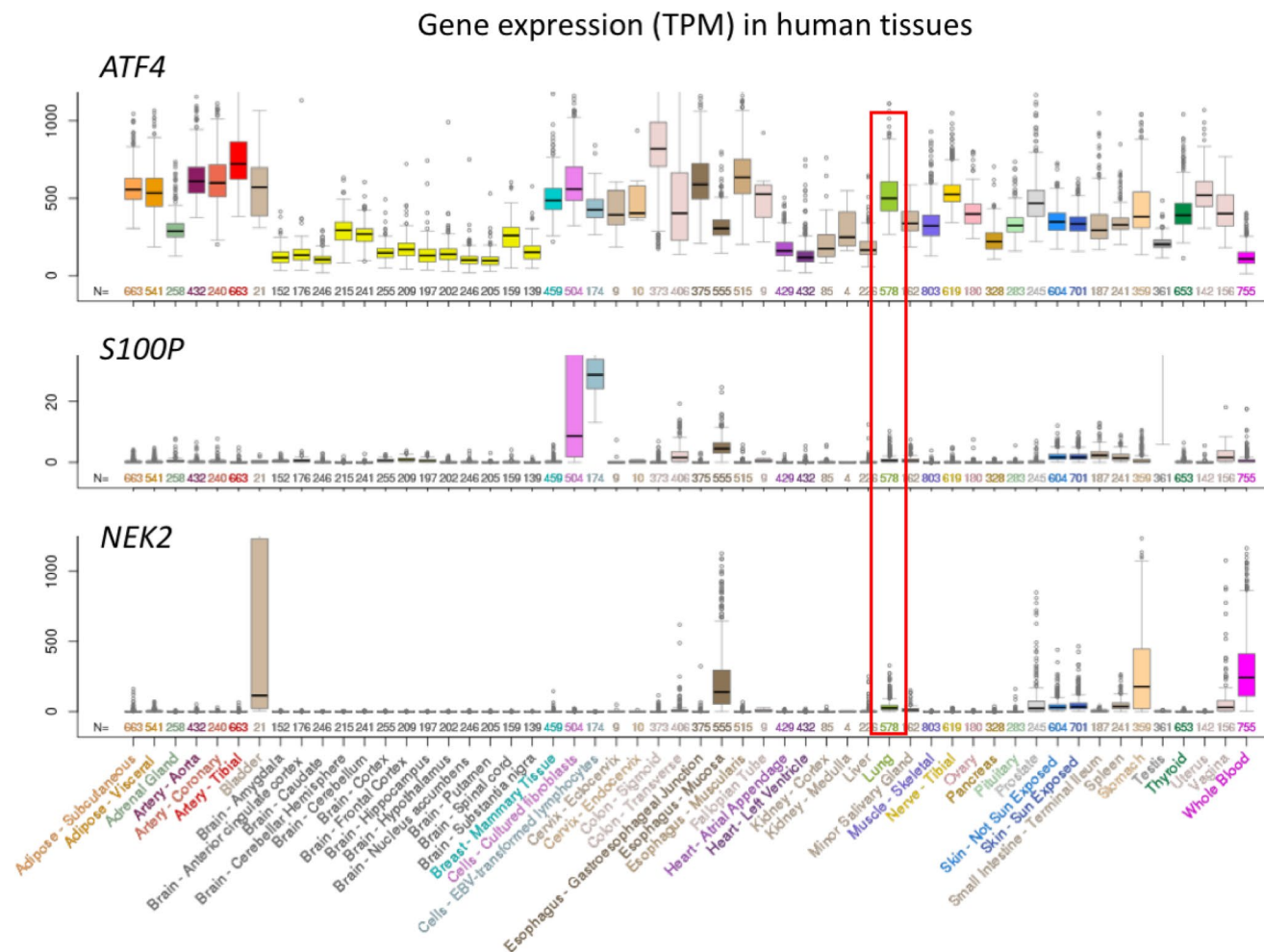
**Fig. 2** Expression and translation of *ATF4* in NSCLC patients. **A** The mRNA and TE foldchange of expressed genes. The genes are ranked by mean foldchange among the 7 NSCLC patients. There are three genes with mRNA and TE foldchange > 0 in all 7 patients. They are *ATF4* (activating transcription factor 4), *S100P* (S100 calcium binding protein P), and *NeK2* (NIMA-related kinase 2). **B** The foldchange

of mRNA expression and TE in seven NSCLC patients compared to normal white blood cells. *ATF4* gene is labeled with red star. **C** Pearson correlation between the mRNA foldchange and TE foldchange of *ATF4* gene in seven NSCLC patients. **D** Correlation between *ATF4* foldchange and the age of NSCLC patients. **E** Relationship between *ATF4* foldchange and the gender of NSCLC patients

particular gene, the up- or down-regulation in one patient does not necessarily represent the tendency in all seven patients. In fact, only very few genes show a consistent direction of foldchange among all seven patients. This phenomenon suggests that the cancer tissue is generally noisy. Under the most stringent criteria, there are three genes with mRNA foldchange > 0 and TE foldchange > 0 in all seven patients. These three genes are up-regulated in NSCLC at both transcription level and translation level: they are *ATF4* (activating transcription factor 4), *S100P* (S100 calcium binding protein P), and *NeK2* (NIMA-related kinase 2). We would check these three genes one by one. We first looked at the well-studied gene *ATF4*. Interestingly, all seven patients showed a higher extent of TE up-regulation than mRNA up-regulation (Fig. 2B). That is to say, although the mRNA level of *ATF4* is already up-regulated in NSCLC, the translation rate is elevated more dramatically. This could lead to much more abundant ATF4 proteins in NSCLC compared

to normal tissues. Moreover, the mRNA and TE foldchange values are positively correlated within the seven patients (Fig. 2C), indicating that the regulation on *ATF4* might be a robust mechanism. In contrast, the extent of *ATF4* up-regulation is not correlated with age (Fig. 2D) or gender (Fig. 2E).

For the other two genes *S100P* and *NeK2*, no striking patterns were observed. The TE foldchange is not always higher or lower than the mRNA foldchange across the seven NSCLC patients. The TE and mRNA foldchange values are not correlated within seven patients (PCC = 0.11,  $p$ -value = 0.72 for *S100P*, and PCC = -0.02,  $p$ -value = 0.93 for *NeK2*), suggesting that the regulation on *S100P* and *NeK2* is not as robust as *ATF4*. The *S100P* and *NeK2* foldchange is not related to age and gender, either. Next, we fully took advantage of the GTEx (genotype-tissue expression) data (Consortium, 2013) and checked the expression of *ATF4*, *S100P*, and *NeK2* (Fig. 3). *ATF4* is generally omnipresent, while *S100P* is poorly expressed in normal



**Fig. 3** Expression (measured by TPM) of genes in human tissues from GTEx data. The rectangle highlights the white blood. These three genes have mRNA and TE foldchange > 0 in all 7 patients.

*ATF4* (activating transcription factor 4), *S100P* (S100 calcium binding protein P), and *NeK2* (NIMA-related kinase 2)

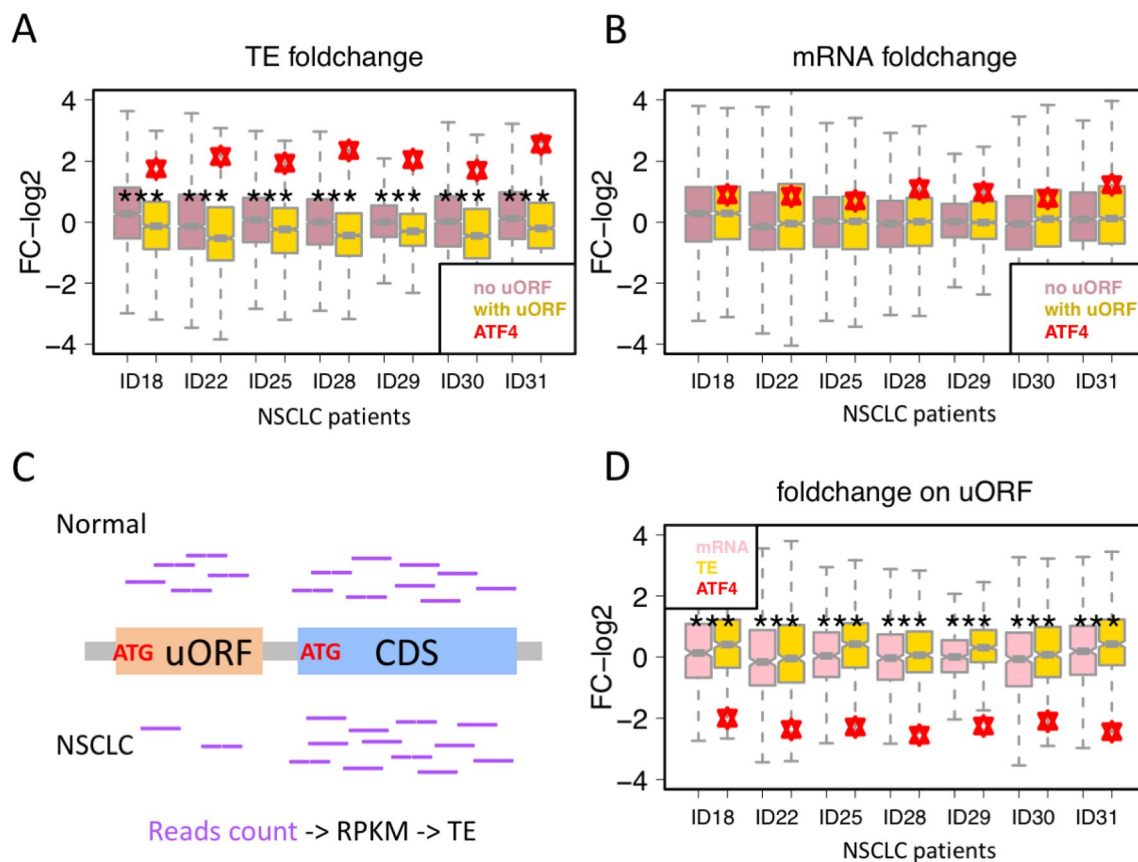
lungs, and *NeK2* is highly expressed in a few tissues including lungs (Fig. 3). We consider that *NeK2* is already highly expressed in normal tissues, so it should not have striking effects when further up-regulated in NSCLC. Therefore, we will focus on *ATF4* in the following analyses.

### Genes with uORFs are Translationally Suppressed Except *ATF4*

We ask what kind of feature leads to the translational up-regulation of *ATF4* in NSCLC. The strongest regulatory *cis* element of translation efficiency is believed to be uORFs located in 5'UTR, upstream of the main CDS. uORFs are also understood as “5' leader sequences” in some literatures (Vasudevan et al., 2020). Coincidentally, *ATF4* has 4 uORFs, whereas most genes only have one uORF or do not have uORFs. The number of uORFs in *ATF4* exceeds the majority of genes. It is extremely likely that the uORFs play a role in regulating the translation of *ATF4* in NSCLC.

Globally, compared to genes without uORFs, the genes with uORFs tend to have lower TE foldchange (Fig. 4A). That is to say, genes with uORFs are translationally suppressed in NSCLC. This agrees with the known concept that uORFs suppress mRNA translation. However, *ATF4* is an exception that is up-regulated in NSCLC (Fig. 4A). To rule out any technical bias, we checked the mRNA foldchange of genes with or without uORFs and found no significant difference (Fig. 4B). This proves that our bioinformatic pipeline does not introduce any biases to the measurement of gene expression level or translation efficiency.

Next, we quantified the reads count, RPKM, TE, and foldchange on uORFs with the same pipeline as on CDS (Fig. 4C). We compared the mRNA foldchange and TE foldchange on uORFs between NSCLC and normal samples and found that the uORF expression is generally unchanged while the translation signals on uORFs are globally increased in NSCLC (Fig. 4D). Note that the uORFs of *ATF4*, which have decreased translation signals in NSCLC, are exceptions compared to other uORFs (Fig. 4D). Indeed,



**Fig. 4** Genes with uORFs are translationally suppressed except *ATF4*. **A** TE foldchange of genes with or without uORFs. *ATF4* gene is labeled with red star. The statistical significance is calculated with KS tests. \*\*\*represents  $p$ -value < 0.001. **B** mRNA foldchange of genes with or without uORFs. *ATF4* gene is labeled with red star. **C** The

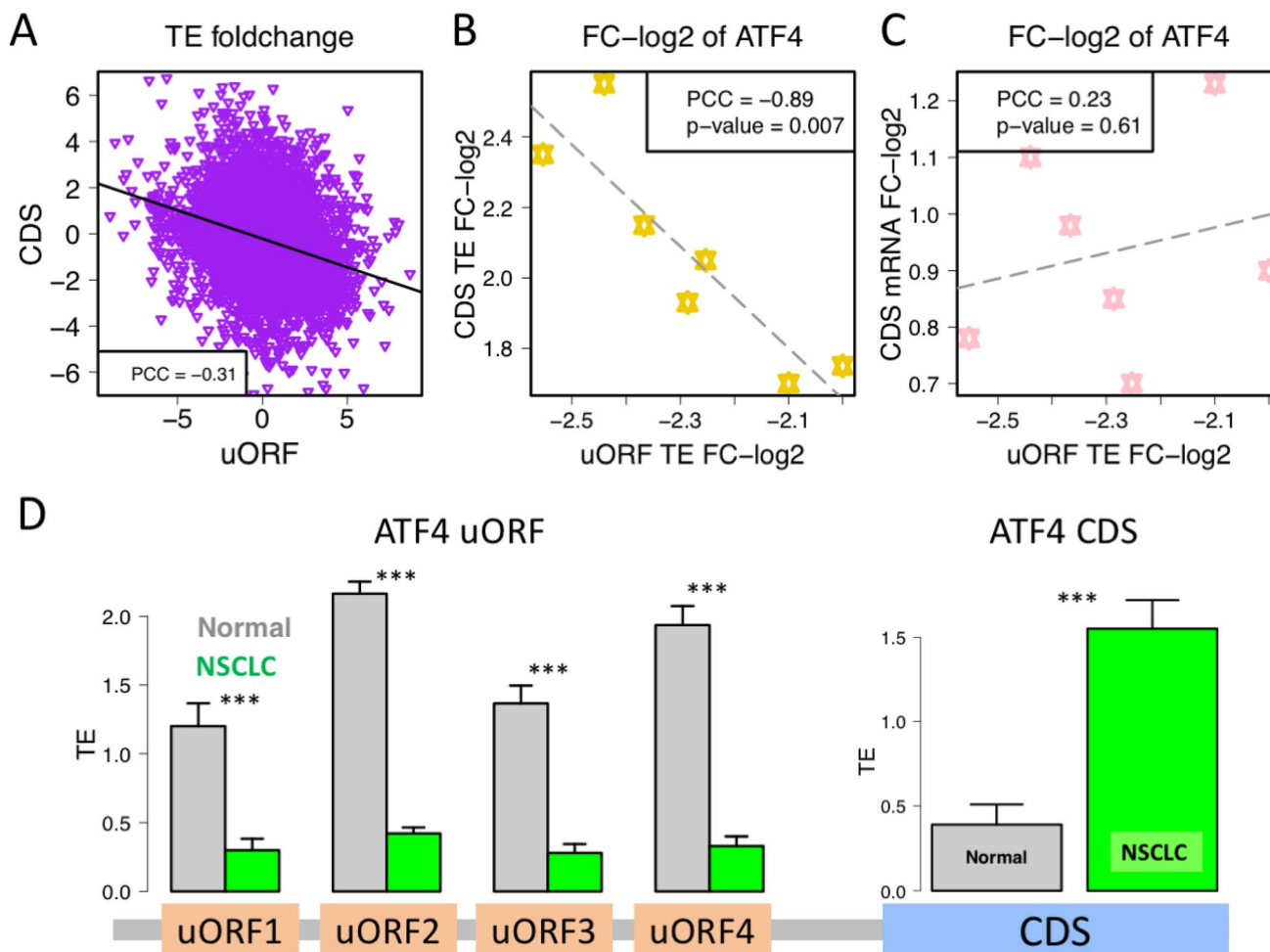
reads count on uORF and CDS to calculate RPKM, TE, and foldchange values. **D** mRNA and TE foldchange of uORFs. *ATF4* gene is labeled with red star. The statistical significance is calculated with KS tests. \*\*\*represents  $p$ -value < 0.001

apart from the uORFs in *ATF4*, there are still many other uORFs that have decreased TE in NSCLC. However, only the uORFs of *ATF4* gene are consistently down-regulated in all seven NSCLC patients. For other uORFs, they displayed inconsistent patterns of up- and down-regulation in different patients. This result also indicates the robust regulation on *ATF4* gene.

### Decreased Translation on uORFs Elevates the Translation of the Main CDS of *ATF4*

It is known that the translation on uORFs would sequester the ribosomes and inhibit the translation of main CDS. For each pair of CDS and uORF (multiple uORFs of the same gene were combined), we compared their TE foldchange

between NSCLC and normal tissues. Expectedly, at the genome-wide level, the foldchange of CDS is negatively correlated with the foldchange of the matched uORF (Fig. 5A). However, it is unexpected that for *ATF4* gene, the CDS TE foldchange and uORF TE foldchange are significantly negatively correlated across the seven NSCLC patients (Fig. 5B). We calculated the Pearson correlation coefficient (PCC) between CDS TE foldchange and uORF TE foldchange across the seven patients, gene by gene, and obtained that the median PCC value of all genes is 0.14, suggesting that most genes do not exhibit a correlation between CDS and uORF TE foldchange across seven patients (although within each sample the two features are negatively correlated). Again, gene *ATF4* is an exception that the CDS TE foldchange and uORF TE foldchange values are negatively correlated across



**Fig. 5** The reduced uORF translation lead to the increased translation in CDS of gene *ATF4*. **A** Pearson correlation between the TE foldchange in each gene and the matched uORF. Multiple uORFs within one gene are combined. Patient ID18 was used to plot this graph. **B** For *ATF4* gene across the seven NSCLC patients, the TE foldchange in CDS is negatively correlated with the TE foldchange in uORF. **C** The mRNA foldchange in CDS is not correlated with the TE fold-

change in uORF. **D** TE of CDS and 4 uORFs in *ATF4* gene. The mean and standard deviation across seven patients were displayed. The gene model is just an example because in reality the uORF could extend into the CDS. The overlapped region between uORF and CDS is not used for reads count and TE calculation. All the comparisons between normal and NSCLC are significant under KS tests ( $p$ -value < 0.001)



seven patients (Fig. 5B). In contrast, the mRNA foldchange of CDS is not correlated with the uORF TE foldchange (Fig. 5C). These results further prove that the down-regulated translation signals on *ATF4* uORFs caused the up-regulated TE on its CDS, agreeing with the known notion that uORFs are translational suppressors of main CDS.

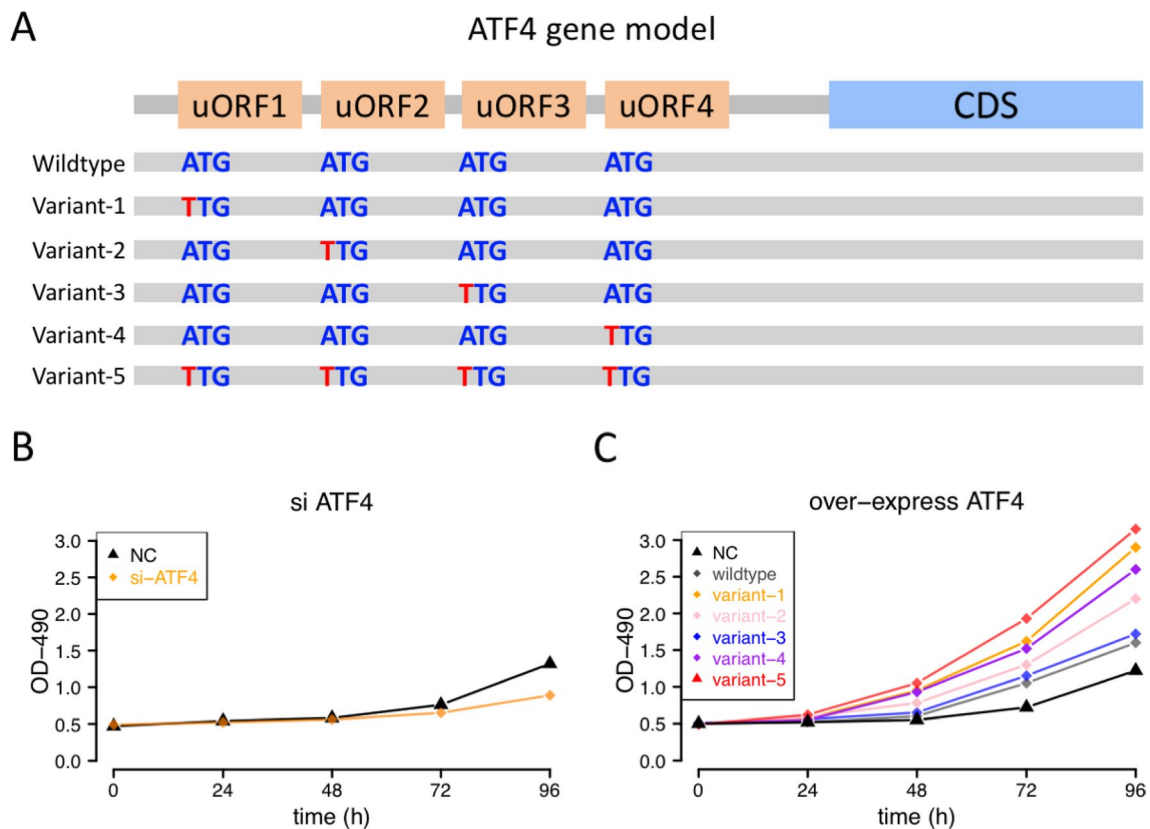
The above analyses have combined the reads of multiple uORFs within the same gene. Since most genes with uORF only have a single uORF, the global anti-correlation between uORF and CDS should be solid. However, genes like *ATF4* have multiple uORFs so those uORFs should be presented separately. Interestingly, we found a robust pattern that all the 4 uORFs of *ATF4* have significantly lower TE in NSCLC, while the TE of main CDS has significantly increased (Fig. 5D).

### Experimental Verification of the uORF-ATF4-NSCLC Axis

We set out to experimentally confirm the role of uORF-mediated translation regulation in NSCLC. We designed five mutant sequences of *ATF4* (Fig. 6A). The ATG of the four uORFs is changed separately (denoted as variant-1 to

variant-4) or changed simultaneously (denoted as variant-5). The start codon ATG of uORF is termed uATG. Note that we only altered the uATGs but did not delete the whole uORF regions for the following reasons: (1) uATG is important for loading the ribosomes onto uORFs. Mutations in uATG are sufficient to abolish the ribosome binding to uORFs; (2) Deleting the whole uORF region would introduce other unpredictable changes like the RNA secondary structure and gene length, which may also affect the translation efficiency of CDS.

We first silenced the *ATF4* gene in human NSCLC cell line (which reduced the *ATF4* expression by 82%) and observed a remarkable reduction in the cell growth rate (Fig. 6B). This preliminary assay demonstrated that *ATF4* is able to promote NSCLC cell growth, but the detailed role and molecular mechanism of uORF are still unclear. Next, we transfected the wildtype and mutant *ATF4* sequences into the cells. In the cells transfected with wildtype *ATF4*, the cell growth slightly increased compared to the negative control (Fig. 6C). In the cells transfected with variants of single-mutation on uATG, the growth rate increased remarkably, suggesting that the abolishment of uORFs has alleviated the ribosome sequestration and thus enhanced the translation



**Fig. 6** The experimental verification of the uORF-ATF4-phenotype axis. **A** Design of five *ATF4* variant sequences. The ATG of the four uORFs is mutated separately (variant-1 to variant-4) or together (vari-

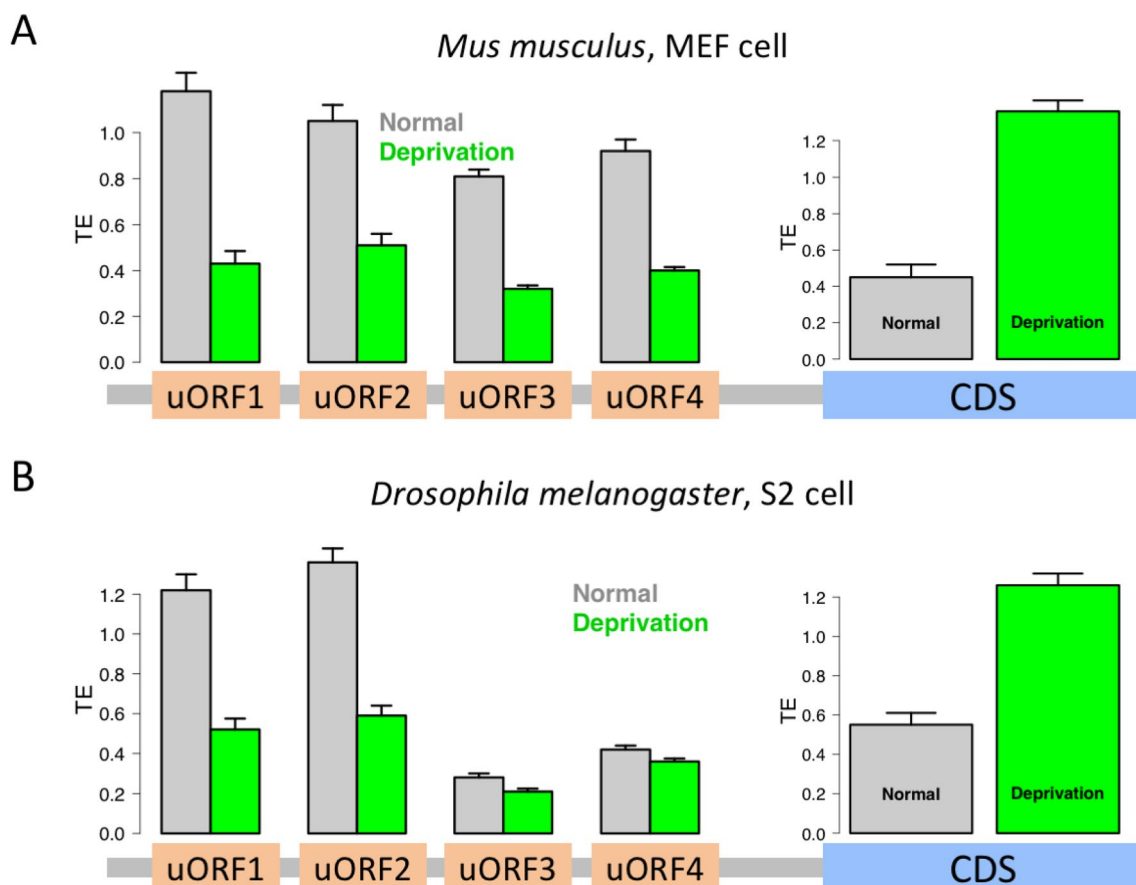
ant-5). **B** Cell growth of human cell line with si-NC (negative control) and the knockdown of *ATF4*. **C** Cell growth under transfection of different *ATF4* variants and NC

of CDS. For the sequence with all four uATG abolished, which is expected to have the strongest translation on *ATF4* main CDS, we observed the highest growth rate among all variants (Fig. 6C). These experimental validations support the uORF-mediated important regulatory role of *ATF4* in oncogenesis.

### The uORF-ATF4-Disease/Stress Axis Is Evolutionarily Conserved in Mouse and Fly

We wonder how general the uORF-ATF4 regulatory mechanism is. Ideally, we should search for samples with similar phenotypes in mammals. Given the rarity of ribosome profiling data compared with the transcriptome data, we were only able to find integrated translome data for cell lines under stress conditions. We selected two representative datasets of mouse embryonic fibroblast (MEF) cells and *Drosophila* S2 cells under normal and nutrient deprivation conditions. In this way, the uORF-ATF4-cancer axis is extended to the uORF-ATF4-disease/stress axis.

In mice, we saw that the translation of *ATF4* uORFs is consistently reduced upon nutrient deprivation while the CDS translation is remarkably elevated (Fig. 7A). This result highlights the conservation of uORF-ATF4 regulation in mammals. In a wider range of species, we looked at the *Drosophila* S2 cells. The *Drosophila* genome also encodes 4 uORFs in gene *ATF4*. This demonstrates the highly conserved *ATF4* sequences across the entire animal kingdom. Upon nutrient deprivation in S2 cells, the translation of all 4 uORFs was reduced while the CDS showed remarkably enhanced TE (Fig. 7B). However, the patterns in *Drosophila* S2 cells are slightly different from what we observed in humans and mice. In S2 cells, only the first two uORFs of *ATF4* were highly translated under normal condition (Fig. 7B), while in mammals, all the 4 uORFs of *ATF4* showed substantial translation signals under normal conditions. The functional regulatory network of *ATF4* may have diverged in invertebrates including insects, where the last two uORFs of *ATF4* lost their translatability as well as their regulatory role in stress response. Nevertheless, the certain



**Fig. 7** The translational changes on uORF and CDS of *ATF4* upon nutrient deprivation. **A** In MEF (mouse embryonic fibroblast) cells, the uORF translation of *ATF4* is reduced while the CDS translation is enhanced upon nutrient deprivation. **B** In *Drosophila* S2 cells, the uORF translation of *ATF4* is reduced while the CDS translation is

enhanced upon nutrient deprivation. The TE is displayed as mean and standard deviation of all the samples and replicates. The gene model is just an example because in reality the uORF could extend into the CDS. The overlapped region between uORF and CDS is not used for reads count and TE calculation

thing is that the uORF-mediated translation regulation of *ATF4* and its downstream effects should be highly conserved across vertebrates and invertebrates.

### Population Data Suggest the Deleteriousness of Abolishing uATGs of uORFs

The translome data showed that *ATF4* translation suppression by uORF is alleviated upon NSCLC or stress. If the uORF-mediated *ATF4* regulation is really essential, then the mutations that abolish uORF should be deleterious. Notably, in theory, abolishing uATG is sufficient to abolish the translation of the whole uORF. We will check the deleteriousness of uATG-loss mutations by using population SNP data.

Population genetics theory dictates that the fitness changes caused by the mutations could be well reflected by the allele frequency (AF) in natural populations (Crow, 1955). In particular, deleterious mutations are usually suppressed to very low frequencies across the population. This golden standard could be used to test the distinct biological consequences of different mutations. Let us check the mutations in 5'UTRs. We classified all mutations in 5'UTR into three distinct groups: (1) mutations in uATGs (the ATG of uORFs), (2) mutations in uORF (the uORF region minus the uATG), and (3) the remaining 5'UTR (the 5'UTR region minus the uORF region) (Fig. 8A). We obtained the population SNP data from multiple species. These SNP data include the human 1000 genome (Kuehn, 2008), *Drosophila* genetic reference panel (Mackay et al., 2012), 1000 genome project of *Arabidopsis thaliana* (Alonso-Blanco et al., 2016; Chu and Wei, 2021a; Wei, 2020), and the SNPs called from millions of global SARS-CoV-2 sequences available in GISAID or relevant literatures (Cai et al., 2022; Li et al., 2020b; Liu et al., 2022; Martignano et al., 2022; Shu and McCauley, 2017; Wei, 2022; Zhao et al., 2022; Zhu et al., 2022; Zong et al., 2022). We utilized these SNP data to test whether the mutations that abolish uORFs are most deleterious. Notably, a mutation in uATG (the start codon of uORF) is sufficient to destroy the uORF. In human, *Drosophila*, *Arabidopsis*, and even SARS-CoV-2 populations, the mutations in uATGs were suppressed to very low allele frequencies compared to the other mutations in 5'UTR (Fig. 8B–E). In contrast, whether the mutations are located in or out of uORF regions did not differ much. For *ATF4* gene (in humans and *Drosophila*), there were several mutations located in its 5'UTR, including the uORF region. The allele frequency of mutation in uATG was extremely low in both humans and *Drosophila* (Fig. 8B, C). These results are perfect indicators of the deleteriousness of mutations in the uATG of uORFs. If the uORFs lose the ability to sequester the ribosomes, then their translational regulatory role would be lost. Also, the results suggest that altering the uATG of uORF rather than the rest of the uORF region was deleterious because

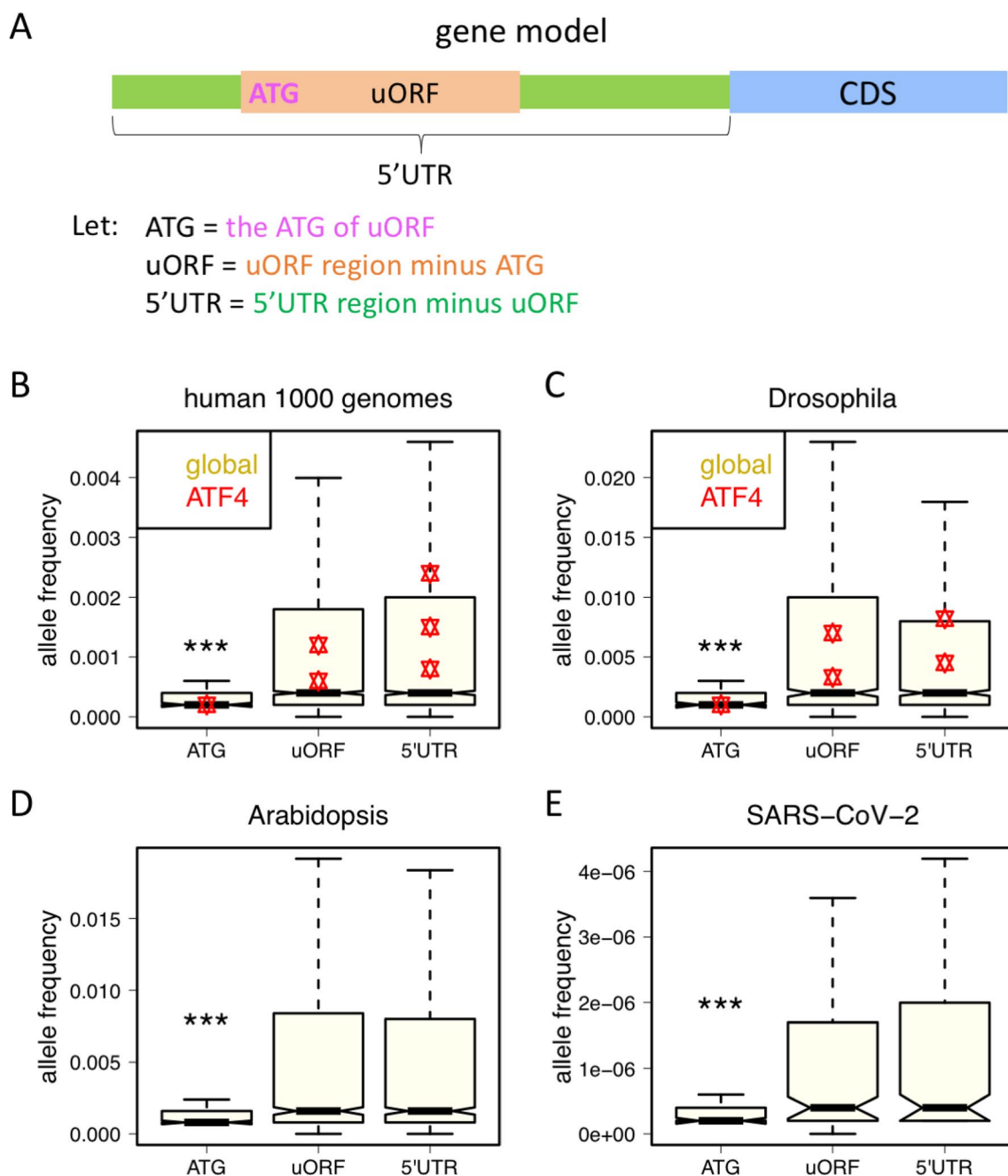
no difference was observed between the mutations in uORF (excluding uATG) and the rest of the 5'UTR (Fig. 8B–E). This result also proves that abolishing uATG is sufficient to abolish the translation (or function) of the whole uORF.

### Discussion

The ribosome profiling technique greatly facilitated translational studies. We have fully utilized this technique. We observed strikingly different patterns in *ATF4* gene compared with other genes under stress conditions or diseases. The CDS translation of the majority of genes was down-regulated due to the increased ribosome sequestration in uORFs. In *ATF4*, however, the ribosome sequestration in uORFs was alleviated under stress or disease, leading to the elevated translation signals in main CDS. This conserved phenomenon between human disease and nutrient deprivation of mice and flies indicates the important regulatory role of uORF and *ATF4* upon diseases or stress.

There are quite a few similarities between cancer like NSCLC and nutrient deprivation stress, where the resource and energy supply is limited. On one hand, the cancer cells require more nutrients and calories to grow and proliferate. On the other hand, upon nutrient deprivation, the cells should save as much energy as they could to maintain the basic requirements. Disease and stress are not only phenotypically similar but also genetically analogous at the molecular level. As we have proposed, virus infection is an excellent example of the connection between disease and stress, where at the individual level it is regarded as a disease but at the cell level, it is treated as stress. The common point between disease and stress is that some unnecessary cellular activities should be shut down to ensure the fundamental needs or fight against the pathogens. This indicates a possible evolutionarily conserved molecular mechanism for the cells to respond to stress and disease. Reducing the translation of most of the genes should be a smart way to avoid unnecessary waste, but meanwhile, a small set of genes like *ATF4* should be up-regulated in response to environmental stimuli. Thus, for *ATF4* gene, the constantly observed uORF translational reduction and the CDS translational enhancement in human disease or mouse/fly stress could be the evolutionarily conserved mechanism that alleviates the food limitation and energy shortage, a possible strategy to get through stress/disease.

Certainly, the cellular system works as an integrated network rather than a simple pathway. Although numerous literatures tried to simplify the cell system as a single pathway, we should admit that the observed changes in *ATF4* are only a small node in the whole cellular network. Other pathways in the network might be equally important.



**Fig. 8** The population genetics analysis on uORFs. **A** Classification of mutations in 5'UTR. **B** Allele frequency of different sets of mutations in the human 1000 genomes. The mutations in *ATF4* gene are highlighted as red stars. The statistical significance is judged by KS test between ATG and the other two sets of mutations. \*\*\*means  $p$ -value < 0.001. **C** Allele frequency of different sets of mutations in the *Drosophila* genetic reference panel. The mutations in *ATF4* gene are highlighted as red stars. The statistical significance is

judged by KS test between ATG and the other two sets of mutations. \*\*\*means  $p$ -value < 0.001. **D** Allele frequency of mutations obtained from the 1000 genome project of *Arabidopsis thaliana*. \*\*\*means  $p$ -value < 0.001. **E** Allele frequency of mutations obtained from the world-wide SARS-CoV-2 sequences. \*\*\*means  $p$ -value < 0.001. These results prove that mutations in uATG are sufficient to abolish uORFs and that these mutations are deleterious

Finally, we carried out population genetics analyses on the mutations in uATGs, uORFs, and 5'UTRs. Evolutionary theories dictate that if a *cis* element is highly functional (such as uORFs, particularly the uATGs), then the mutations

that abolish these *cis* elements would be deleterious so that the allele frequency of such mutations should be very low. From our analyses of the uORF-mediated translational regulation on *ATF4*, we already know that uORFs have crucial



functions in stress response and disease. Therefore, it is intuitive to predict that the mutations that abolish uORF (particularly uATG) should be deleterious. The population SNP data collected by us range from viruses to eukaryotes, including humans, *Drosophila*, *Arabidopsis*, and SARS-CoV-2. These species are sufficient to represent different evolutionary clades. We constantly observed the suppression of mutations in uATGs, suggesting that the abolishment of uORF function is deleterious in all species.

Our study proposes an evolutionarily conserved pattern that enhances the *ATF4* translation by uORFs upon stress or disease. While generalizing the concept of disease and stress which may share similar molecular mechanisms, our results also propose a novel angle to alleviate the stress response or diseases like NSCLC.

**Acknowledgements** We thank all group members for the support to this project.

**Author Contributions** YS, JX, NZ, and LD contributed to acquisition, analysis, and interpretation of data and drafted the manuscript. WX contributed to conception, design of the work, and revision of the manuscript.

**Funding** No funding was received for this current study.

**Data Availability** The transcriptome and translome data were obtained from NCBI with accession IDs ERP105150 (human NSCLC patients), SRP263114 (mouse embryonic fibroblasts), and SRP101682 (*Drosophila* S2 cell). There were seven anonymous NSCLC patients with transcriptome and translome. The reference genomes of humans (*Homo sapiens*), mice (*Mus musculus*), and flies (*Drosophila melanogaster*) were got from the UCSC genome browser website (Kent et al., 2002). The human cell line is pursued from the cell bank of the Chinese Academy of Sciences. The human 1000 genome SNPs were downloaded from the link (<ftp://ftp.1000genomes.ebi.ac.uk/>). The latest version was used. The *Drosophila* SNPs were downloaded from the *Drosophila* genetic reference panel (<http://dgrp2.gnets.ncsu.edu/>) with the latest update.

## Declarations

**Conflict of Interest** The authors declare they have no conflict of interest.

**Ethical Approval** Not applicable.

**Consent for Publication** Not applicable.

## References

- Akulich KA, Sinitcyn PG, Makeeva DS, Andreev DE, Terenin IM, Anisimova AS, Shatsky IN, Dmitriev SE (2019) A novel uORF-based regulatory mechanism controls translation of the human MDM2 and eIF2D mRNAs during stress. *Biochimie* 157:92–101
- Alonso-Blanco C, Andrade J, Becker C, Bemm F, Bergelson J, Borgwardt KM, Cao J, Chae E, Dezaan TM, Ding W et al (2016) 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* 166:481–491
- Arella D, Dilucca M, Giansanti A (2021) Codon usage bias and environmental adaptation in microbial organisms. *Mol Genet Genomics* 296:751–762
- Cai H, Liu X, Zheng X (2022) RNA editing detection in SARS-CoV-2 transcriptome should be different from traditional SNV identification. *J Appl Genet*. <https://doi.org/10.1007/s13353-022-00706-y>
- Chan CP, Kok KH, Tang HM, Wong CM, Jin DY (2013) Internal ribosome entry site-mediated translational regulation of ATF4 splice variant in mammalian unfolded protein response. *Biochim Biophys Acta* 1833:2165–2175
- Chew GL, Pauli A, Schier AF (2016) Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nat Commun* 7:11663
- Chu D, Wei L (2019) Characterizing the heat response of *Arabidopsis thaliana* from the perspective of codon usage bias and translational regulation. *J Plant Physiol* 240:153012
- Chu D, Wei L (2020) Reduced C-to-U RNA editing rates might play a regulatory role in stress response of *Arabidopsis*. *J Plant Physiol* 244:153081
- Chu D, Wei L (2021a) Context-dependent and -independent selection on synonymous mutations revealed by 1,135 genomes of *Arabidopsis thaliana*. *BMC Ecol Evol* 21:68
- Chu D, Wei L (2021b) Direct in vivo observation of the effect of codon usage bias on gene expression in *Arabidopsis* hybrids. *J Plant Physiol* 265:153490
- Consortium GT (2013) The genotype-tissue expression (GTEx) project. *Nat Genet* 45:580–585
- Crow JF (1955) General theory of population genetics: synthesis. *Cold Spring Harb Symp Quant Biol* 20:54–59
- Desrosiers R, Tanguay RM (1986) Further characterization of the post-translational modifications of core histones in response to heat and arsenite stress in *Drosophila*. *Biochem Cell Biol* 64:750–757
- Du X, Pang TY, Mo C, Renoir T, Wright DJ, Hannan AJ (2015) The influence of the HPG axis on stress response and depressive-like behaviour in a transgenic mouse model of Huntington's disease. *Exp Neurol* 263:63–71
- Feder ME, Krebs RA (1997) Ecological and evolutionary physiology of heat shock proteins and the stress response in *Drosophila*: complementary insights from genetic engineering and natural variation. *EXS* 83:155–173
- Fujikake N, Nagai Y, Popiel HA, Kano H, Yamaguchi M, Toda T (2005) Alternative splicing regulates the transcriptional activity of *Drosophila* heat shock transcription factor in response to heat/cold stress. *FEBS Lett* 579:3842–3848
- Ghosh S, Chan CK (2016) Analysis of RNA-Seq data using top hat and cufflinks. *Methods Mol Biol* 1374:339–361
- Harding HP, Novoa I, Zhang Y, Zeng H, Wek R, Schapira M, Ron D (2000) Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol Cell* 6:1099–1108
- Hellsten SV, Eriksson MM, Lekholm E, Arapi V, Perland E, Fredriksson R (2017) The gene expression of the neuronal protein, SLC38A9, changes in mouse brain after in vivo starvation and high-fat diet. *PLoS ONE* 12:e0172917
- Hoeijmakers L, Ruigrok SR, Amelianchik A, Ivan D, van Dam AM, Lucassen PJ, Korosi A (2017) Early-life stress lastingly alters the neuroinflammatory response to amyloid pathology in an Alzheimer's disease mouse model. *Brain Behav Immun* 63:160–175
- Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Sci* 324:218–223
- Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147:789–802

- Jin J, Li K, Qin J, Yan L, Wang S, Zhang G, Wang X, Bi Y (2021) The response mechanism to salt stress in Arabidopsis transgenic lines over-expressing of GmG6PD. *Plant Physiol Biochem* 162:74–85
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) The human genome browser at UCSC. *Genome Res* 12:996–1006
- Kilian J, Whitehead D, Horak J, Wanke D, Weigl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K (2007) The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J* 50:347–363
- Kosuge Y, Osada N, Shimomura A, Miyagishi H, Wada T, Ishige K, Shimba S, Ito Y (2018) Relevance of the hippocampal endoplasmic reticulum stress response in a mouse model of chronic kidney disease. *Neurosci Lett* 677:26–31
- Kuehn BM (2008) 1000 genomes project promises closer look at variation in human genome. *JAMA* 300:2715
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079
- Li Y, Yang X, Wang N, Wang H, Yin B, Yang X, Jiang W (2020a) GC usage of SARS-CoV-2 genes might adapt to the environment of human lung expressed genes. *Mol Genet Genomics* 295:1537–1546
- Li Y, Yang X, Wang N, Wang H, Yin B, Yang X, Jiang W (2020b) Mutation profile of over 4500 SARS-CoV-2 isolations reveals prevalent cytosine-to-uridine deamination on viral RNAs. *Future Microbiol* 15:1343–1352
- Li Y, Yang XN, Wang N, Wang HY, Yin B, Yang XP, Jiang WQ (2020c) The divergence between SARS-CoV-2 and RaTG13 might be over-estimated due to the extensive RNA modification. *Future Virol* 15:341–347
- Li Q, Li J, Yu CP, Chang S, Xie LL, Wang S (2021) Synonymous mutations that regulate translation speed might play a non-negligible role in liver cancer development. *BMC Cancer* 21:388
- Liu Y, Pei L, Xiao S, Peng L, Liu Z, Li X, Yang Y, Wang J (2020) AtPPRT1 negatively regulates salt stress response in Arabidopsis seedlings. *Plant Signal Behav* 15:1732103
- Liu X, Liu X, Zhou J, Dong Y, Jiang W, Jiang W (2022) Rampant C-to-U deamination accounts for the intrinsically high mutation rate in SARS-CoV-2 spike gene. *RNA* 28:917–926
- Lu PD, Jousse C, Marciniak SJ, Zhang Y, Novoa I, Scheuner D, Kaufman RJ, Ron D, Harding HP (2004) Cytoprotection by pre-emptive conditional phosphorylation of translation initiation factor 2. *EMBO J* 23:169–179
- Lukoszek R, Feist P, Ignatova Z (2016) Insights into the adaptive response of Arabidopsis thaliana to prolonged thermal stress by ribosomal profiling and RNA-Seq. *BMC Plant Biol*. <https://doi.org/10.1186/s12870-016-0915-0>
- Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM et al (2012) The drosophila melanogaster genetic reference panel. *Nature* 482:173–178
- Martignano F, Di Giorgio S, Mattiuz G, Conticello SG (2022) Commentary on “Poor evidence for host-dependent regular RNA editing in the transcriptome of SARS-CoV-2.” *J Appl Genet* 63:423–428
- Rzymiski T, Milani M, Pike L, Buffa F, Mellor HR, Winchester L, Pires I, Hammond E, Ragoussis I, Harris AL (2010) Regulation of autophagy by ATF4 in response to severe hypoxia. *Oncogene* 29:4424–4435
- Schulz J, Mah N, Neuenschwander M, Kischka T, Ratei R, Schlag PM, Castanos-Velez E, Fichtner I, Tunn PU, Denkert C et al (2018) Loss-of-function uORF mutations in human malignancies. *Sci Rep* 8:2395
- Shu Y, McCauley J (2017) GISAID: global initiative on sharing all influenza data—from vision to reality. *Euro Surveill*. <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494>
- Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105–1111
- Vasudevan D, Neuman SD, Yang A, Lough L, Brown B, Bashirullah A, Cardozo T, Ryoo HD (2020) Translational induction of ATF4 during integrated stress response requires noncanonical initiation factors eIF2D and DENR. *Nat Commun* 11:4677
- Vattem KM, Wek RC (2004) Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc Natl Acad Sci USA* 101:11269–11274
- Wang Y, Gai Y, Li Y, Li C, Li Z, Wang X (2021) SARS-CoV-2 has the advantage of competing the iMet-tRNAs with human hosts to allow efficient translation. *Mol Genet Genomics* 296:113–118
- Wei L (2020) Selection on synonymous mutations revealed by 1135 genomes of Arabidopsis thaliana. *Evol Bioinform Online* 16:1176934320916794
- Wei L (2022) Reconciling the debate on deamination on viral RNA. *J Appl Genet*. <https://doi.org/10.1007/s13353-022-00706-y>
- Wu D, Cui M, Hao Y, Liu L, Zhou Y, Wang W, Xue A, Chingin K, Luo L (2019) In situ study of metabolic response of Arabidopsis thaliana leaves to salt stress by neutral desorption-extractive electrospray ionization mass spectrometry. *J Agric Food Chem* 67:12945–12952
- Xu G, Yuan M, Ai C, Liu L, Zhuang E, Karapetyan S, Wang S, Dong X (2017) uORF-mediated translation allows engineered plant disease resistance without fitness costs. *Nature* 545:491–494
- Yang R, Hong YC, Ren ZZ, Tang K, Zhang H, Zhu JK, Zhao CZ (2019) A role for PICKLE in the regulation of cold and salt stress tolerance in Arabidopsis. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2019.00900>
- Ye J, Kumanova M, Hart LS, Sloane K, Zhang H, De Panis DN, Bobrovnikova-Marjon E, Diehl JA, Ron D, Koumenis C (2010) The GCN2-ATF4 pathway is critical for tumour cell survival and proliferation in response to nutrient deprivation. *EMBO J* 29:2082–2096
- Yu YY, Li Y, Dong Y, Wang XK, Li CX, Jiang WQ (2021) Natural selection on synonymous mutations in SARS-CoV-2 and the impact on estimating divergence time. *Future Virol* 16:447–450
- Zhang Y, Jin X, Wang H, Miao Y, Yang X, Jiang W, Yin B (2021a) compelling evidence suggesting the codon usage of SARS-CoV-2 adapts to human after the split from RaTG13. *Evol Bioinform Online* 17:11769343211052012
- Zhang YP, Jiang W, Li Y, Jin XJ, Yang XP, Zhang PR, Jiang WQ, Yin B (2021b) Fast evolution of SARS-CoV-2 driven by deamination systems in hosts. *Future Virol* 16:587–590
- Zhang Y, Jin X, Wang H, Miao Y, Yang X, Jiang W, Yin B (2022) SARS-CoV-2 competes with host mRNAs for efficient translation by maintaining the mutations favorable for translation initiation. *J Appl Genet* 63:159–167
- Zhao MM, Li CX, Dong Y, Wang XK, Jiang WQ, Chen YG (2022) Nothing in SARS-CoV-2 makes sense except in the light of RNA modification? *Future Virol*. <https://doi.org/10.2217/fvl-2022-0043>
- Zhu L, Wang Q, Zhang W, Hu H, Xu K (2022) Evidence for selection on SARS-CoV-2 RNA translation revealed by the evolutionary dynamics of mutations in UTRs and CDSs. *RNA Biol* 19:866–876
- Zong J, Zhang Y, Guo F, Wang C, Li H, Lin G, Jiang W, Song X, Zhang X, Huang F et al (2022) Poor evidence for host-dependent regular RNA editing in the transcriptome of SARS-CoV-2. *J Appl Genet* 63:413–421

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.