THE ROLES OF CODON USAGE IN TRANSLATIONAL AND TRANSCRIPTIONAL REGULATION ON GENE EXPRESSION

APPROVED BY SUPERVISORY COMMITTEE

Yi Liu, Mentor

Benjamin Tu, Chair

Melanie Cobb, Ph.D.

Philip Thomas, Ph.D.

Michael Buszczak, Ph.D.

DEDICATION

To my parents, for their endless support and trust.

To my friends, for their accompany on this journey.

THE ROLES OF CODON USAGE IN TRANSLATIONAL AND TRANSCRIPTIONAL REGULATION ON GENE EXPRESSION

by

QIAN YANG

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

December, 2020

THE ROLES OF CODON USAGE IN TRANSLATIONAL AND TRANSCRIPTIONAL REGULATIONS ON GENE EXPRESSION

Publication No. _____

Qian Yang, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2020

Supervising Professor: Yi Liu, Ph.D.

Codon usage bias refers to the universal feature of genome that synonymous codons are used at different frequency. Codon usage plays critical roles in determining gene expression levels through impacting multiple fundamental cellular processes. Although the correlation between codon usage and gene expression level has long been observed, the underlying mechanisms are still largely unclear. In the first part, I demonstrate the mechanism of codon usage determining translation efficiency. The slow decoding rate of nonoptimal codon causes ribosome stalling on mRNA, which leads to premature termination of translation and reduced protein production. This process is conserved from *Neurospora* to *Drosophila*. In addition, I demonstrate that the premature termination of translation is

mediated by the canonical release factor eRF1, which recognizes ribosomes stalled at nonoptimal codons. Together, I propose a model that explains the impact of codon usage on translation efficiency. In the second part, I investigate the role of promoter in codon-dependent gene expression at transcription level. I show that codon usage and sequence downstream of core promoter act in concert to determine the transcription level. Moreover, I identify the regulatory element in *Hsp70* promoter that upregulates the transcription of genes containing nonoptimal codons. The differential transcriptional level is achieved through epigenetic regulations affecting nucleosome density and H3K27ac level and premature termination pathway mediated by Ars2, NEXT complex and nuclear exosome. Collectively, these results show that codon usage and promoter impact transcription through multilayer regulatory mechanisms.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

Yang, Q., Lyu, X., Zhao, F. and Liu, Y., 2020. Promoter determines the codon usage effect on gene expression levels. (submitted)

Yang, Q., Yu, C.H., Zhao, F., Dang, Y., Wu, C., Xie, P., Sachs, M.S. and Liu, Y., 2019. eRF1 mediates codon usage effects on mRNA translation efficiency through premature termination at rare codons. *Nucleic acids research*, 47(17), pp.9243-9258.

Lyu, X., **Yang, Q**., Li, L., Dang, Y., Zhou, Z., Chen, S. and Liu, Y., 2020. Adaptation of codon usage to tRNA I34 modification controls translation kinetics and proteome landscape. *PLoS Genetics*, 16(6), p.e1008836.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Ars2 – arsenic resistance protein 2

bp – base pair

CAI – codon adaptation index

CDE – codon usage-dependent element

ChIP – chromatin immunoprecipitation

CTD – carboxyl-terminal domain

DPE – downstream core promoter element

DSCP – *Drosophila* synthetic core promoter

E. coli – Escherichia coli

eGFP – enhanced green fluorescent protein

eRF1 – eukaryotic release factor 1

H3 – histone 3

H3K27ac – acetylation of histone H3 Lys27

H3K27me3 – tri-methylation of histone H3 Lys27

H3K9me3 – tri-methylation of histone H3 Lys49

HDAC – histone deacetylase

IDR – immediate downstream region

Inr – initiator

KD – knockdown

kDa – kilodalton

KO – knockout

Luc – luciferase

NEXT – nuclear exosome targeting

Per – Period

Pol II – polymerase II

QA – quinic acid

TSS – transcription start site

UAS – upstream activating sequence

WT – wild type

YFP – yellow fluorescent protein

CHAPTER ONE
Introduction

1-1 Codon usage bias

Codon usage bias is a universal phenomenon in all organisms(Ikemura, 1985; Plotkin & Kudla, 2011; Sharp et al., 1986). The 20 amino acids are encoded by 61 codons. Except methionine and threonine, each amino acid has two to six encoding codons, known as the degeneracy of genetic code. The synonymous codons in the genome are used at different frequency, a phenomenon called codon usage bias. Codon usage is a key factor in determining gene expression level through impacting multiple fundamental cellular processes, including co-translational folding (Fu et al., 2016; Yu et al., 2015; Zhao et al., 2017; Zhou et al., 2013), translation fidelity (Kramer & Farabaugh, 2007), SRP recognition (Pechmann et al., 2014), polyadenylation (Zhou et al., 2018) and is related to multiple human diseases (Sauna & Kimchi-Sarfaty, 2011; Shah et al., 2015).

Several indexes were introduced to quantitatively measure codon usage bias. The most commonly used is Codon Adaptation Index (CAI), which reflects the relative adaptiveness of the codon in a gene to codons of the mostly expressed genes(Sharp & Li, 1987). Codon Bias Index (CBI) is another measure of codon usage bias, which compares the codon usage to the most frequently used codons and shows the extent to which a gene uses a subset of optimal codons(Bennetzen & Hall, 1982). Since codon usage bias and tRNA levels are tightly correlated, the tRNA adaptation index (tAI), which includes weights of tRNA

pool, wobble interactions and properties of the ribosome, is now widely used to better reflect the codon usage bias from a mechanistic view(dos Reis et al., 2004).

Codon usage and gene expression

Initially, the synonymous codon was proposed to be functionally redundant(Crick et al., 1961; Nirenberg et al., 1965). However, distinctive patterns of synonymous codons were observed to be used by each organism and also each gene within the same genome, which led to the hypothesis that synonymous codons have a regulatory role in gene expression(Bennetzen & Hall, 1982; Goel et al., 1972; Gouy & Gautier, 1982; Grantham et al., 1981; Ikemura, 1981, 1985; Zhang et al., 1991). Early observations on codon frequency and gene expression level gave rise to the model that gene expression level is correlated with codon frequency. Highly expressed genes usually contain large portion of frequently used codons, whereas genes containing more rare codons are expressed at low level(Duret & Mouchiroud, 1999; Konigsberg & Godson, 1983; Kurland, 1991; Sharp et al., 1986; Sørensen et al., 1989). Subsequent genome-wide sequencing studies highlighted the breadth of this phenomenon, supported with molecular and biochemical assays on individual reporters or disease-related genes(dos Reis et al., 2004; Fu et al., 2018; Jeacock et al., 2018; Lampson et al., 2013; Lavner & Kotlar, 2005; Lithwick & Margalit, 2003; Newman et al., 2016; Shah et al., 2015; Tamir Tuller et al., 2010). Since the discovery of codon usage, numerous studies have shown that codon usage is a major determinant of gene expression through processes including not only translation and co-translational events, such as protein folding, but also mRNA decay and transcription(Bazzini et al., 2016; Boel et al., 2016;

Guimaraes et al., 2014; Mordstein et al., 2020; Presnyak et al., 2015; Zhou et al., 2016; Zhou et al., 2018).

Despite the correlation between codon usage and gene expression level has long been observed, the underlying mechanisms are been revealed recently. During the past decades, many groups dedicated to uncovering the underlying mechanisms of codon usage regulating gene expression have revealed that codon usage participates in almost all the critical processes that are related to gene expression. Conserved from yeast to mammal, codon usage bias is strongly correlated with mRNA half-life and plays a critical role in determining mRNA stability in translation dependent manner(Bazzini et al., 2016; Presnyak et al., 2015; Q. Wu et al., 2019). Slow translation elongation rate triggers mRNA decay via Ccr4-Not complex and Dhh1 de-capping factor(Buschauer et al., 2020; Radhakrishnan et al., 2016). Furthermore, codon usage also impacts mRNA level through regulation during transcription in translation independent ways(Zhou et al., 2016; Zhou et al., 2018). Moreover, codon usage regulating protein synthesis via modulating elongation speed and initiation rate(Chu et al., 2014; Yu et al., 2015).

Codon usage and translation speed

Translation is the primary pathway determining protein synthesis, and it is mainly regulated through initiation and elongation. Since the discovery of the strong correlation of codon usage and gene expression level, a basic model was proposed wherein codon usage regulates translation speed to change protein synthesis rate. Early evidences from reporter assays in model organisms or in cell-free translation assays showed that the synthesis rate of

codon optimal mRNAs is higher than nonoptimal mRNAs, and rare codons are decoded slower than frequent synonymous codons(Bonekamp & Jensen, 1988; Pedersen, 1984; Sørensen et al., 1989; Varenne et al., 1984). In consistent with this model, later computational modeling also emphasis the major role of codon usage in determining translation speed(Brockmann et al., 2007; Man & Pilpel, 2007). With the development of ribosome profiling, which maps the position of translation ribosomes at codon level, it is possible to more accurately measure decoding speed(Ingolia et al., 2009). Although early studies using ribosome profiling did not reveal a clear correlation of codon usage and decoding time(Ingolia et al., 2011; Qian et al., 2012), possibly due to biases introduced during sample preparation(Gerashchenko & Gladyshev, 2014; Hussmann et al., 2015; Santos et al., 2019). However, later improvement on the experimental method along with sophisticated bioinformatic and statistical analyses revealed that codon usage indeed correlate with decoding rate(Dana & Tuller, 2014b; Gardin et al., 2014; Hussmann et al., 2015; Weinberg et al., 2016; C. C.-C. Wu et al., 2019b; Yu et al., 2015). Although this correlation may be weaker in different organisms, generally different codons are decoded at different speeds and it is determined by codon usage(Gobet et al., 2020; Li et al., 2014; C. Pop et al., 2014). Furthermore, biochemical assays measuring translation speed of rare or frequent codons using reporters or on disease-related genes led to similar conclusion that translation speed correlates with codon frequency(Kirchner et al., 2017; Spencer et al., 2012; Yu et al., 2015; Zhao et al., 2017). Moreover, in vivo real-time imaging tracking individual ribosome using fluorescent labeling and super-resolution microscopy further supported that mRNAs containing more optimal codons are translated faster than nonoptimal ones(Wang et

al., 2016; Yan et al., 2016). Despite the abundant evidences showing codon usage determines elongation speed, it remains an open question of how slow elongation leads to low protein level. It is generally proposed that slow decoding rate can generally cause traffic jam and ribosome queuing, and thus inhibit translation(Michael A Ferrin & Arvind R Subramaniam, 2017; Mitarai et al., 2008).

Codon usage and tRNA level

Since the rate-limiting step in elongation cycle is the search and selection of cognate tRNA, whereas transpeptidation and ribosome translocation are usually fast, the effect of codon usage on translation elongation speed is mediated by tRNAs. The decoding rate of each codon is mainly affected by tRNA concentration, charging level and codon-anticodon wobble decoding. Although which factor is dominant is still a matter of debate, it is generally thought tRNA concentration is the primary factor in determining overall decoding speed. Soon after the discovery of nonuniform usage of synonymous codons, a correlation of codon frequency and cognate tRNA level was observed(Bulmer, 1987; Ikemura, 1981, 1985; Moriyama & Powell, 1997). Since tRNA concentrations also correlate with tRNA gene copy number in the genome, tRNA repertoires and codon usage are believed to have coevolved to achieve balance between demand and supply, thus efficient gene expression(Bulmer, 1987; dos Reis et al., 2004; Duret, 2000, 2002; Hershberg & Petrov, 2008).

In consistent with this model, recent high-throughput studies provided more accurate quantification of cellular tRNA concentration and better understanding of the relationship of codon usage and tRNA level, and showed strong correlation of codon usage, tRNA

concentration and decoding time(Dana & Tuller, 2014a; C. C.-C. Wu et al., 2019b). In addition of genome-wide studies, data from individual reporters using molecular and biochemical assays in vivo and in vitro led to similar conclusion that tRNA concentration determines decoding rate. Studies using reporter assays showed that the time needed to finish translating mRNA indeed depends on codon usage, and frequent codons are decoded faster than rare codons(Yu et al., 2015; Zhao et al., 2017). Since the effect of codon usage on translation is caused by different tRNA concentration, changing tRNA level would lead to corresponding changes on protein synthesis rate and gene expression. Indeed, overexpressing matching cognate tRNAs would overcome the inhibitory effect on translating mRNAs enriched in rare codons caused by insufficient tRNA(Kirchner et al., 2017; Letzring et al., 2010; Lin et al., 2019; Spencer et al., 2012).

1-2 Ribosome stalling and premature termination of translation

Perturbation on translation elongation leads to production of misfolded or truncated nonfunctional proteins. Translation of problematic mRNAs also leads to generation and accumulation of truncated or extended non-functional proteins. This process is monitored by ribosome as a hub to interacting with various factors and ensuring immediate cleanup of detrimental mRNAs and proteins. Once ribosome encounter aberrant translation events such as stalling or collision, downstream quality control pathways are activated to degrade nascent peptides by proteasome. The faulty mRNAs are degraded by Xrn1 and exosome through no-go decay pathway(Doma & Parker, 2006). Moreover, the translation initiation of the same mRNA is inhibited to further prevent accumulation of aberrant translation products.

Causes of ribosome stalling

Various factors ranging from exocellular stress to cis elements on the mRNA can cause ribosome stalling(Guydosh & Green, 2014; Ishimura et al., 2014; LaRiviere et al., 2006; Simms et al., 2014). The intrinsic sequence of mRNA is critical for control elongation process. Strong mRNA secondary structure is inhibitory to ribosome movements(Mao et al., 2014; Cristina Pop et al., 2014(Chen, 2013 #195)). Poly-adenosine tract and tandem CGA codons in yeast cause ribosome pausing(Dimitrova et al., 2009; Doma & Parker, 2006). Positively charged amino acids interacting with negative charges in the ribosome exit tunnel also slow down ribosome movement(Charneski & Hurst, 2013; Lu & Deutsch, 2008). Peptide bond formation rates vary among amino acids and are especially slow for proline, which leads to translational pauses at polyproline motifs(Gutierrez et al., 2013; Huter et al., 2017; Pavlov et al., 2009; Ude et al., 2013). Moreover, special regulatory sequences including E.coli SecM and *Neurospora* arginine attenuator also stall ribosomes(Fang et al., 2004; Sabi & Tuller, 2015; Tanner et al., 2009).

Rare codon is an important factor causing ribosome pausing or stalling. Due to the low abundance or inefficient decoding of corresponding tRNA, A site of ribosome remains empty. Multiple studies addressed that rare codons combines with upstream and downstream sequence context, which further slowdowns ribosome, cause ribosome stalling at rare codons(Fang et al., 2000; Kwon et al., 2016; Lindhout et al., 1985; Wolin & Walter, 1988). The well-studied rare codon CGA for arginine in S.cerevisiae is known to cause ribosome stalling (Curran, 1995; Letzring et al., 2013). Moreover, di-codons also show a regulatory role in determining elongation speed(Cannarozzi et al., 2010; Gamble et al., 2016).

Dom34 in ribosome quality control pathway

In no-go decay pathway, the A site of stalled ribosomes are recognized by Dom34 and Hbs1, which are homologs of eukaryotic release factor 1 (eRF1) and eRF3(Arango et al.). Dom34 enter the A-site of ribosome to facilitate ribosome dissociation, whereas Hbs1 promotes this process in a GTP dependent way(Pisareva et al., 2011; Shoemaker et al., 2010). Although Dom34 is homolog of eRF1, it lacks the codon recognition motif that is responsible for discriminating between sense and stop codons, and lacks the full extension of the M domain that hydrolysis tRNA and nascent peptide, thus Dom34/Hbs1 complex promotes ribosome subunit dissociation, but not hydrolysis of the peptidyl-tRNA, in a codon-independent manner(Graille et al., 2008; Lee et al., 2007; Shoemaker et al., 2010). Recent progress on the cause of recruitment of Dom34/Hbs1 showed that the ribosome collision of disome is the minimum unit and substrate for activation of quality control and no-go decay pathways(M. A. Ferrin & A. R. Subramaniam, 2017; Ikeuchi et al., 2019; Juszkiewicz et al., 2018; Simms et al., 2017).

Dom34/Hbs1 facilitates no-go decay and ribosome associated quality control, wherein nascent peptides are degraded. The collided disome creates an interface that is recognized by the E3 ligase Hel2 homolog for ubiquitination of the ribosomal proteins. The ribosome is dissociated into subunits by quality control factors. The nascent peptidyl-tRNA is ubiquitinated by E3 ubiquitin ligase and peptidyl-tRNA linkage in hydrolyzed by Vms1(Garzia et al., 2017; Juszkiewicz & Hegde, 2017; Matsuo et al., 2017; Shao et al., 2015; Shao & Hegde, 2014; Shao et al., 2013; Sundaramoorthy et al., 2017; Verma et al.,

2018; Zurita Rendon et al., 2018) . The nascent peptide chains are then extracted from the 60S subunit by the ATPase Cdc48– Ufd1–Npl4 and presented to the 26S proteasome for degradation(Defenouillere et al., 2013; Lyumkis et al., 2014).

More recently, progress on the impacts on other cellular functions showed that this pathway also feedbacks to translation initiation. Ribosome collision activates GCN2 during elongation, which further phosphorylate eIF2α and inhibits initiation(Meydan & Guydosh, 2020; Wu et al., 2020). Moreover, it also inhibits the initiation of the same mRNA via a distinct pathway facilitated by GIGYF2 and 4EBP(Juszkiewicz et al., 2020; Sinha et al., 2020).

eRF1 mediated translation termination

In eukaryotes, canonical release factors eRF1 and eRF3 are only involved in translation termination and is known to specifically recognize stop codons and discriminate sense codons. eRF1 enters ribosome A-site and is responsible for stop codon recognition, whereas eRF3 stimulates this process in a GTP-dependent manner(Bertram et al., 2000; Fan-Minogue et al., 2008; Song et al., 2000). After stop codon recognition, eRF1 takes conformational change and the GGQ domain swings into the catalytic center of the large subunit(Frolova et al., 1999; Wong et al., 2012). eRF1 induces hydrolysis of peptidyl-tRNA, stimulated by ATP-binding cassette protein ABCE1. After peptide release, eRF1 remains bound to post-termination complexes and together with the ABCE1 splits them into 60S and tRNA/mRNA-associated 40S subunits, which is coupled by ATP hydrolysis(Pisarev et al., 2010; Shoemaker & Green, 2011). Subsequent dissociation of tRNA from small subunit are

promoted by factors including eIFs 3, 1 and 1A, Ligatin, or MCT1/DENR(Skabkin et al., 2010; Skabkin et al., 2013).

1-3 Genetic and epigenetic determinants of transcription

Transcription is heavily regulated by a huge number of molecules working in concert and larger networks involving multilayer pathways and factors acting in trans or cis. Trans-acting factors are transcription factors and long noncoding RNAs, which impacts the on and off of transcription and transcription rate. Cis-acting factors are regulatory regions in the chromosome, including sequence elements in the genome such as enhancers, and epigenetic marks(Shlyueva et al., 2014; Zabidi & Stark, 2016). These factors control the initiation of transcription, which is the recruitment and assembly of RNA polymerase and cofactors, the pause release of polymerase to elongation, and the processing and termination of transcription. From initiation to termination, every step during transcription impacts the final level of mature transcripts.

Core promoter

Promoter determines the transcription on and off, whereas enhancers amplify the initiation of transcription at promoter region(Banerji et al., 1981; Shlyueva et al., 2014). Core promoter is the hub for assembly of RNA polymerase II (RNA Pol II) and its associated general transcription factors(Hampsey, 1998). It ranges from about -40 to +40 nucleotides relative to the transcription start site (TSS), and comprises combination of regulatory

elements of TATA box, initiator (Inr), and downstream core promoter element (DPE). The TATA box a is A/T-rich region and the binding site for the transcription initiation factor TATA-binding protein (TBP) subunit of the TFIID complex(Breathnach & Chambon, 1981). Inr is the region encompassing the transcription start site(Smale & Baltimore, 1989; Vo Ngoc et al., 2017). In promoters lacking TATA box, the DPE is often present and is at strict spacing to Inr for the binding of subunits of the TFIID complex (Burke & Kadonaga, 1997; Butler & Kadonaga, 2001; Kutach & Kadonaga, 2000; Louder et al., 2016).

Upstream of core promoter are regions that may contain proximal regulatory elements including enhancers, silencers and insulators. Together with distal regulatory elements that are recruited by conformational changes in the 3-dimensional structure of chromatin, these cis-acting factors modulate and fine tune the activity of Pol II at core promoter(Blackwood & Kadonaga, 1998; Butler & Kadonaga, 2001; Lemon & Tjian, 2000; Zabidi & Stark, 2016).

Epigenetic regulations

Chromatin is composed of DNA and proteins where the fundamental subunit is nucleosome. 147 bp of DNA wraps an octamer of histone proteins (H3, H4, H2A and H2B). Nucleosome packaging restricts proteins such as transcription factors and polymerase to bind to DNA and impacts transcription. For example, nucleosome at promoter region is usually repressive and prevents the initiation of transcription(WASYLYK & CHAMBON, 1979; Wasylyk et al., 1979). Multiple protein complexes function as chromatin remodelers to reposition the nucleosome and facilitates protein-DNA interactions. These chromatin remodelers are classified into four families; SWI/SNF, ISWI, NURD/Mi-2/CHD and INO80.

They all depends on ATP hydrolysis to unwrap, mobilize, exchange or eject the nucleosome(Cairns, 2005; Clapier & Cairns, 2009).

Post-translational modifications on histone proteins also influence the nucleosome interaction with DNA. Histone proteins are recognized and covalently marked by histone modifying enzymes. These modifications include acetylation, methylation, phosphorylation, ribosylation and ubiquitination. Active promoters is associated with high level of nucleosomes bearing tri-methylation of histone H3 Lys4 (H3K4me3) and acetylation of H3 Lys27 (H3K27ac)(Barski et al., 2007; Roh et al., 2004; Schübeler et al., 2004). Despite the strong correlation, the exact role of H3K4me3 and H3K27ac in transcription activation is unclear.

Transcription repression is mediated by two histone modifications: tri-methylation of histone H3 Lys49 (H3K9me3) and tri-methylation of histone H3 Lys27 (H3K27me3). H3K9me3 is typically enriched in constitutive heterochromatin region, and also found in loci of silenced genes and tandem repeat elements(Grewal & Jia, 2007; Huisinga et al., 2006). Heterochromatin protein 1 family proteins bind to methylated H3K9 and lead chromatin condensation(Bannister et al., 2001; Lachner et al., 2001; Nakayama et al., 2001). Heterochromatin protein 1 also recruit repressors to modulate gene silencing(Motamedi et al., 2008). H3K27me3 is generally enriched in the facultative heterochromatin region, which is developmentally regulated(Margueron & Reinberg, 2011; Morey & Helin, 2010). H3K27me3 is added by polycomb repressive complex 2(Cao et al., 2002). H3K27me3 further recruits PRC1 to monoubiquitination of H2A at K119 and inhibits Pol II transcription(Stock et al., 2007; Wang et al., 2004). Although H3K9me3 and H3K27me3 are

both silencing mark of gene expression, they are mutually exclusive in the same gene loci(Nestorov et al., 2013).

1-4 Early termination of transcription

The level of intact and mature mRNA is interfered by processes including aberrant splicing and premature termination of transcription. Premature termination of transcription is common and usually happens at the proximal sites of TSS. The premature transcripts are either quickly degraded or polyadenylated to become stable noncoding RNAs. The best studied premature termination of transcription pathway is the Nrd1–Nab3–Sen1(NNS) complex in S. cerevisiae, which is not conserved in metazoans(Arigo et al., 2006; Thiebaut et al., 2006; Vasiljeva & Buratowski, 2006). Recently, progresses on identifying similar pathways in metazoans are made and show that complexes interacting with Ars2 are playing important roles in determining the fate of nascent transcripts.

Ars2 mediated early termination

Arsenic resistance protein 2 (Ars2) is a conserved protein involved in RNA metabolism through its productive and destructive roles in the nucleus. Ars2 binds to the nuclear cap binding complex (CBC) and facilitates cap-proximal termination in transcription of many RNA species(Gruber et al., 2009). Its targets include 3′extended snRNAs and snoRNAs, promoter-upstream transcripts (PROMPTs), enhancer RNAs (eRNAs), and also within the first intron of protein coding genes(Gruber et al., 2012; O'Sullivan et al., 2015). Ars2 acts as a scaffold, interacts with diverse complexes and determines the fate of the RNA.

For example, ARS2 interacts with DROSHA and Dicer-2, impacting siRNA biogenesis in *Drosophila*(O'Sullivan et al., 2015; Sabin et al., 2009). Ars2 also interacts with phosphorylated adapter for RNA export protein (PHAX) for snRNA nuclear export, and ALYREF for mRNA export(Fan et al., 2017; Gromadzka et al., 2016; M. Hallais et al., 2013; Ohno et al., 2000). On the other hand, Ars2 interacts with exosome cofactors including the poly(A) tail exosome targeting (PAXT) and Nuclear Exosome Targeting (NEXT) complex via the ZnF protein ZC3H18 for RNA degradation(Andersen et al., 2013; S. Falk et al., 2016; C. Iasillo et al., 2017; N. Meola et al., 2016). The mutually exclusive interaction of Ars2 with PHAX/ALYREF or NEXT/PAXT determines whether a nuclear RNA is degraded or exported(S. Giacometti et al., 2017; Silla et al., 2018).

NEXT complex delivers premature terminated transcripts to nuclear exosome for degradation. NEXT complex is composed of three proteins: the RNA helicase Mtr4, the zinc-knuckle protein ZCCHC8 and the RNA-binding protein RBM7. ZCCHC8 is the scaffold protein that mediates the interaction of RBM7 and Mtr4 in the complex(Nicola Meola et al., 2016). RBM7 is RNA binding protein with polyuridine-specific RNA recognition motif(Dominika Hrossova et al., 2015). Mtr4 is the RNA helicase that unwinds the RNA substrate presents it to the exosome(Weick et al., 2018).

CHAPTER TWO
Codon usage impacts translation efficiency via eRF1 mediated premature termination at rare codons

Introduction

Codon usage and gene expression levels are tightly correlated. Codon usage impacts functional protein level through processes in transcription, mRNA decay, translation and co-translational protein folding. For decades, it has been proposed that codon usage is a major determinant of translation efficiency and impacts functional protein level. Indeed, genes containing more optimal codons are usually expressed at higher levels than nonoptimal codons(Gingold & Pilpel, 2011; T. Tuller et al., 2010). It is a common practice to optimize genes based on host codon usage in heterologous gene expression in order to increase yield(Burgess-Brown et al., 2008; Gustafsson et al., 2004). However, the evidence for a substantial causality of codon content and translation efficiency is often contradictory.

Codon usage bias is directly correlated with the concentration of cognate tRNA(Bulmer, 1987; Duret, 2000; Moriyama & Powell, 1997). We previously used *Neurospora*, *Drosophila* and mammalian reporter-based cell-free translation systems combined with genome-wide ribosome profiling to show that rare codons are decoded at slower rate than optimal codons(Fu et al., 2018; Yu et al., 2015; Zhao et al., 2017). Codon usage bias may impact final protein production in two ways. Firstly, due to the slow elongation rate caused by rare codons, optimal genes finish translation faster and thus accumulate more protein. Alternatively, rare codons cause ribosome stalling due to the longer

waiting time for A-site tRNA, which lead to abolish of translation and decrease of full-length protein products.

Once the ribosome is stalled on the mRNA, it needs to be dissociated and released from the mRNA. Two proteins, eRF1 and Dom34, are both mimics of tRNA and capable to enter A site and mediate the dissociation of ribosome on mRNA in eukaryotes. The eRF1/eRF3 and Dom34/Hbs1 complexes share many similarities with the tRNA selection pathway(Becker et al., 2011). The stringent stop codon recognition by eRF1 is achieved by combined interaction of multiple domains in eRF1, 18s rRNA, and ribosomal proteins. Extensive mutational and genetic analyses identified motifs GTS31-33, E55, TASNIKS58-64, and YxCxxxF125-130 (numbering for human eRF1) motifs located at the apex of the N-domain in eRF1 are essential for stop codon recognition (Brown et al., 2015; Frolova et al., 2002; Kryuchkova et al., 2013). Early studies also proposed that eRF1 can release ribosomes at sense codons with a much lower efficiency (Chavatte et al., 2003; Svidritskiy et al., 2018; Vivanco-Dominguez et al., 2012). More recently, multiple studies proposed an alternative role of eRF1 on rescuing stalled ribosomes at sense codons(Chiabudini et al., 2014; Shcherbik et al., 2016; Wada & Ito, 2019).

Insights into substrate selection by Dom34 mostly came from studies showing a strict dependence on the mRNA 3′ length of the stall site. The less 3’ mRNA present, the more potent the releasing mediated by Dom34 (Guydosh & Green, 2014; Pisareva et al., 2011; Shoemaker et al., 2010). Dom34 is mainly involved in no-go decay and non-stop decay pathways, wherein ribosomes are stalled at the end of truncated mRNAs, or endonucleolytically cleaved faulty mRNAs (Doma & Parker, 2006). Recent structure of

Dom34 also proposed an empty mRNA tunnel model wherein Dom34 enters ribosomes stalled at 3' end (Hilal et al., 2016). Although eRF1 is proposed to

In this chapter, I used in vitro cell-free translation assay and in vivo reporters to study the underlying mechanism of codon usage determining translation efficiency. I show that rare codon cause ribosome stalling and premature termination of translation, which reduced the production of full-length protein. This process is conserved from *Neurospora* to *Drosophila*. I further reveal that the premature termination process is mediated by canonical termination factor eRF1. Together, these results demonstrate how codon usage impacts translation efficiency and the role of eRF1 in mediating codon-dependent premature termination of translation.

Results

2.3.1 Nonoptimal codons cause ribosome stalling and premature termination of translation in cell-free assay

Previously we showed that codon decoding time is strongly correlated with codon usage(Yu et al., 2015). We reasoned that the difference between optimal and nonoptimal codon decoding time may affect the elongation process and impact translation efficiency. In order to determine the effect of codon usage on translation elongation process, I examined the profile of nascent peptides using *Neurospora* cell-free translation assay. Endogenous mRNAs were depleted from the whole-cell lysate by MNase digestion. The nascent peptides generated from exogenously supplied mRNAs were labeled by [^{35}S]-methionine. Two

versions of firefly luciferase (*Luc*) mRNA using either wild-type codon (WT) and optimized codon (OPT) were added into the reaction. In OPT *Luc*, all except the first ten codons of the luciferase open reading frame were changed to the most preferred codons in *Neurospora crassa*. After 12 min of translation, reaction was halted, and translation products were compared. As shown in Figure 2-1A, the translation products of the WT *Luc* and OPT *Luc* mRNAs exhibited different profiles that products generated from WT *Luc* showed huge amount of low molecular weight products.

The truncated translation products can be generated from ribosome stalling or premature termination during translation. Ribosome stalling produces peptidyl-tRNA, whereas premature termination produces truncated nascent peptides. In order to characterize the processes that generated these translation products, I used RNase A treatment to distinguish peptidyl-tRNA and nascent peptides. On neutral pH SDS-PAGE (NuPAGE) that prevents alkaline hydrolysis of peptidyl-tRNA, the peptidyl-tRNA species are around 20 kDa larger than their corresponding nascent peptides. RNase A cleaves the tRNA in peptidyl-tRNA species and leads to mobility shifts on NuPAGE gel. As shown in Figure 2-1B, Translation products generated from OPT *Luc* mRNA showed obvious mobility shifts of many large molecular weight bands to lower positions after RNase A treatment, indicating they are mostly peptidyl-tRNAs. However, translation products generated from WT *Luc* mRNA showed only minor mobility shift, indicating they are premature terminated translation products.

In order to further characterize these truncated translation products, I separated peptidyl-tRNAs and nascent peptides using sucrose cushion and centrifugation. Peptidyl-

tRNAs associated with ribosomes would be pelleted after centrifugation, whereas nascent peptides would be enriched in supernatant fraction. By comparing OPT and WT fractions, we found there are more products in WT supernatant fraction, indicating more premature termination events during translation of WT *Luc* mRNA (Figure 2-1C). In addition, there are more products in WT pellet fraction, indicating ribosomes frequently stall during translation of WT *Luc* mRNA. Taken together, these results indicate that nonoptimal codons cause ribosome stalling and premature termination of translation in cell-free translation system.

2.3.2 Poor codon usage causes reduced full-length protein level in Neurospora crassa

In order to examine the codon usage effect on translation efficiency in vivo, I expressed eGFP-2A-Luc reporters in *Neurospora* (Figure 2-2A). To avoid the potential impacts of codon usage manipulations on transcription, mRNA decay and translation initiation, and to accurately measure the effect on translation elongation, the P2A peptides was inserted between eGFP and luciferase. The P2A sequence triggers release of nascent peptide during translation and allows the ribosome to continue translating downstream open reading frame. Thus, eGFP level can be used to normalize changes on mRNA level and translation initiation. Three versions of luciferase with different codon usage were used: full-length optimized codons (2A-OPT), optimized codons with the exception of wild-type codons in the middle of the protein (2A-M-WT), and optimized codons in the N-terminal region and wild-type codons in the C-terminal region of the protein (2A-N-OPT). The expression cassettes were inserted at the his-3 locus of $dim5^{KO}$ strain, which reduced the codon usage impact on transcription level(Zhou et al., 2016). As shown in Figure 2-2B and

C, Western blot of flag-tag that simultaneously detects protein levels of eGFP and Luc showed that 2A-M-WT Luc/eGFP ratio is lower that 2A-OPT, and the ratio is further reduced in 2A-N-OPT. In consistent with previous studies, the increase of nonoptimal codon in 2A-M-WT and 2A-N-OPT resulted in decreased levels of mRNA (Figure 2-2D)(Presnyak et al., 2015; Radhakrishnan et al., 2016; Zhou et al., 2016). Collectively, these results showed that translation efficiency is dependent on codon usage of the mRNA, and nonoptimal codons reduce translation efficiency in *Neurospora crassa*.

2.3.3 Poor codon usage causes reduced full-length protein level in Drosophila cell culture

Codon usage bias and its impact on gene expression is universal. In order to determine whether the codon usage effect on translation efficiency is conserved among species, I expressed eGFP-2A-Luc reporters in *Drosophila* S2 cells (Figure 2-3A). The T2A peptides, which are more efficient than P2A in *Drosophila* cells, was inserted between eGFP and luciferase. Two versions of luciferase with different codon usage were used: full-length optimized codons (2A-OPT) and wild type codons (2A-WT). The reporter constructs were transfected into S2 cells, and protein levels of eGFP and luciferase were simultaneously measured by western blot detecting Myc-tag. WT Luc showed dramatically decreased protein level comparing to OPT, whereas the eGFP levels were similar between two constructs (Figure 2-3B). The Luc/eGFP ratio further reflected that poor codon usage of WT Luc decreased translation efficiency (Figure 2-3C). Consistent with the similar levels of eGFP protein between the two constructs, RNA levels measured by qRT-PCR showed the OPT and

WT mRNA levels are similar (Figure2-3D). These results showed that the impact of codon usage on translation efficiency is conserved in *Drosophila melanogaster*.

2.3.4 Nonoptimal codons cause ribosome stalling and premature termination in Drosophila cell culture

To examine the role of nonoptimal codons on ribosome stalling and premature termination, I expressed two versions of luciferase in *Drosophila* S2 cells: full-length optimized codons (OPT) and N-terminal and middle region deoptimized codons (NDe) (Figure 2-4A). Translation products were analyzed using sucrose cushion and centrifugation to separate peptidyl-tRNAs and nascent peptides. As expected, full-length protein level of NDe Luc was much lower than OPT Luc (Figure 2-4B). Moreover, low molecular weight products corresponding to the nonoptimal codon region in NDe Luc were observed in NDe pellet and supernatant fractions. These results demonstrate that nonoptimal codons cause ribosome stalling and premature termination of translation in *Drosophila* cells.

2.3.5 eRF1 mediates the premature termination process

In eukaryotes, ribosomes are dissociated from mRNA either by termination on stop codons via recognition of eRF1, or by rescue pathway on coding or noncoding regions via Dom34. Stalled ribosomes at sense codons are usually recognized by Dom34 in the no-go decay, non-stop decay and ribosome quality control pathways. However, recent studies also proposed a role of the canonical termination factor eRF1 in mediating premature termination within open reading frame. In order to investigate the mechanism of premature termination at

nonoptimal codons, I expressed and purified eRF1 and Dom34 recombinant proteins using E. coli and added them into cell-free lysate expressing WT *Luc* mRNA. To better reflect the translation elongation process, harringtonine was added to inhibit translation initiation after 6 min of translation, and the reaction was halted after an additional 9 min. As shown in Figure 2-5A, addition of eRF1 but not Dom34 into the reaction increased the portion of low-molecular-weight translation products and decreased full-length luciferase. RNase A or puromycin treatment showed that the truncated protein products are majorly nascent peptides generated by premature termination. The same result is plotted using ImageJ to show the band intensity difference between different groups (Figure 2-5A, right panel). In addition, cell-free lysate made from Dom34 knockout (*dom34*^{KO}) strain showed similar translation profile of OPT or WT *Luc* mRNAs comparing to lysate made from wild type (WT) strain, indicating Dom34 has little impact on producing premature termination translation products (Figure 2-5B).

To further confirm that eRF1 mediates premature termination at nonoptimal codons, I compare the effect of increased eRF1 level on translation profile of WT and OPT *Luc* mRNAs. Adding recombinant eRF1 protein into translation reactions only resulted in increased premature termination products on WT *Luc* mRNA, but not OPT mRNA (Figure 2-5C). These results demonstrated that eRF1 mediates premature termination preferentially at nonoptimal codons.

2.3.6 Mutant eRF1 proteins show reduced effect on premature termination

eRF1 is proposed to be highly specific for stop codon recognition(Song et al., 2000). In order to further study the role of eRF1 in premature termination at nonoptimal codons, I expressed and purified two eRF1 mutant proteins. To test whether the stop codon recognition domain is also essential for premature termination at nonoptimal codons, the critical amino acids carrying stop codon recognition function (E55 and NIK61-63) were mutated to alanine. In addition, catalytically dead eRF1 mutant was generated where the residues (GGQ) required for hydrolysis of peptidyl-tRNA and nascent peptide release were mutated to AAQ. As shown in Figure 2-6, addition of wild type eRF1 recombinant proteins into translation reaction of WT *Luc* mRNA showed dosage dependent increase of premature termination products and decrease of full-length protein. However, the effect is dramatically decreased when adding E+NIK mutant eRF1, and completely abolished when adding GGQ mutant eRF1. Consistent with its role in termination at stop codons, adding eRF1 mutants caused aberrant termination at stop codon, where full-length peptidyl-tRNA level increased at the cost of decreasing full-length protein (Figure 2-6A and C). Overall, these results demonstrated that premature termination at nonoptimal codons requires the GGQ catalytical domain and is partially dependent on the stop codon recognition domain of eRF1.

2.3.7 eRF1 knockdown strain show reduced premature termination at nonoptimal codons

In order to further confirm the role of eRF1 in premature termination at nonoptimal codons, I generated eRF1 knockdown strain using inducible RNAi. Hairpin RNA that silences eRF1 is expressed by switching the culture to quinic acid containing culture. Lysate from the WT and eRF1 knockdown strains were used in cell-free translation assay to

compare the translation products of WT *Luc* mRNA under different eRF1 concentration. As shown in Figure 2-7A, eRF1 knockdown resulted in dramatic decrease of premature termination products. Moreover, supplying WT eRF1 recombinant protein into the lysate from eRF1 knockdown strain caused increase of premature termination products and decrease of full-length protein (Figure 2-7B). These results further demonstrate the role of eRF1 in mediating premature termination of translation at nonoptimal codons.

2.3.8 eRF1 knockdown strain show increased translation efficiency of nonoptimal codons

In order to confirm the role of eRF1 in vivo, I expressed two versions of luciferase in eRF1 knockdown strain with *dim5* knockout background, codon fully optimized (OPT) and N and C terminal region codon optimized (M-WT) (Figure 2-8A, top penal). The presence of quinic acid (Adalat et al.) increased M-WT protein level but did not change OPT level (Figure 2-8A and B). Furthermore, the addition of QA did not affect mRNA levels of M-WT and OPT (Figure2-8C). Collectively, these results demonstrate that eRF1 affects translation efficiency in codon-dependent manner and mediates premature termination only at nonoptimal codons.

Discussion

The correlation between codon usage and translation efficiency has been observed for decades. However, much remains unknown about the molecular mechanisms connecting synonymous codon usage to efficient protein biogenesis. Here I used *Neurospora crassa* and *Drosophila* assays to demonstrate that nonoptimal codons cause ribosome stalling and

decrease translation efficiency by premature termination of translation. Furthermore, I showed that eRF1 is the dominant factor that mediates premature termination at nonoptimal codons.

Translation efficiency is regulated to achieve the optimal synthesis level of proteins. Early studies focused on the diverse regulation on initiation, since initiation is thought to be the rate-limiting step and primary determinant of translation efficiency(Jackson et al., 2010; Sonenberg & Hinnebusch, 2009). When ribosome abundance is limited during fast cell growth stage, for most actively translated mRNAs, translation efficiency is primarily determined by ribosome loading at 5' ends(Chu et al., 2011; Shah et al., 2013; Weinberg et al., 2016). However, with emerging evidences, the field has come to appreciate that codon-dependent regulation on elongation also determines translation efficiency(Cambray et al., 2018; Fu et al., 2018; Subramaniam et al., 2014; Varenne et al., 1984). Here we demonstrate that nonoptimal codons cause premature termination of translation, which directly decreases translation efficiency. Codon usage regulating elongation speed is a critical factor impacting translation efficiency.

Stalled ribosome is usually dissociated from mRNA via ribosome quality control pathway, where Dom34 enters the A site of stalled ribosome at 3' end of truncated or faulty mRNAs(Guydosh & Green, 2014). Here I showed that eRF1 releases stalling ribosomes at nonoptimal codons in *Neurospora crassa*. Supporting this model, multiple studies showed that eRF1 also dissociates ribosomes from sense codons(Chiabudini et al., 2014; Wada & Ito, 2019). Thus, it is likely that different factors are recruited to rescue stalled ribosomes caused by different situations.

Materials and methods

Plasmid construction

Vectors for expression of *Luc* mRNA in *Neurospora crassa* and *Drosophila melanogaster* with different codon-optimized regions were from previous studies(Yu et al., 2015; Zhao et al., 2017). The codons of firefly luciferase were optimized or deoptimized based on the *Neurospora crassa* codon usage frequency table from (https://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=5141). The *Drosophila melanogaster* codon usage frequency table was obtained from (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=7227).

Expect for the first 10 amino acids, the full-length OPT-*Luc* and DeOPT-*Luc* constructs contain the most or the least preferred codons, respectively. For constructs used in *Neurospora* and *Drosophila* cell-free assays, the luciferase sequence was cloned into pJI204 vector containing a T7 promoter and 30-nt poly-A sequence at BamHI/XbaI sites. Other Luc-based constructs were created by homologous recombination-based cloning (In-Fusion HD cloning kit, Clontech) using the WT-*Luc*, OPT-*Luc*, or DeOPT-*Luc* as parental templates.

The *Neurospora* eGFP-2A-Luc constructs for in vivo study were generated in pMF272 plasmids, which contain the ccg-1 promoter and his-3 targeting sequence by insertion at the AscI/XbaI. *Drosophila* S2 in vivo constructs were created on the backbone of pAc-STABLE1-neo(González et al., 2011). Different versions of *Luc* sequences were inserted into the pAc vector at NheI/XhoI sites.

Neurospora crassa strains and culture conditions

The *Neurospora* strains used in this study for cell-free lysate preparation were wild-type 74A-OR23-1VA (*A*). *Dom34*^{KO} strain was generated by replacing the dom34 gene with the hygromycin resistance gene by homologous recombination. The dim-5^{KO} (*bd, his-3*) strain was used as the host strain for all his-3 targeting constructs to monitor eGFP-2A-Luc reporter expression in vivo(Zhou et al., 2016). The eRF1-KD strain was generated in the 301-6 (*bd, his-3, a*) background by introducing a construct expressing a eRF1-specific RNA hairpin under the control of the qa-2 promoter(Cheng et al., 2005). To induce eRF1 silencing, *Neurospora* mycelium was cultured in 0.01 M quinic acid with 1× Vogel's, 0.1% glucose, and 0.17% arginine overnight at 28 °C.

In vitro transcription

To prepare the templates for in vitro transcription, the plasmids were linearized by treatment with NheI followed by successive phenol-chloroform extraction and ethanol precipitation. The capped and poly(A) tailed mRNA transcripts were then synthesized using HiScribe T7 Quick High Yield RNA Synthesis Kit (NEB) supplemented with 3'-O-Me-m7G(5')ppp(5')G anti-reverse cap structure analog (NEB) per manufacturer's instructions. The integrity and quantity of the synthesized transcripts were evaluated by denaturing agarose gel electrophoresis. The mRNA concentrations were determined using a Nanodrop spectrophotometer (ThermoFisher Scientific).

Preparation of Neurospora cell-free translation extract and in vitro translation reactions

The preparation of wild-type translation extract was performed following a protocol previously described(Wu et al., 2018). *Neurospora* conidia were harvested and inoculated at a concentration of 1×10^7 conidia/ml in 1× Vogel's with 2% sucrose. A 1-L culture was incubated at 32 °C with shaking at 200 rpm for 6.5 h. Geminating conidia were harvested by vacuum filtration and snap frozen in liquid nitrogen. The mycelial pads were homogenized using a mortar and pestle with gradual addition of 1 volume of Buffer A (30 mM HEPES-KOH, pH 7.6, 100 mM KOAc, 3 mM $Mg(OAc)_2$, 2 mM DTT, 1× protease inhibitor cocktail (A32963, ThermoFisher Scientific)). The lysate was centrifuged for 15 min at 30,000 × g at 4 °C. The supernatant was carefully collected and centrifuged again. Small molecular weight molecules were removed using Zeba Desalt Spin Columns (Pierce). For preparation of extract from eRF1-KD strains, the mycelium mat was cultured in 0.01 M quinic acid with 1× Vogel's, 0.1% glucose, and 0.17% arginine overnight at 28 °C. The lysate was aliquoted, snap frozen in liquid nitrogen, and stored at -80 °C.

A reaction mixture (5 µL of cell lysate with 1 µL of 10× energy mix, 0.06 µL of 10 U creatine phosphate kinase, 0.35 µL of 2 M KOAc, 0.1 µL of 0.1 M $Mg(OAc)_2$, 0.1 µL of 1 mM amino acids mix, 0.1 µL of RNase inhibitor (ThermoFisher Scientific), and 2.4 µL of RNase-free water) was prepared, and 60 ng of mRNA was added to achieve a total volume of 9.5 µL. [^{35}S]Met (0.5 µL/reaction) was added to label all protein products. The reaction was incubated at 26 °C for various time. For RNase A treatment, 1 µL of 10 mg/ml RNase A (ThermoFisher Scientific) was added to reaction and incubated for 15 min at 37 °C. For puromycin treatment, 1 µL of 10 mg/ml puromycin (Sigma) was added, and the reaction was

incubated at 26 °C for 15 min. For harringtonine treatment, 1 μL of 10 mg/ml harringtonine (Sigma) was added after 6 min, and the reaction was incubated at 26 °C for an additional 9 min. To determine eRF1 function, 0.6 μg recombinant protein was added to 20 μL in vitro translation reaction (0.6 μM). The eRF1 concentration used in our assays is similar to previous studies (Eyler et al., 2013; Frolova et al., 1994). The reactions were stopped by placing on ice and were analyzed by electrophoresis on 10% NuPAGE gel (ThermoFisher Scientific).

S2 cell culture and transfection

Drosophila S2 cells were cultured in Schneider's *Drosophila* Medium (Gibco) supplemented with 10% heat-inactivated fetal bovine serum (Gibco) and 1% penicillin-streptomycin (10,000 units penicillin and 10 mg streptomycin/ml, Sigma) at 27 °C. For transfection, cells were plated at 5×10^6 cells per well of a 12-well plate. For expression of reporter genes, 1 μg plasmid was transfected into each well using Lipofectamine 3000 (ThermoFisher Scientific) following the manufacturer's instructions. Cells were harvested after 48 h for further analysis.

Isolation of ribosome-associate nascent peptides

The in vitro translation reactions were terminated by adding cycloheximide to a final concentration of 0.5 mg/ml. For cell culture, S2 cells were treated with cycloheximide (0.01 mg/ml) for 10 min and harvested. S2 cells were lysed in hypotonic buffer and homogenized by a Dounce homogenizer. Translation product or cell lysate was carefully layered on top of

a sucrose cushion (0.5 M sucrose, 25 mM HEPES, pH 7.5, 80 mM KOAc, 1 mM $Mg(OAc)_2$) and then centrifuged in a TL100.3 rotor (Beckman Coulter) at 95,000 rpm for 15 min at 4 °C. The ribosomal pellets are washed once with 25 mM HEPES (pH 7.5), 80 mM KOAc, and 1 mM $Mg(OAc)_2$ and then resuspended in lysis buffer for further analysis.

Protein expression and purification

The plasmid pQE2 (Qiagen) was used for recombinant protein expression. The eRF1 open reading frame was PCR amplified from *Neurospora* cDNA and inserted into pQE2 at the NheI/NcoI site. Site directed mutagenesis by PCR was used to generate mutant eRF1. The constructs were overexpressed in E. coli BL21. The recombinant protein was purified by affinity chromatography on nickel beads (Sigma) according to the manufacturer's protocol. The protein concentration was determined using a Nanodrop spectrophotometer.

Protein analysis

Neurospora tissue was ground in liquid nitrogen with a mortar and pestle. Tissue powder was suspended in ice-cold extraction buffer (50 mM HEPES, pH 7.4, 137 mM NaCl, 10% glycerol, 1 μg/ml pepstatin A and 1 μg/ml leupeptin). S2 cells were lysed in 1× Passive Lysis Buffer (Promega) according to the manufacturer's instruction. For Western blot analyses, equal amounts of total protein were loaded in each lane. After electrophoresis, proteins were transferred onto PVDF membrane and Western blot analysis was performed. Anti-Flag antibody (F3165, Sigma) was used to detect Flag-tagged proteins, anti-luciferase

antibody (L2164, Sigma) was used to detect luciferase, and anti-Myc antibody (M4439, Sigma) was used to detect Myc-tagged proteins.

RNA analysis

RNA was extracted with Trizol (Ambion) in accordance with the manufacturer's protocol. For qRT-PCR, the primer sequences used for amplification of *Luc* in *Neurospora* targeting the 5'UTR and the flag-tag region, or in *Drosophila* targeting the 5'UTR and the myc-tag region. The *Neurospora* gene coding for β-tubulin was used as an internal control. The *Drosophila* gene coding for actin was used as an internal control.

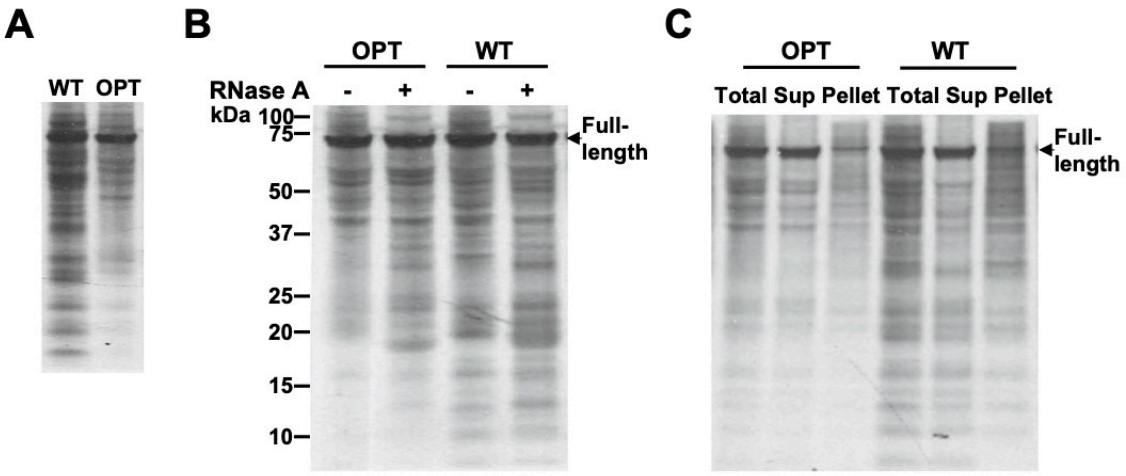

Figure 2-1. Nonoptimal codons cause ribosome stalling and premature termination of translation. (A) [^{35}S]-Met-labeled total translation products analyzed by electrophoresis on 10% NuPAGE gel. Translation of WT and OPT *Luc* mRNAs was carried out for 12 min. (B) [^{35}S]-Met-labeled total translation products of WT and OPT *Luc* mRNA with and without RNase A treatment. After 10 min of translation, samples were treated with RNase A for 15 min at 37 °C. (C) Pellet- and supernatant-associated [^{35}S]-Met-labeled translation products of WT and OPT *Luc* mRNAs. After 10 min of translation, reactions were terminated by addition of cycloheximide (0.5 mg/ml final concentration). Ribosome-associated nascent peptides were separated by sucrose cushion centrifugation. Sup: supernatant.

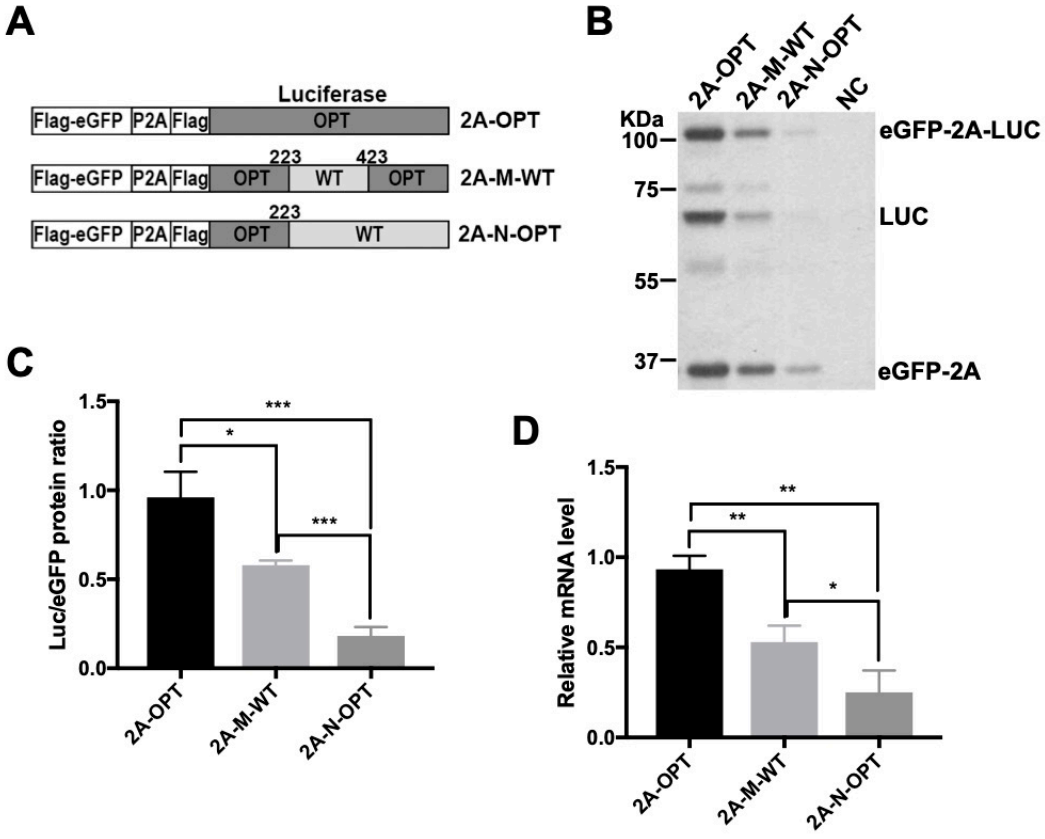

Figure 2-2. Poor codon usage causes reduced full-length protein level in *Neurospora*. (A) Diagrams of eGFP-2A-Luc reporter constructs. (B) Representative western blot analysis of the eGFP-2A-Luc translation products. *Neurospora* strains carrying the indicated reporter construct were grown, and proteins were extracted for analysis. The positions of the eGFP-2A, Luc, and full-length eGFP-2A-Luc proteins are labeled. NC: negative control, which was expression of empty vector. (C) Quantification of western blot analysis shown in panel (B) and two additional independent experiments. The ratio of luciferase to eGFP is plotted. Data

are means ± SD. *P < 0.05. ***P < 0.001. (D) Relative mRNA levels from eGFP-2A reporter constructs measured by RT-qPCR. Data are means ± SD. *P < 0.05. **P < 0.01.

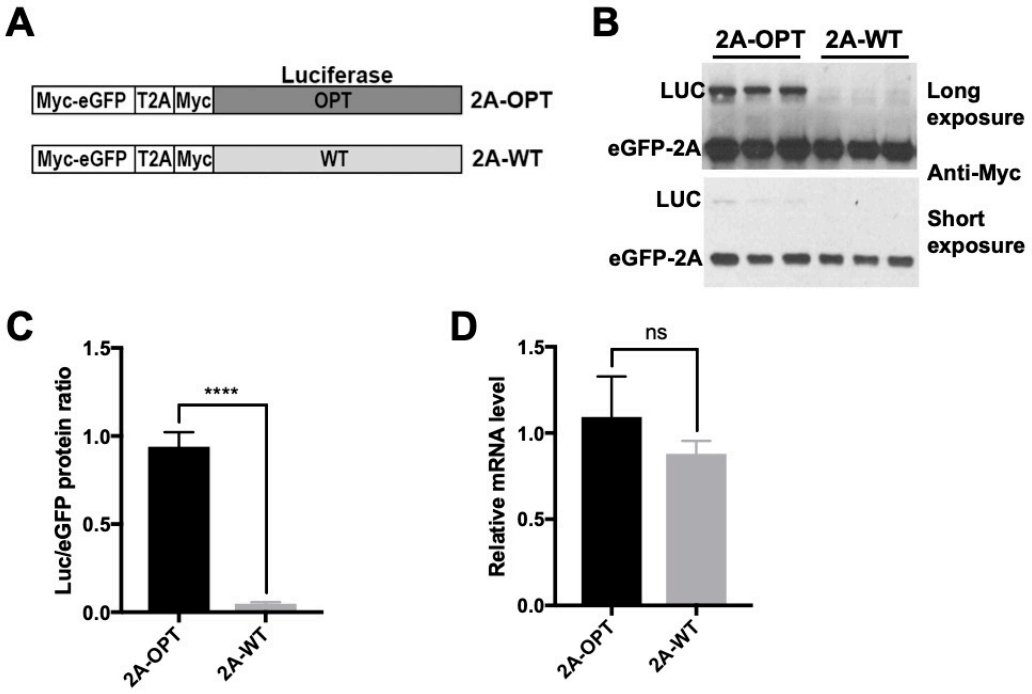

Figure 2-3. Poor codon usage causes reduced full-length protein level in *Drosophila* cell culture. (A) Diagrams showing the eGFP-2A-Luc reporter constructs used in S2 cell experiments. (B) Representative western blot analysis of products from eGFP-2A-Luc reporter constructs detected using c-Myc antibody. Two different exposures are shown. (C) Quantification of Luc/eGFP protein ratio from western blot analysis in panel (B) and two additional independent experiments. Data are means ± SD. ****P < 0.0001. (D) Relative mRNA levels from eGFP-2A-OPT and eGFP-2A-WT reporter constructs measured by RT-qPCR. Data are means ± SD. ns: not significant.

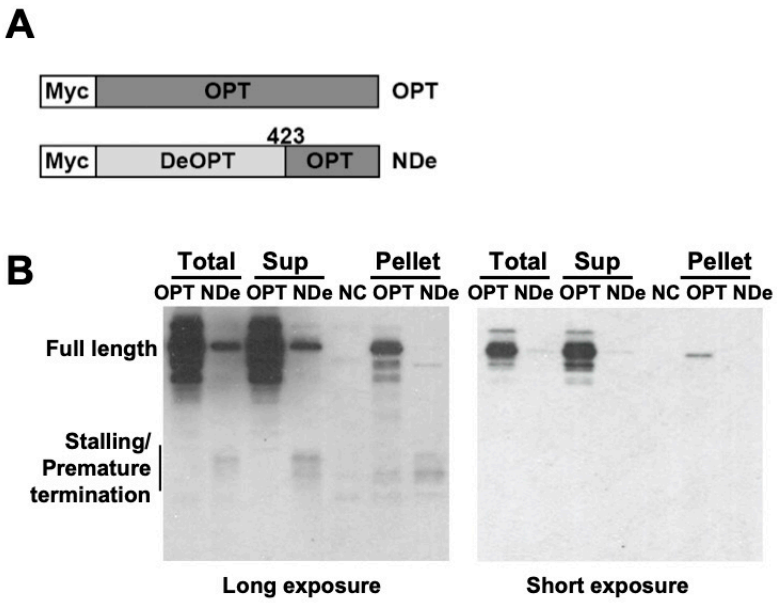

Figure 2-4. Nonoptimal codons cause ribosome stalling and premature termination. (A) Diagrams showing the *Luc* reporter constructs that were transfected individually into S2 cells. (B) Western blot analysis of total cell extracts and isolated ribosome-associated nascent peptides using c-Myc antibody; two exposures are shown. Sup: supernatant.

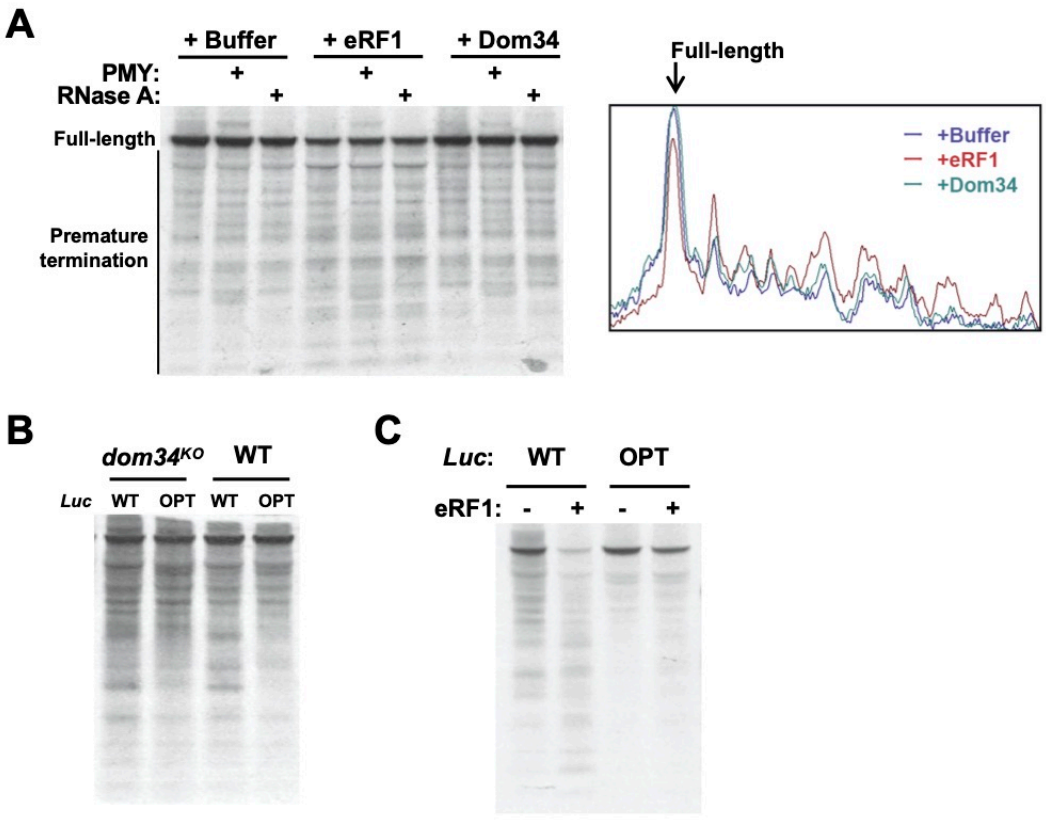

Figure 2-5. eRF1 mediates the premature termination process. (A) Left: [^{35}S]-Met-labeled total translation products of the WT *Luc* mRNA with or without the addition of recombinant eRF1 or Dom34. RNase A (1 µg/µL) or puromycin (PMY, 1 µg/µL) was added after 15 min of translation. Right panel: ImageJ analysis of the translation products in the western blot lanes 1, 4, and 7. (B) [^{35}S]-Met-labeled total translation products of WT or OPT *Luc* mRNA from experiments in cell-free translation extracts prepared from the wild-type or dom34 knockout strain (*dom34^{ko}*). (C) [^{35}S]-Met-labeled total translation products of the WT or OPT *Luc* mRNA with or without the addition of recombinant eRF1 (0.6 µg eRF1 protein in a 20-µL reaction).

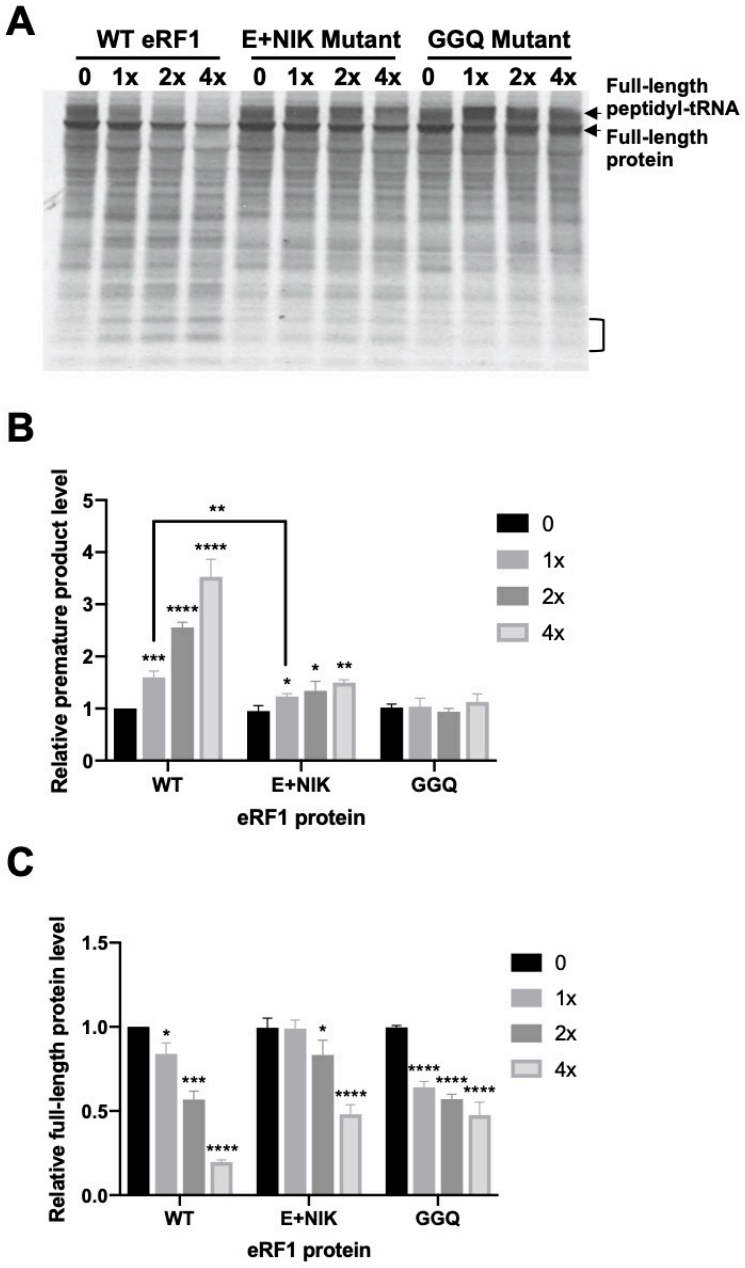

Figure 2-6 Mutant eRF1 proteins showed reduced effect on premature termination. (A) [^{35}S]-Met-labeled total translation products of the WT *Luc* mRNA with the addition of different

concentrations of recombinant wild-type or mutant eRF1 protein (1x is 0.6 μg/20 μL reaction). (B) Quantification of premature termination based on bands with the region labeled by the bracket in panel (A) from three independent experiments. Data are means ± SD. *P < 0.05. **P < 0.01. ***P < 0.001. **** P < 0.0001. (C) Quantification of full-length luciferase levels from independent experiments. Data are means ± SD. *P < 0.05. ***P < 0.001. **** P < 0.0001.

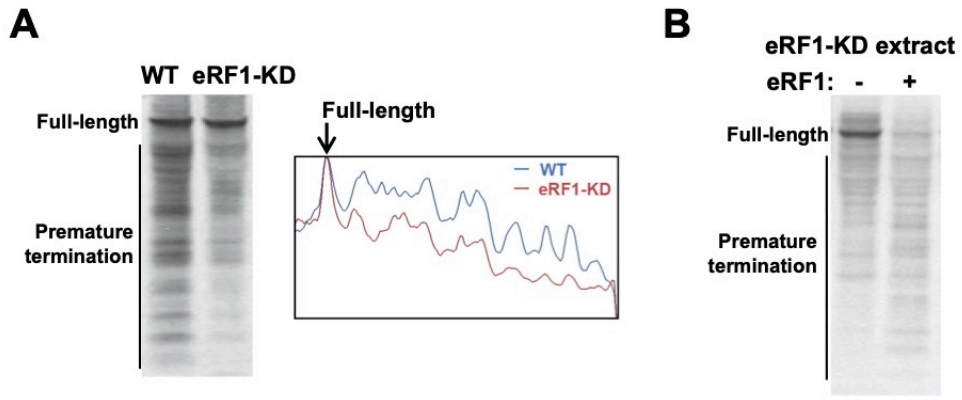

Figure 2-7. eRF1 knockdown strain showed reduced premature termination at nonoptimal codons. (A) Left panel: [^{35}S]-Met-labeled total translation products of WT *Luc* mRNA from experiments in cell-free translation extracts prepared from the wild-type or eRF1-depleted strain (eRF1-KD). Right panel: ImageJ analysis of the translation products. (B) [^{35}S]-Met-labeled total translation products of the WT *Luc* mRNA with or without the addition of recombinant eRF1 in cell-free translation extracts prepared from the eRF1-depleted strain.

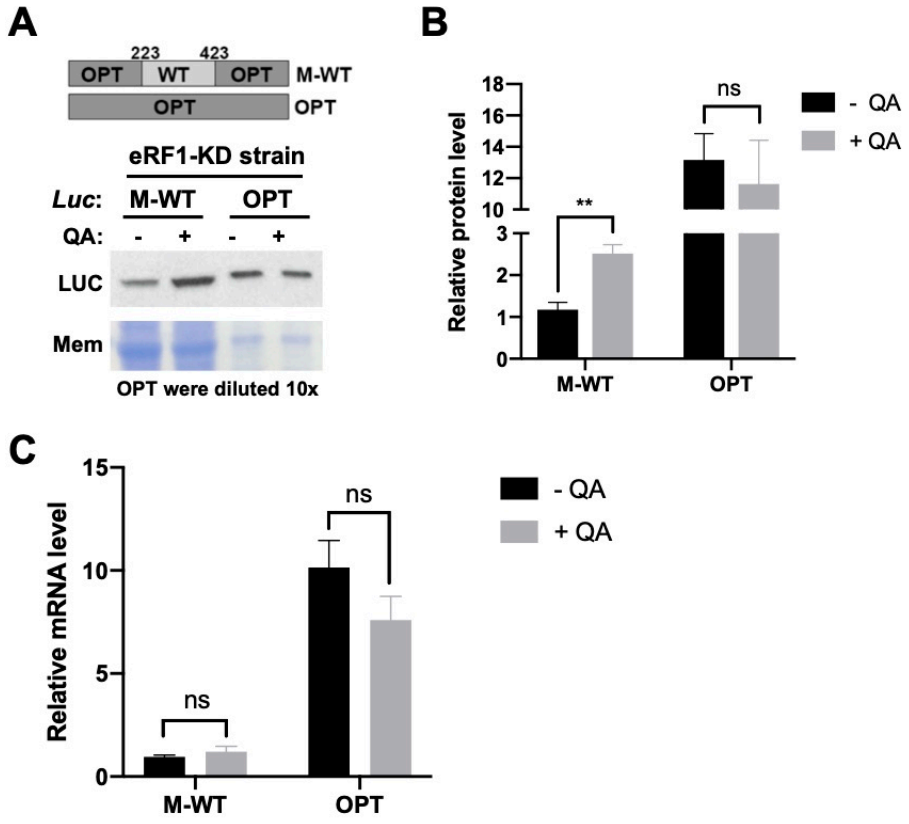

Figure 2-8. eRF1 knockdown strain showed increased translation efficiency of nonoptimal codons. (A) Top panel: Diagrams showing two luciferase-expressing constructs used for *Neurospora* transformation into the eRF1-KD strain. Bottom panel: Representative western blot analysis using luciferase antibody showing the levels of luciferase protein in the strains. cultureed with or without 1mM QA. (B) Quantification of protein levels from western blot analysis in panel (A) and two additional independent experiments. Data are means ± SD. **P < 0.01. ns: not significant. (D) Relative mRNA levels from M-WT and OPT luciferase reporters measured by RT-qPCR. Data are means ± SD. ns: not significant.

Table 2-1 Primer list

Name	Sequence	Note
Flag-F	ACCCCTCACATCAACCAAAGG	RT-qPCR, luciferase
Flag-R	GCCGCCCTTGTCATCGTCATC	RT-qPCR, luciferase
β-tubulin-F	ATAACTTCGTCTTCGGCCAG	RT-qPCR, *Neurospora*
β-tubulin-R	ACATCGAGAACCTGGTCAAC	RT-qPCR, *Neurospora*
Myc-F	TGATATCATCGATTTAAAGCAATGGAG	RT-qPCR, luciferase
Myc-R	GTCGCCCAAGCTCTCCATTTCATT	RT-qPCR, luciferase
Actin-F	GCACACCCACAAGCTTACACA	RT-qPCR, *Drosophila*
Actin-F	TTGCGCTTTGGGAAATATCTTC	RT-qPCR, *Drosophila*

CHAPTER THREE
codon usage impacts transcription level via epigenetic modification and early termination of transcription

Introduction

Codon usage is a key factor in determining gene expression level through impacting multiple fundamental cellular processes(Bazzini et al., 2016; Fu et al., 2016; Lyu et al., 2020; Mordstein et al., 2020; Presnyak et al., 2015; Xu et al., 2013; Yu et al., 2015; Zhou et al., 2013; Zhou et al., 2016; Zhou et al., 2018). Using reporter assays in *Neurospora*, we previously showed that change of codon usage directly impacts the transcription level of different genes(Zhou et al., 2016; Zhou et al., 2018). This change is achieved by chromatin remodeling and histone modification H3K9me3(Zhou et al., 2016). On the other hand, certain codon combinations are avoided in the genome to prevent cryptic transcription termination(Zhou et al., 2018). Since the concept of codon is applicable only in translation process, the codon usage effect on transcription is likely related to GC content. Similarly, in human cells, codon usage, GC3 and GC content of exons determine mRNA level via transcription(Kudla et al., 2006; Lemaire et al., 2019; Pouyet et al., 2017). However, the underlying mechanism is largely unknown.

Transcription level is mainly determined by enhancer-promoter interaction and core promoter strength(Field & Adelman, 2020; Smale & Kadonaga, 2003). Little is known about the role of sequences downstream of core promoter in regulating transcription level. In

addition, epigenetic regulatory elements such as nucleosome density and histone modifications are critical epigenetic factors that affect transcription levels(Cairns, 2005; Clapier & Cairns, 2009). Moreover, after the productive initiation of RNA polymerase II, multiple events still may impact transcription level during the early stage of transcription elongation(Arigo et al., 2006; M. Hallais et al., 2013).

Arsenic resistance protein 2 (Ars2) is a conserved protein involved in transcription elongation steps(Marie Hallais et al., 2013; Claudia Iasillo et al., 2017; Speth et al., 2018). It interacts with the nuclear cap binding complex (CBC) and facilitates cap-proximal termination in transcription of many RNA species, including 3′extended snRNAs and snoRNAs, promoter-upstream transcripts (PROMPTs), enhancer RNAs (eRNAs), and also within the first intron of protein coding genes(Andersen et al., 2013; D. Hrossova et al., 2015; Lubas et al., 2015; Lubas et al., 2011; N. Meola et al., 2016; Preker et al., 2008). Ars2 acts as a scaffold, interacts with diverse complexes and determines the fate of the RNA(Schulze & Cusack, 2017; Schulze et al., 2018). One of the interacting complexes is Nuclear Exosome Targeting (NEXT) complex, which associates with Ars2 via ZC3H18(Lubas et al., 2011). NEXT complex is composed of three proteins: the RNA helicase Mtr4, the zinc-knuckle (ZnK) protein ZCCHC8 and the RNA-binding protein RBM7(Sebastian Falk et al., 2016; Lubas et al., 2015; Lubas et al., 2011). NEXT complex promotes the degradation of short early terminated RNAs by recruiting nuclear exosome(Fan et al., 2017; Lubas et al., 2011).

In this chapter, I investigated the role of promoter in codon-dependent differential gene expression in *Drosophila melanogaster*. I discovered that the region downstream of

core promoter is critical for promoting the expression of nonoptimal codon genes. This effect is achieved at transcription level, via regulations on nucleosome density and histone 3 lysine 27 acetylation (H3K27ac). Furthermore, the Ars2 and NEXT complex pathway is involved in early termination of transcription at 5' end of nonoptimal codon genes. In addition, the downstream region of core promoter contains element that inhibits the Ars2 mediated early termination. Together, these results show that codon usage impacts transcription through multilayer regulations.

Results

3.2.1 Codon usage effect on mRNA level is promoter-dependent

In order to investigate the role of promoter in differential expression caused by codon usage bias, I used luciferase reporters driven by various promoters and expressed them in *Drosophila* S2 cells (Figure 3-1A). Two versions of luciferase with different codon usage, full-length codon optimized (Opt *Luc*), and full-length codon de-optimized (De *Luc*) were tagged with 5xMyc tag at 5' end. Core promoters of *Actin5C*, *Drosophila* Synthetic Core Promoter (DSCP), *Tubulin*, *Hsp70*, *Myc* and *Period* and their 100-150 bp downstream region were identified based on *Drosophila* promoter database (https://labs.biology.ucsd.edu/Kadonaga/DCPD.htm) and used to drive luciferase expression. The 5x UAS sequence from the UAS/Gal4 system was placed in front of core promoter and used as uniform enhancer. A separate cassette expressing eYFP was used to normalize plasmid copy number during transfection.

The expression level of Opt and De *Luc* mRNAs were measured by northern blot probing the 5' end 5xMyc tag sequence. As shown in Figure 3-1B and C, under promoters from *Act5C*, DSCP and *Tub*, the difference between De and Opt *Luc* mRNAs are more than 3-fold. In contrast, under promoters from *Hsp70*, *Myc* and *Per*, the differences are much smaller, which are less than 2-fold. Surprisingly, under Per promoter, De *Luc* showed higher level than Opt *Luc*.

To rule out any potential effect from the unidentified regulatory elements in Myc-tag and luciferase coding sequences, I changed the coding region to YFP with two versions of codon usage, full-length codon optimized (Opt), and full-length codon de-optimized (De) (Figure 3-1D). Similarly, Opt and De YFP expression is promoter dependent. Under promoters from *Act5C* and DSCP, De YFP level is more than 5-fold lower than Opt, whereas under *Hsp70* promoter, De YFP level is strongly elevated (Figure 3-1E and F). Collectively, these results show that codon usage effect on mRNA level is promoter dependent.

3.3.2 Core promoter has little impacts on mRNA stability

Steady state mRNA level is mainly determined by transcription and mRNA decay rates. Codon usage is known to have a major impact on mRNA stability(Presnyak et al., 2015; Q. Wu et al., 2019). However, whether change of promoter and downstream noncoding sequence alter mRNA stability is unclear. Thus, I examined the mRNA stability of Opt and De *Luc* under different promoters by inhibiting RNA polymerase II with actinomycin D. Consistent with previous studies, De *Luc* is less stable than Opt *Luc* (Figure 3-2A and B). Moreover, the decay rates of Opt and De *Luc* remained unchanged under different promoters.

Thus, this result rules out the effect of different mRNA stability caused by different promoters on mRNA levels of Opt and De *Luc*.

3.3.3 The region downstream of core promoter enhances nonoptimal codon expression

Core promoter determines the transcription strength. The core promoter in *Drosophila* is well defined region that usually is composed of mainly three sequence motifs: TATA box, initiator (Inr) and downstream core promoter element (DPE)(Burke & Kadonaga, 1997; Kutach & Kadonaga, 2000; Smale & Kadonaga, 2003). In order to investigate whether core promoter or its immediate downstream region (Svidritskiy et al.) is responsible for differential Opt and De *Luc* expression, I switched the IDR sequences between different promoters (Figure 3-3A). Based on the DSCP, its IDR was changed to the IDRs of *Hsp70*, *Myc* or *Per*. Surprisingly, IDR from *Hsp70*, *Myc* or *Per* drastically enhanced De *Luc* expression, indicating the IDR contains elements that promote nonoptimal codon expression (Figure 3-3B and C). To further test the effect of IDR, I also changed the IDR of *Hsp70* to IDR of DSCP. Consistently, IDR from DSCP reduced the De *Luc* level, supporting the role of IDR in regulating differential codon-dependent RNA levels (Figure 3-3D and E). These chimeric promoters showed that IDR is the key region that determines RNA level of different codon usage. In addition, these results indicate that IDR of *Hsp70*, *Myc* and *Per* contain regulatory element that enhance nonoptimal codon expression.

3.3.4 A 20 bp element in Hsp70 promoter is sufficient to promote nonoptimal codon expression

In order to further identify the regulatory element that enhances De *Luc* expression, I checked the sequence in the 5'UTR region of *Hsp70*. Based on the chimeric promoter composed of DSCP core promoter and *Hsp70* IDR, every nonoverlapping 20 bp sequence in *Hsp70* IDR was converted to the sequence at the corresponding position of DSCP (Figure 3-4A). As shown in Figure 3-4B and C, replacing 0-20 bp (Mut1), 20-40 bp (Mut2), or 60-80 bp (Mut4) region of *Hsp70* IDR showed little impact on Opt or De *Luc* levels, whereby mutant constructs showed similar expression levels as WT *Hsp70* 5'UTR. In contrast, mutation of the 40-60 bp region of *Hsp70* IDR (Mut3) caused dramatic decrease of De expression. Similarly, converting the 20 bp sequence region in *Hsp70* promoter to the sequence at the corresponding position of DSCP IDR also decreased the De *Luc* expression. This result further ruled out the effect of core promoter and confirmed that the 20 bp sequence is required for enhanced expression of nonoptimal codons. Due to its role in regulating codon usage-dependent gene expression, the 20 bp region of *Hsp70* (GTAAAGTGCAAGTTAAAGTG) was named as the codon usage-dependent element (CDE).

The three main elements in core promoters, TATA box, Inr and DPE are spaced at certain distance(Smale & Kadonaga, 2003). Disruption of the spacing of these elements would result in decrease and abolish of their regulatory function in transcription. In order to see whether the CDE also requires being positioned at certain distance from transcription start site, I replaced every nonoverlapping 20 bp sequence in the IDR of DSCP to CDE (Figure 3-3D). Presence of CDE at 0-80 bp positions downstream DSCP core promoter is sufficient to promote De *Luc* expression, indicating the regulatory effect of CDE is

independent of its distance to TSS (Figure 3-3E and F). Taken together, the 20 bp CDE element from *Hsp70* promoter is responsible for the elevated mRNA level of nonoptimal codon reporter.

3.3.5 Different promoters determine nuclear RNA level of nonoptimal codon reporter

Since the mRNA decay rate did not alter under different promoters, I hypothesized that the effect could be on transcriptional or post-transcriptional processes inside the nucleus. To test this hypothesis, I first measured the nuclear RNA levels of Opt and De *Luc* under *Act5C*, DSCP or *Hsp70* promoters. Nuclear RNA level showed similar profile as total RNA that *Act5C* and DSCP expressed much lower De than Opt *Luc*, whereas *Hsp70* showed enhancement on De *Luc* expression (Figure 3-5A and B). I also noticed that the ratio of De to Opt *Luc* RNA level is higher in the nuclear fraction, possibly due to the short half-life of De *Luc* in the cytoplasm.

In order to further determine whether the effect of CDE on mRNA level is through processes inside nucleus, I next checked the nuclear RNA level of Opt and De *Luc* under promoters with or without CDE. Inserting CDE into DSCP (CDE3) increased De *Luc* in both total and nuclear fractions. On the other hand, mutating the CDE in *Hsp70* decreased De *Luc* in both total and nuclear fractions, indicating CDE impacted transcriptional or post-transcriptional processes inside nucleus to improve De *Luc* expression (Figure 3-5C and D). These results indicate that the impact of promoter on codon-dependent expression is through transcriptional or post-transcriptional processes inside the nucleus.

3.3.6 Promoter regulates transcription level to affect codon-dependent differential expression

In order to test the hypothesis that the effect of promoter is on transcription level, I used nuclear run-on assay to measure the transcription level. Nuclei were isolated and transcription was resumed to label all nascent RNA transcripts with BrUTP. Labeled transcripts were further purified by pull down using anti-BrUTP antibody and measured by qPCR using primers targeting the 5' end of the transcripts. Comparing to Opt *Luc*, De *Luc* under *Act5C* and DSCP promoters showed decreased level of nascent transcripts, whereas under *Hsp70* promoter, De showed similar level to Opt *Luc* (Figure 3-6A). To further examine the role of CDE in transcription level, nascent transcripts produced by promoters with or without CDE were also compared. As expected, inserting CDE into DSCP (CDE3) resulted in increased nascent transcripts level of De *Luc* (Figure 3-6B). On the other hand, replacement of CDE in *Hsp70* promoter caused decrease of nascent transcripts level of De *Luc*.

To further investigate the impact on transcription, I next examined RNA polymerase II levels by chromatin immunoprecipitation assay (ChIP) using Pol II CTD antibody. For this purpose, I generated stable cell lines and replaced eYFP with hygromycin resistant gene (*hph*) for selection. The Pol II level was measured by a set of primers targeting the 5'end of luciferase using qPCR and normalized to the first intron of *Act5C*. Under *Act5C* and DSCP promoters, Pol II level is lower at the 5' end of De *Luc* than Opt *Luc*, whereas under *Hsp70* promoter, Pol II level is similar between Opt and De *Luc* (Figure 3-6C). Overall, these results show that promoter and CDE sequence impact De *Luc* level through transcription.

3.3.7 codon usage and promoter together determine nucleosome density and histone modification

To further investigate the underlying mechanism of promoter regulating codon dependent gene expression, I examined the nucleosome density and histone modifications using ChIP assays. Stable cell lines expressing Opt or De *Luc* under DSCP or *Hsp70* promoters were used to detect the relative levels of histone 3 (H3) and active (H3K27ac) or inhibitory (H3K9me3 and H3K27me3) histone modification markers at the 5'end of luciferase using qPCR. H3 level reflects the nucleosome positioning and density, which is correlated with transcription level. Under DSCP promoter, Opt and De *Luc* showed similar levels of H3, indicating the nucleosome densities are comparable (Figure 3-7A). Surprisingly, under *Hsp70* promoter, De *Luc* has less H3 than Opt *Luc*, suggesting the active transcription is partially contributed by more flexible and open chromatin structure. In addition, supporting the nuclear run-on and Pol II ChIP results, H3K27ac level is lower in De *Luc* than Opt *Luc* under DSCP promoter (Figure 3-7B). In contrast, H3K27ac levels are similar between Opt and De under *Hsp70* promoter. However, the inhibitory histone modifications, H3K9me3 and H3K27me3 showed no difference in Opt *Luc* and De *Luc* under DSCP and *Hsp70* promoter (Figure 3-7E and F).

In order to confirm the effect of H3K27ac on codon dependent differential expression, I treated cells with TSA, a potent inhibitor of class I and II histone deacetylase (HDAC). Treatment with TSA increased De level driven by DSCP promoter, but not *Hsp70* promoter, indicating acetylation is critical in regulating De *Luc* expression (Figure 3-7C and

D). Collectively, these results indicate that the promoter dependent boost on nonoptimal codon genes is through epigenetic regulations.

3.3.8 Nonoptimal codons cause early termination of transcription through Ars2-NEXT pathway

In order to further investigate the underlying mechanism of promoter and codon dependent differential expression, I tested pathways involved in transcriptional and post-transcriptional regulations using RNAi. Ars2 is known to cause early termination of transcription of many RNA species, including early termination within the first intron of protein coding genes(Simone Giacometti et al., 2017; Marie Hallais et al., 2013). Ars2 recruits NEXT complex, which interacts with nuclear exosome and causes degradation of early terminated transcripts(Schulze & Cusack, 2017; Schulze et al., 2018). NEXT complex contains three proteins: Mtr4, ZCCHC8 and RBM7(Lubas et al., 2011). To determine the role of Ars2-NEXT pathway in codon dependent expression, I used dsRNA to knockdown key genes in this pathway. As shown in Figure 3-8, under DSCP promoter, silencing of Ars2 resulted in increase of full length De *Luc*, but no impact on Opt *Luc*. Silencing of Mtr4, ZCCHC8 from NEXT complex downstream of Ars2 resulted in appearance of smear of low molecular weight of RNA species, indicating these are early terminated transcripts that are degraded under normal condition. Silencing of Rrp40 from exosome complex also showed similar low molecular weight of RNA species, further confirm that these short transcripts are degraded by NEXT and exosome complexes. In contrast, under *Hsp70* promoter, knockdown of these factors showed little changes on Opt or De *Luc* expression, indicating *Hsp70*

promoter promoted De *Luc* to avoid the Ars2-NEXT pathway to produce high level of full-length RNA.

3.3.9 CDE prevents early termination of transcription of nonoptimal codon reporter

Based on the RNAi result, I hypothesized that CDE in *Hsp70* enhanced De *Luc* expression by preventing early termination of transcription. To test this hypothesis, I inserted one to three copies of CDE into DSCP IDR and expressed these constructs in S2 cells using transient transfection (Figure 3-9A). Since transient transfection of large amount of plasmid is overexpression of luciferase in *Drosophila* cells, the small fragment of RNA in De luciferase under DSCP promoter is easy to detect using northern blot. This is possibly due to inefficient clean up by nuclear exosome. As expected, increase of CDE copy number in DSCP IDR leads to decrease and disappearance of small RNA species, indicating that CDE inhibited Ars2-NEXT pathway to promote De *Luc* expression (Figure 3-9B).

Discussion

Codon usage has been shown to impact transcription level(Zhou et al., 2016; Zhou et al., 2018). Here I explored the role of promoter in codon-dependent regulation at transcription level. Using reporters expressed under different promoters, I showed that coding sequence and promoter-proximal region together determine the transcription level in *Drosophila* S2 cells. I also identified the regulatory element in *Hsp70* promoter. This regulation is through change on nucleosome density and H3K27ac levels. Moreover, I

discovered that Ars2-NEXT-exosome pathway also plays a role in the expression of nonoptimal codon genes.

The concept of codon usage is used in translation process. Codon usage effect on transcription is possibly through change of GC content, supported by other studies that GC3 and GC content also impact transcription level and mRNA decay in cytoplasm(Kudla et al., 2006; Lemaire et al., 2019; Q. Wu et al., 2019). Since codon usage preference for G/C ending codons and the increase of GC content in metazoans may have coevolved, it is not a surprise to learn that codon usage affect transcription independent of translation.

Transcription level is believed to be mainly determined by enhancer-promoter interaction and core promoter strength(Field & Adelman, 2020; Shlyueva et al., 2014). It was largely unknown whether codon usage of coding sequence would impact transcription level in metazoans. Due to the expansion of intronic noncoding sequence during evolution, the distance between coding sequence and promoter is also increased in metazoans. It is possible that the effect is not limited to coding sequences. Ars2 is known to cause early termination at early elongation stage and its effect is dependent on the distance to promoter(S. Giacometti et al., 2017; Gruber et al., 2009; Marie Hallais et al., 2013). Therefore, the coding sequence at 5' end may have stronger effect than 3' end through Ars2-NEXT-exosome pathway. Ars2 effect is not limited to coding sequence but any sequences near promoter(Gruber et al., 2012; Gruber et al., 2009; Marie Hallais et al., 2013; Sabin et al., 2009). Based on the codon usage and GC preference of the genome, it is possible that low GC content sequence may more likely to lead to Ars2 mediated early termination.

Materials and methods

Codon manipulation and plasmid construction

Opt *Luc* and De *Luc* genes were from previous study (Zhao et al., 2017). The codons of firefly luciferase were optimized or deoptimized based on the *Drosophila melanogaster* codon usage frequency table (https://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=7227). Opt YFP is the commonly used eYFP sequence. De YFP codons were deoptimized based on the *Drosophila* codon usage frequency table.

All plasmids were generated based on the Ac5-STABLE1-neo plasmid. A multiple cloning site was first inserted between Ac5 promoter and SV40 poly(A) signal between KpnI and BamHI sites. eYFP, SV40 terminator, gypsy insulator, 5xUAS and 5xMyc-luciferase were inserted into the multiple cloning sites. Different promoters were amplified by PCR from the genome and inserted between 5xUAS and 5xMyc-luciferase at NotI/XhoI sites. Gal4 expression vector was generated by inserting the Gal4 gene after Ac5 promoter using EcoRI/HindIII sites, followed by another cassette expressing puromycin resistant gene. For YFP expression plasmids, the region from Ac promoter to gypsy insulator were replaced by a linker DNA. 5xMyc-luciferase was replaced by Opt or De YFP. For luciferase reporter plasmids used in generating stable cell lines, upstream eYFP was replaced by hygromycin resistant gene (hph).

S2 Cell culture, transfection and generation of stable cell lines

Drosophila S2 cells were cultured at 25 °C in Schneider's *Drosophila* medium (Thermo Fisher Scientific 21720024), supplemented with 10% (v/v) fetal bovine serum (Thermo Fisher Scientific 10438026) and 1% (v/v) penicillin-streptomycin (Sigma P4333). For transfection, cells were plated at 1x 10^6 cells per well of a 24-well plate. 500 ng total plasmid, in which Gal4 expression plasmid and reporter plasmid were mixed at 1:1 molar ratio, was transfected into each well using Lipofectamine 3000 (Thermo Fisher Scientific L3000001) following the manufacturer's instructions. Cells were harvested after 48 hours for further analysis.

To generate stable cell lines, cells were plated at 2x 10^6 cells per well of a 12-well plate. cells were transfected with Gal4 and luciferase plasmids at 1:1 molar ratio using Lipofectamine 3000 (Thermo Fisher Scientific L3000001) following the manufacturer's instructions. After 48 h, cells were passaged and selected with 10 ng/mL puromycin (Sigma P8833) and 300 ng/mL hygromycin B (Thermo Fisher Scientific 10687010) for a week.

RNAi

Templates of double-stranded RNAs from the DRSC (*Drosophila* RNAi Screening Center) were first PCR amplified from genomic DNA using primers with T7 promoter sequence on both ends. dsRNAs were synthesized using HiScribe T7 High Yield RNA Synthesis Kit (NEB E2040). Knockdown experiments were performed in 24 well plates. 4x10^5 cells were incubated with 5 μg dsRNA in 200 μL serum free medium for 30 min. Then 600 μL medium supplied with 10% FBS were added. Cells were harvested after 3 days for further analysis.

Nuclear run-on

Cells were lysed in lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, and 0.5% NP-40). Nuclei were isolated by centrifuge at 2000x g for 5 min and suspended in 40 µL freezing buffer (50 mM Tris-HCl, pH 7.4, 5 mM $MgCl_2$, 30% glycerol, and 1 mM DTT), and 60 µL transcriptional buffer (20 mM Tris-HCl, pH 8.0, 5 mM $MgCl_2$, 150 mM KCl, 2 mM DTT, 500 µM ATP, 500 µM CTP, 500 µM GTP, 500 µM BrUTP, and 200 U/mL Superase-in). After incubation at 25 °C for 30 min, 500 µL TRIzol (Invitrogen) was added to each reaction to stop transcription. RNA was isolated and resuspended in 100 µL IP buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.05% NP-40, and 1 mM EDTA). Anti-BrUTP antibody (Santa Cruz IIB5) and Protein G beads (Thermo Fisher Scientific 10003D) were incubated with RNA for 2 hours at 4 °C, followed by washing beads with IP buffer and isolation of RNA using TRIzol. Newly transcribed RNA level was measured by quantitative real-time PCR (RT-qPCR). RT-qPCR primers were amplifying the 5' end of Myc-tag sequence to measure luciferase level, and Actin intron for normalization.

Chromatin Immunoprecipitation Assay

Cells were suspended and fixed by 1% formaldehyde in PBS buffer for 5 min at room temperature with gentle agitation. Fixation was quenched by adding glycine to a final concentration of 125 mM and incubation for 5 min at room temperature. Cells were pelleted by centrifugation at 1000 x g and resuspended in sonication buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 0.5% SDS). Chromatin was fragmented to 200 - 500 bp by

Bioruptor (Diagenode), with 7 cycles of 30 s on/ off at high power. Samples were precleared using Protein G beads at 4 °C for 1 hour. 1/20 volume of sample was used as input. The antibodies were incubated with sample overnight at 4 °C: Pol II CTD (Abcam ab26721), H3 (Active Motif 39763), H3K27ac (Active Motif 39133), H3K9me3 (Active Motif 61013), H3K27me3 (Active Motif 61017). Protein G (Thermo Fisher Scientific 10003D) or Protein A beads (Thermo Fisher Scientific 10001D) were incubated with sample for 2 hours at 4 °C and washed by RIPA buffer (10 mM Tris-HCl, pH 8.0, 140 mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% Na Deoxycholate), high salt buffer (10 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% Na Deoxycholate), LiCl buffer (10 mM Tris-HCl, pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% NP-40, 0.5% Na Deoxycholate) and TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). Chromatin was eluted with elution buffer (0.1 M $NaHCO_3$, 1% SDS) and de-crosslinked, and DNA was extracted by phenol. Immunoprecipitated DNA was quantified by real-time quantitative PCR (RT-qPCR). RT-qPCR primers were amplifying the 5' end of Myc-tag sequence to measure luciferase level, Actin intron for normalization of Pol II CTD and H3K27ac, 3L:10228701 for normalization of H3K27me3, and F22 for normalization of H3K9me3.

Northern blotting

RNA was extracted with Trizol (Ambion) in accordance with the manufacturer's protocol. The same amount of RNA was loaded and separated on 1% agarose gel with MOPS buffer. RNA was then transferred onto a nylon membrane in 10 × SSC and was crosslinked to the membrane by UV cross-linking. Membranes were prehybridized for 1 h at 60°C in

hybridization buffer (50 mM Tris, pH 7.5, 2 × Denhardt's, 50% formamide, 1 M NaCl, 0.1% Na pyrophosphate and 0.5 mg/mL denatured salmon sperm DNA). Riboprobes were transcribed in vitro from PCR products by MEGAscript™ T7 Transcription Kit (Thermo Fisher Scientific AMB13345) following the manufacturer's protocol, except UTP is substituted with [^{32}P]-UTP. Hybridizations with riboprobes were performed at 60°C overnight. Membranes were washed three times for 20 min each in washing buffer (0.1× SSC and 0.1% SDS) at 65°C. Membranes were exposed to film at −80°C.

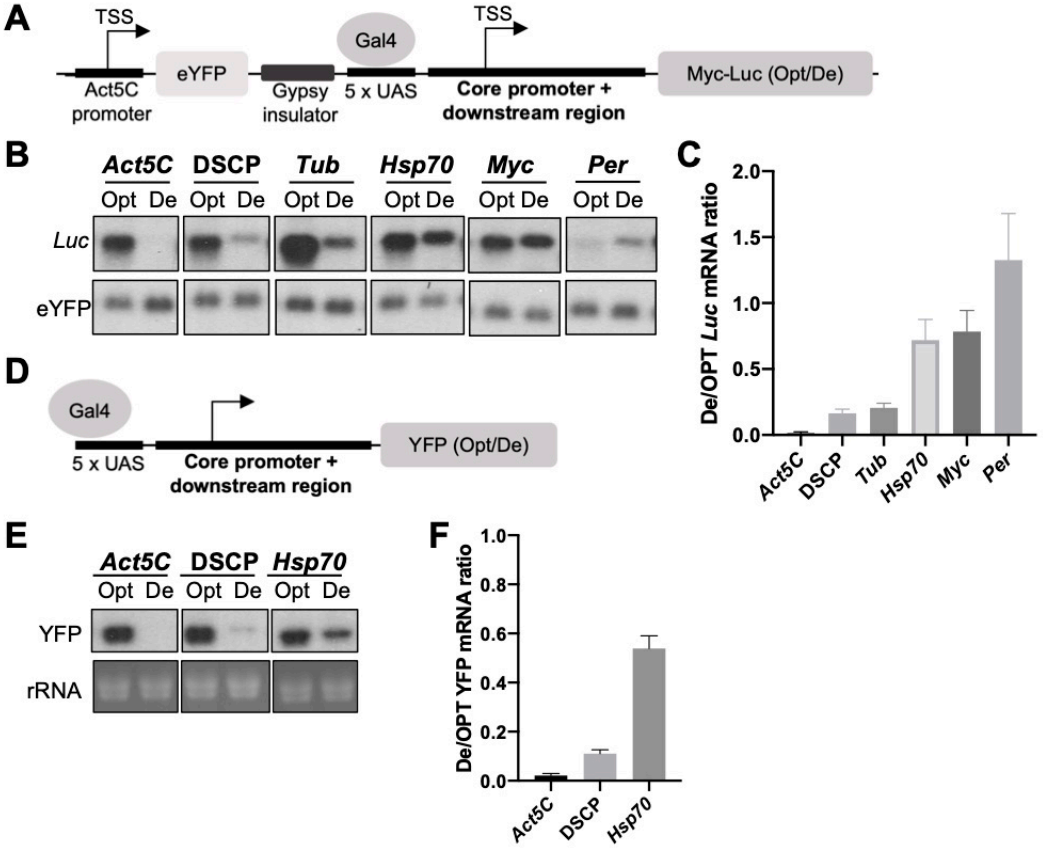

Figure 3-1. Codon usage effect on mRNA level is promoter dependent. (A) Diagram of the luciferase reporter construct. Core promoter and immediate downstream 100 - 150 bp sequence from different promoters are used to drive the expression of the Opt *Luc* or De *Luc*. (B) Northern blot analysis of mRNA levels of *Luc* under different promoters. Probes hybridize the common 5' Myc-tag sequence of *Luc* or the 5' region of eYFP gene. (C) Quantification of De/Opt luciferase mRNA ratios normalized by eYFP level from results in panel B (n=3). (D) Diagram of the *YFP* reporter construct. (E) Northern blot analysis of the mRNA levels of *YFP* under different promoters. The probe hybridizes the common 3' UTR

sequence of *YFP* genes. (F) Quantification of De/Opt *YFP* mRNA ratio from results in panel E. (n=3).

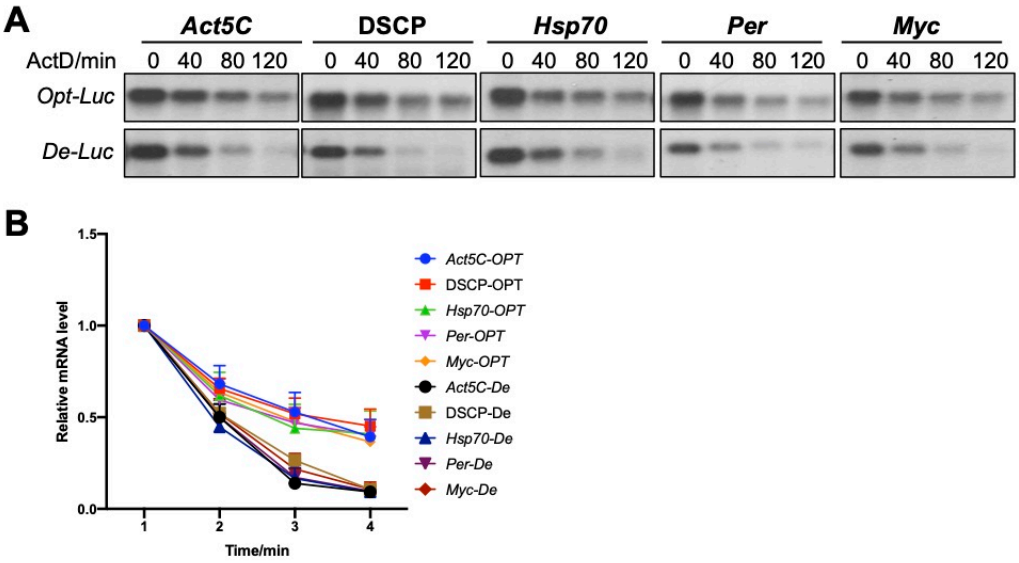

Figure 3-2. Core promoter impacts little on mRNA stability. (A) Northern blot analyses of mRNA decay of the Opt/De *Luc* under different promoters. Actinomycin D was added into cell culture at 10 ng/mL at time 0. Cells were harvested at indicated time points. (B) Quantification of Opt /De *Luc* mRNA decay rate from results in panel G (n=3).

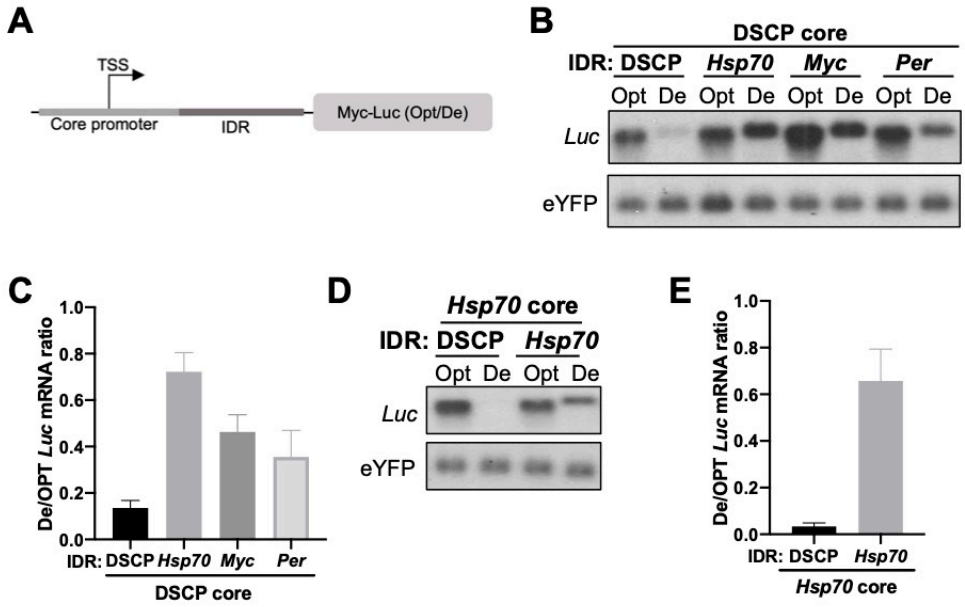

Figure 3-3. The region downstream of core promoter enhances nonoptimal codon expression. (A) Diagram of promoter region. The region labeled as 5' UTR is changed to 5' UTR from other promoters. (B) Northern blot analysis of mRNA levels of luciferase under different chimeric promoters using DSCP core promoter. Probes hybridize 5' Myc-tag sequence for luciferase and 5' region for eYFP. (C) Quantification of De/Opt *Luc* mRNA ratio normalized by eYFP for panel B. (n=3) (D) Northern blot analysis of mRNA levels of luciferase under different chimeric promoters using *Hsp70* core promoter. Probes hybridize 5' Myc-tag sequence for luciferase and 5' region for eYFP. (E) Quantification of De/Opt *Luc* mRNA ratio normalized by eYFP for panel D. (n=3)

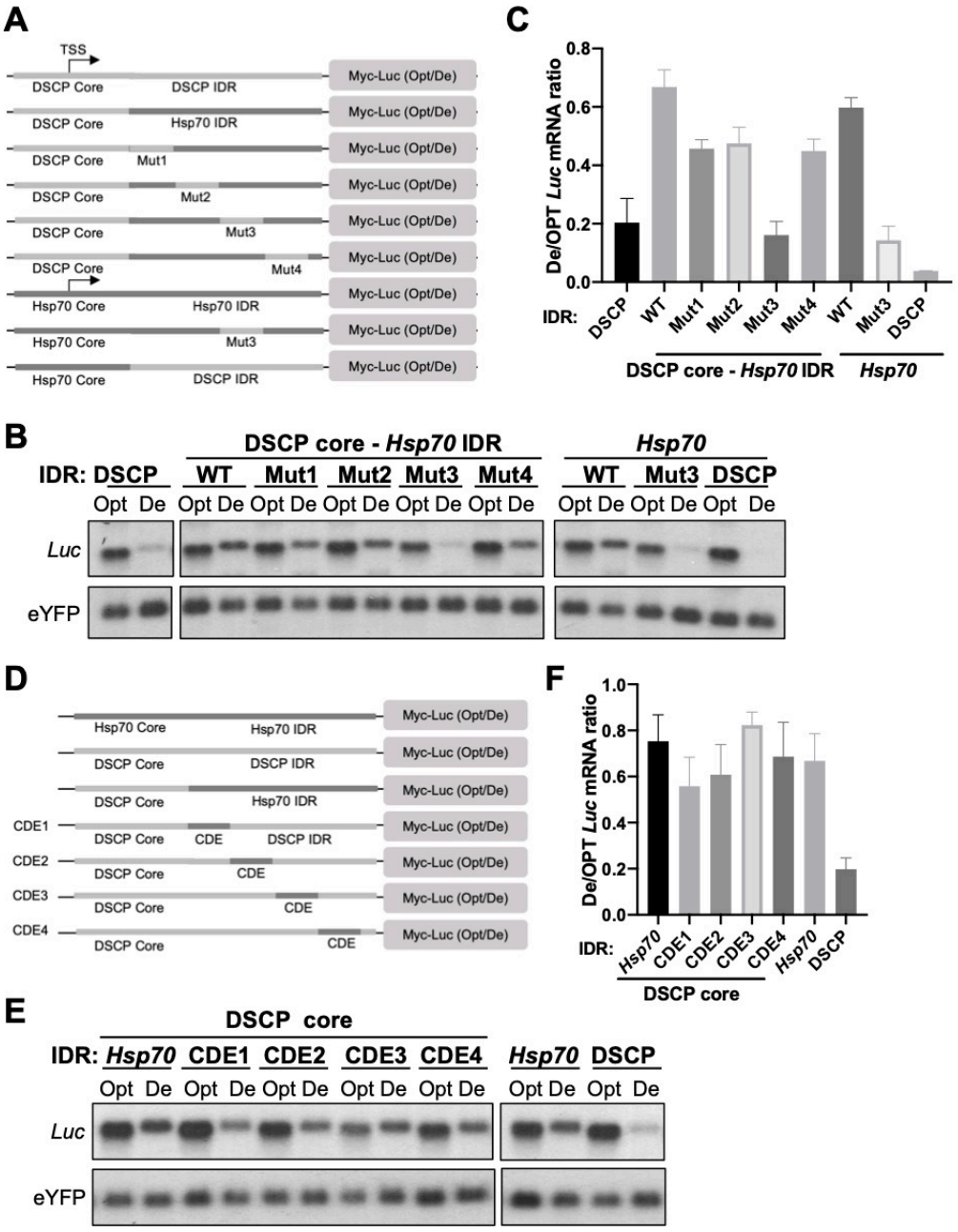

Figure 3-4. A 20 bp element in *Hsp70* IDR is sufficient to promote nonoptimal codon expression. (A) Diagram of the promoter region of the reporter constructs. 20 bp scanning

constructs are used to identify the regulatory element in *Hsp70* 5'UTR. (B) Northern blot analysis of mRNA levels of luciferase under different promoters. Probes hybridize 5' Myc-tag sequence for luciferase and 5' region for eYFP. (C) Quantification of De/Opt mRNA *Luc* ratio normalized by eYFP for panel B. (n=3) (D) Diagram of the promoter region of the reporter constructs. The 20 bp sequence CDE is placed at different distance to core promoter in the 5' UTR of DSCP promoter. (E) Northern blot analysis of mRNA levels of luciferase under different promoters. Probes hybridize 5' Myc-tag sequence for luciferase and 5' region for eYFP. (F) Quantification of De/Opt *Luc* mRNA ratio normalized by eYFP for panel E. (n=3)

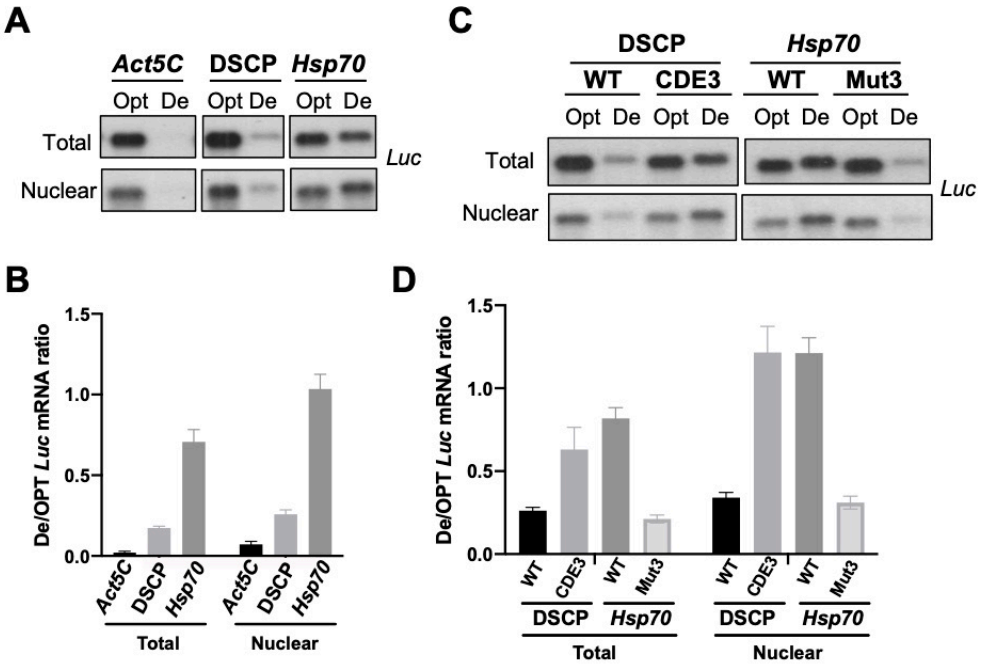

Figure 3-5. Different promoters impact nuclear RNA levels of nonoptimal codons. (A) Northern blot analysis of mRNA levels of luciferase under different promoters in total and nuclear RNA fractions. Probe hybridizes 5' Myc-tag sequence for luciferase. (B) Quantification of De/Opt mRNA ratio for panel A. (n=3). (C) Northern blot analysis of mRNA levels of luciferase under different promoters in nuclear and total RNA fraction. Probe hybridizes 5' Myc-tag sequence for luciferase. (D) Quantification of De/Opt *Luc* mRNA ratio for panel C. (n=3).

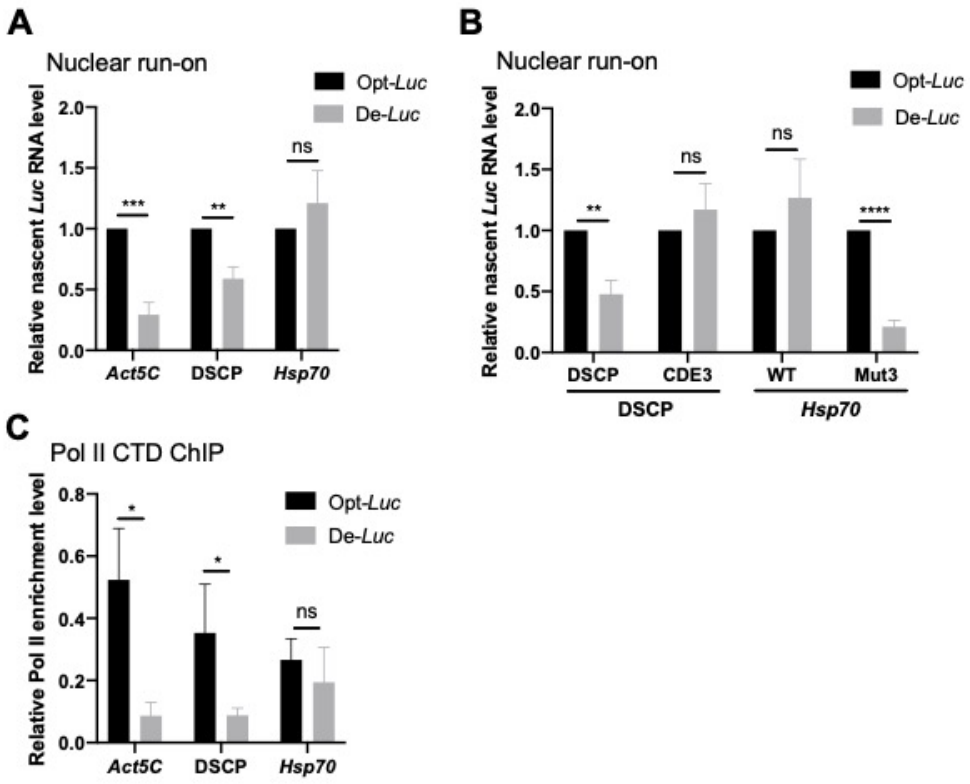

Figure 3-6. Promoter regulates transcription level to affect codon-dependent differential expression. (A) Relative nascent *Luc* RNA level detected by qPCR in nuclear run-on assay. qPCR primers amplify the 5' Myc-tag sequence and is normalized by *Act5C*. (n=4). (B) Relative nascent *Luc* RNA level detected by qPCR in nuclear run-on assay. qPCR primers amplify the 5' Myc-tag sequence and is normalized by *Act5C*. (n=4). (C) Pol II CTD ChIP assay detecting Pol II level at the 5' Myc-tag region of different constructs. qPCR primers amplify the 5' Myc-tag sequence and is normalized by *Act5C*. (n=3). Data are means ± SD. **P* < 0.05. ***P* < 0.01. ****P* < 0.001. *****P* < 0.001. ns: not significant.

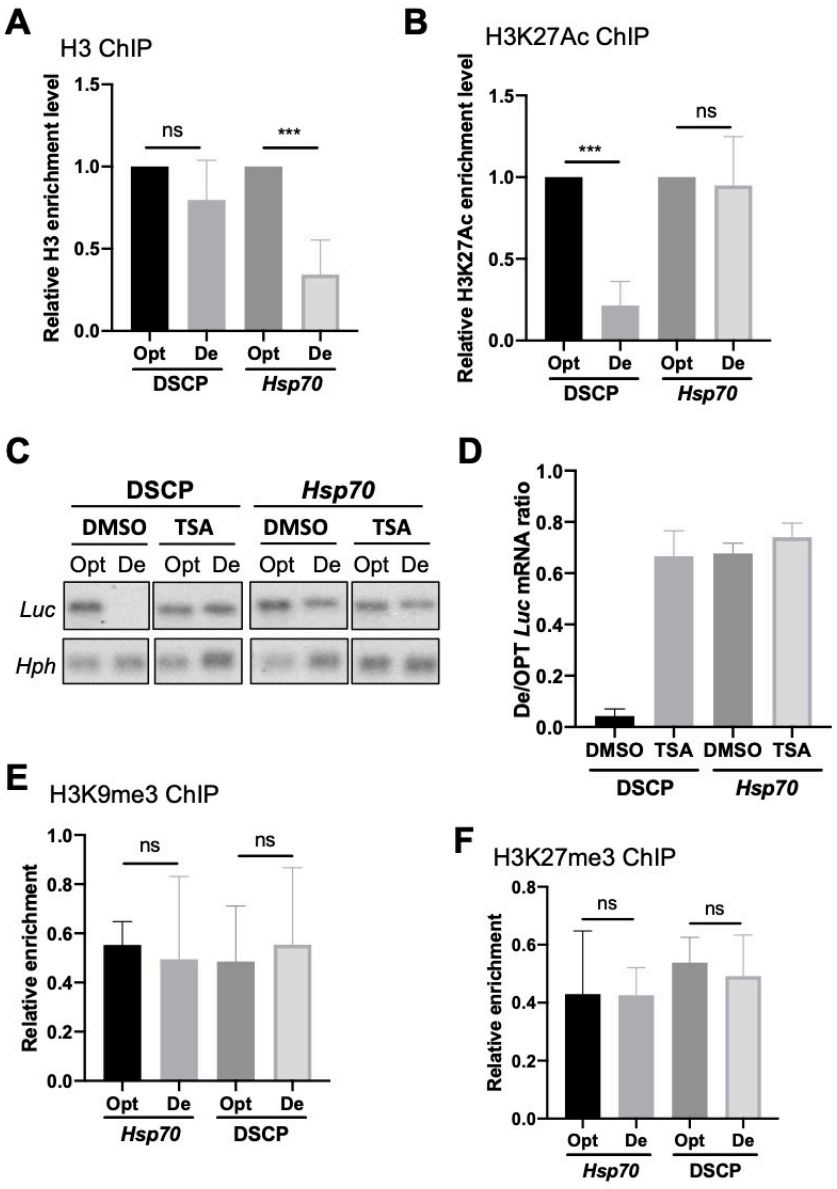

Figure 3-7. codon usage and promoter together determine histone modification and nucleosome density. (A) Histone H3 ChIP assay detecting nucleosome level at the 5' Myc-tag region of different constructs. qPCR primers amplify the 5' Myc-tag sequence and is normalized by a transcription inactivate region 3L1. (n=4). (B) H3K27ac ChIP assay

detecting H3K27ac level at the 5’ Myc-tag region of different constructs. qPCR primers amplify the 5’ Myc-tag sequence and is normalized by *Act5C*. (n=4) (C) Northern blot analysis of mRNA levels of luciferase under different promoters with DMSO or 150 μM TSA treatment. Probe hybridizes 5’ Myc-tag sequence for luciferase and 5’ region for hph. (D) Quantification of De/Opt *Luc* mRNA ratio normalized by *hph* for panel (C). (n=3). Data are means ± SD. $^{**}P < 0.01$. ns: not significant. (E) H3K9me3 ChIP assay detecting H3K9me3 level at the 5’ Myc-tag region of different constructs. qPCR primers amplify the 5’ Myc-tag sequence and is normalized by F22. (n=3). (F) H3K27me3 ChIP assay detecting H3K27me3 level at the 5’ Myc-tag region of different constructs. qPCR primers amplify the 5’ Myc-tag sequence and is normalized by a region in 3L. (n=3).

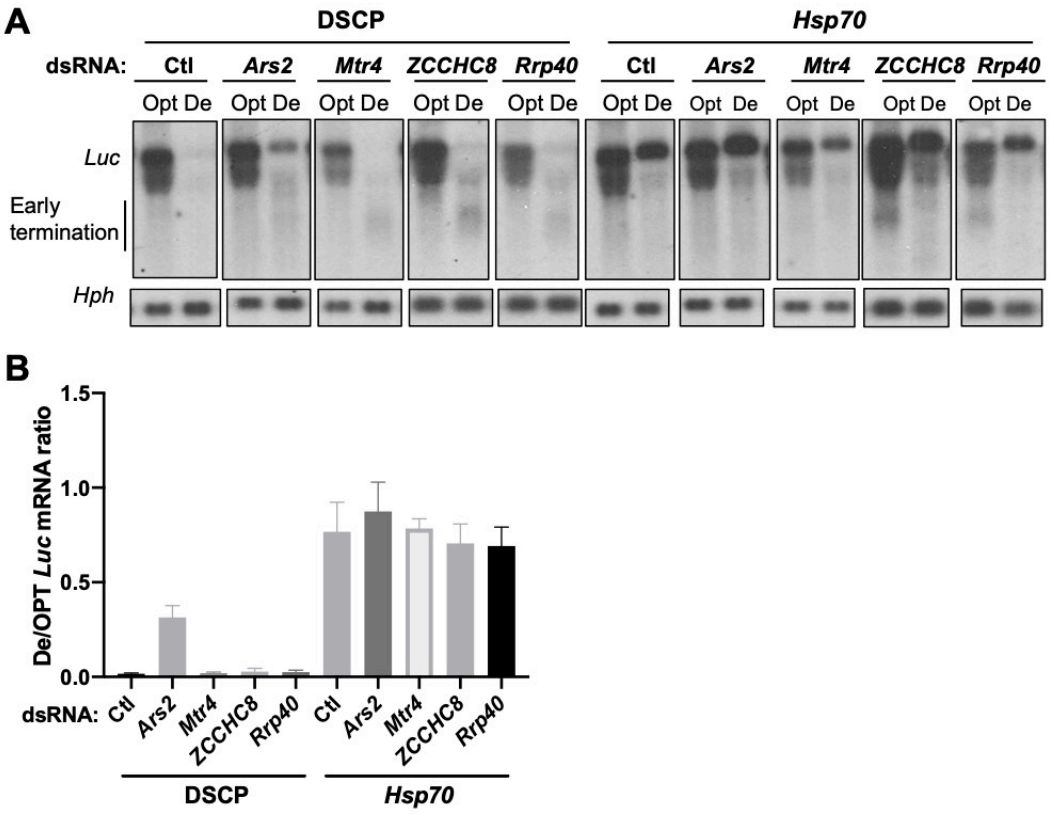

Figure 3-8. Nonoptimal codons cause early termination of transcription mediated by Ars2-NEXT pathway. (A) Northern blot analysis of mRNA levels of luciferase under different promoters in cells treated with dsRNAs targeting Ars2, Mtr4, ZCCHC8, Rrp40 or no dsRNA as mock control. Probes hybridize 5' Myc-tag sequence for luciferase and 5' region for *hph*. (B) Quantification of De/Opt *Luc* mRNA ratio normalized by *hph* for panel A. (n=3).

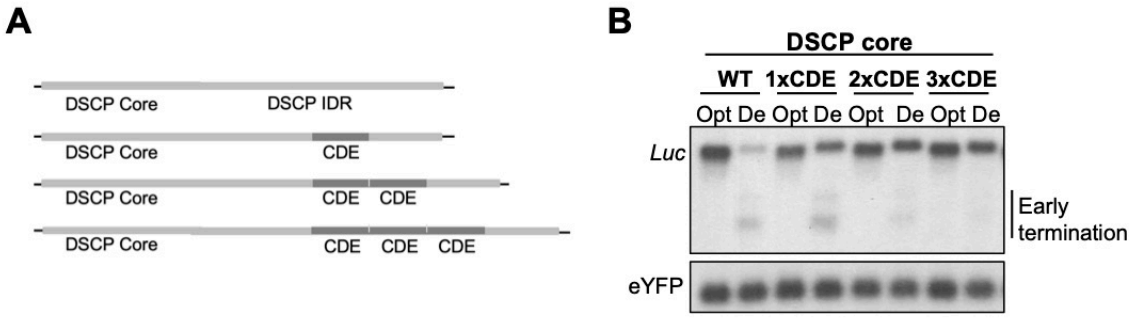

Figure 3-9. CDE prevents early termination of transcription of nonoptimal codon. (A) Diagram of the promoter region of the reporter constructs. Different copies of the CDE sequence is placed in the 5' UTR of DSCP promoter. (B) Northern blot analysis of mRNA levels of luciferase under different promoters. Probes hybridize 5' Myc-tag sequence for luciferase and 5' region for eYFP.

Table 3-1. Primer list

Name	Sequence	Note
Ars2KD-F	TAATACGACTCACTATAGGGCTAGCTCAGATG AAGAGAAAC	RNAi
Ars2KD-R	TAATACGACTCACTATAGGGATACCAACGACG CTCCAC	RNAi
Mtr4KD-F	TAATACGACTCACTATAGGGGTGCTCACCGAG GAGGAT	RNAi
Mtr4KD-R	TAATACGACTCACTATAGGGCAGTGCAGCTTG ATTTTGG	RNAi
ZCCHC8KD-F	TAATACGACTCACTATAGGGTCAGCCTTCGAAA GAAGTAG	RNAi
ZCCHC8KD-R	TAATACGACTCACTATAGGGTGCATTAAACAG AGCTATGC	RNAi
Rrp40KD-F	TAATACGACTCACTATAGGGCAGCCTCCATATC GTATCTC	RNAi
Rrp40KD-R	TAATACGACTCACTATAGGGCGAGTTGACGCA GACCA	RNAi
Myc-F	GATATCATCGATTTAAAGCA	RT-qPCR, ChIP-qPCR
Myc-R	CATGTCGCCCAAGCTCTCCAT	RT-qPCR, ChIP-qPCR
ActinIntron-F	GAGAAAAGCCGCGGAAAATGTGTG	RT-qPCR, ChIP-qPCR
ActinIntron-R	TCAATACAATAACTCTTTAGCTCG	RT-qPCR, ChIP-qPCR
3L-F	GCACACGGTAATTGCTTATT	ChIP-qPCR
3L-R	TCGCATCTCGGTTCTTTC	ChIP-qPCR
F22-F	CAGTTGATGGGATGAATTTGG	ChIP-qPCR
F22-R	TGCCTGTGGTTCTATCCAAAC	ChIP-qPCR
MycProbe-F	TGATATCATCGATTTAAAGCA	Northern probe
MycProbe-R	TGCTAATACGACTCACTATAGGGAGCATGTCGC CCAAGCTCTCCAT	Northern probe
eYFPProbe-F	ATGGTGAGCAAGGGCGAGGAGCTG	Northern probe
eYFPProbe-R	TGCTAATACGACTCACTATAGGGAGCCTCGCCG GACACGCTGAAC	Northern probe
YFP3UProbe-F	TCTTTGTGAAGGAACCTTAC	Northern probe
YFP3UProbe-R	TGCTAATACGACTCACTATAGGGAGTTCATCAG TTCCATAGGTTG	Northern probe

Ars2Probe-F	TATCGGTTACCATGATTTGG	Northern probe
Ars2Probe-R	TGCTAATACGACTCACTATAGGGAGTCGCACTTTATTTCCTTGTG	Northern probe
Mtr4Probe-F	CGACTGGAGGAGCTTCTGCG	Northern probe
Mtr4Probe-R	TGCTAATACGACTCACTATAGGGAGATTTACAAAGGTTGTTAGGG	Northern probe
ZCCHC8Probe-F	GCTCGTCGCCTTGGAAGGCG	Northern probe
ZCCHC8Probe-R	TGCTAATACGACTCACTATAGGGAGTCCAGCCAAGCAGGCGGATA	Northern probe
Rrp40Probe-F	GAGCTGGTCTGCGTCAACTC	Northern probe
Rrp40Probe-R	TGCTAATACGACTCACTATAGGGAGTAAGGGTAGTACTATCTTTC	Northern probe
Hsp70Probe-F	ATGCCTGCTATTGGAATCGATCTG	Northern probe
Hsp70Probe-R	TGCTAATACGACTCACTATAGGGAGCCGATCAGTCGCTTGGCGTCAAAC	Northern probe
HphProbe-F	ATGAAAAAGCCTGAACTCACCG	Northern probe
HphProbe-R	TGCTAATACGACTCACTATAGGGAGTTTGTAGAAACCATCGGCGCAGC	Northern probe

Table 3-2 Promoter list (Core promoter is underlined region)

> Act5C promoter
GCTAAGCGGGCTTTATAAAACGGGCTGCGGGACCAGTTTTCATATCACTACCG TTTGAGTTCTTGTGCTGTGTGGATACTCCTCCCGACACAAAGCCGCTCCATCAG CCAGCAGTCGTCTAATCCAGAGACACCAAACCGAAAGACTTAATTTATATTTA TTTAATTAATTTTAATAAAACACACCAAAT
> DSCP promoter
CGCCCGGGGATCGAGCGCAGCGGTATAAAAGGGCGCGGGGTGGCTGAGAGCA TCAGTTGTGAATGAATGTTCGAGCCGAGCAGACGTGCCGCTGCCTTCGTTAAT ATCCTTTGAATAAGCCAACTTTGAATCACAAGACGCATACCAAACGGCATTCC GGTACTGTTGGTAAAGCCACC
> Tub promoter
TGGCCACACTGCGGCCATCGTATAAAAGCCCGCGCTCTCCAAAGCGAATGCAC TAATTTTTCCAAGTGTGTGAAGCGGACAGTTTGTGTTGTGTTCGACTGCTATAA GCGAAAGATAATTTTAAGTTAAAAGTTCAAAAGCCTAAAACAAATACAAAAT GAGGGAAATCGTTCACATCCAAGCTGGTCAGTGCGGCAACCAAATCGGCGCCA AG
> Hsp70 promoter
CGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAATTCAAA CAAGCAAAGTGAACACGTCGCTAAGCGAAAGCTAAGCAAATAAACAAGCGCA GCTGAACAAGCTAAACAATCTGCAGTAAAGTGCAAGTTAAAGTGAATCAATTA AAAGTAACCAGCAAC
> cMyc promoter
TACGATAGAAATTTTATTTAAGCCACAGACAACATGAAACGGGCACTATTTCT GTGGCGTCGCGTGTTCAGTTCACCGCGGGTAATTCAGAGAATCGCTTTGTGGA TTGGATTTTTGCCTGTTTTCCGCCCGATACAAAAAAAAAAAACCAAACGCTAT ATAAATAGTTCTGTAGTAAAACCTGAAGCAACACGTTTTAAAATATACAACTA CTACTAACAACTGTCACAGCCAAGTTACAAAAGTGCTAAATC
> Per promoter
TGCGAGTGCACGGGGGCTTATACACCGGAGCTTGCGATGCGAGCACTAGTGTT CGTGCGAATTTAGAGCCAGAAGGTCGACAACGCGTTCGTCGCACAGTAGAAA GTATTGCCATATACCAAATAAATCGCAACAACGACCTGTCGAATCACGCAAAC AAACAGCTTCGTTTTTTTTTTTTTAATTTCGACCAGAAACAAACCGTTATTCTTT

CHAPTER FOUR
Discussion and future directions

The role of eRF1 in termination at sense codons

The canonical translation termination factor eRF1 is known to specifically recognize stop codons(Brown et al., 2015). Multiple motifs including GTS31-33, E55, TASNIKS58-64, and YxCxxxF125-130 located at the apex of the N-domain in eRF1 are essential for stop codon recognition(Bertram et al., 2000; Frolova et al., 2002; Frolova et al., 1999; Song et al., 2000). It could lead to translation efficiency decrease and global protein production issue if eRF1 has low discrimination ability for sense codons. However, in chapter two, I showed that eRF1 mediates premature termination at rare codons and reduces translation efficiency. The results support that eRF1 also dissociates ribosome at sense codons, which is also in consistent with recent discoveries(Chiabudini et al., 2014; Shcherbik et al., 2016; Wada & Ito, 2019). It is possible that the specificity of stop codon recognition is also dependent on the ribosome pausing time. Supporting this hypothesis, results from ribosome profiling assays showed that ribosome has a dramatically longer time at stop codon than sense codons(C. C.-C. Wu et al., 2019a). More further investigations are needed to test whether ribosome pausing is a critical determinant in the specificity of eRF1 in codon recognition. Moreover, based on our results, the interaction of decoding domain that recognize stop codons and sense codons is not required but still critical. Further studies on how eRF1 recognize sense codon and trigger ribosome release will lead to better understanding of eRF1 and codon interaction.

In Dom34 mediated ribosome quality control pathway, the nascent peptides generated from truncated or faulty mRNAs are ubiquitinated and degraded by proteosome(Lyumkis et al., 2014; Shao et al., 2015). Although multiple studies showed the role of eRF1 on releasing nascent peptides at sense codon, the fate of these nascent peptides are unknown(Shcherbik et al., 2016; C. C.-C. Wu et al., 2019a). If these peptides are released to cytoplasm without labeling them as degradation substrates, then excess production of these truncated peptides would be detrimental to cells. Thus, further studies are needed to investigate the pathways that degrade the truncated peptides generated by eRF1 mediated premature termination of translation.

Using in vitro cell-free translation assay and reporters in vivo expressed in *Neurospora*, I showed that eRF1 is the key factor in regulating codon usage dependent translation efficiency. Although I also demonstrated that nonoptimal codon also decreases translation efficiency using reporters in *Drosophila* S2 cells, the role of eRF1 in *Drosophila* is unclear. Due to the difficulty of manipulating expression of essential genes in higher organisms, simple knockdown of eRF1 by RNAi in *Drosophila* cells did not show impact on reporter gene expression. Thus, whether this mechanism is conserved in higher organisms needs additional tests.

Regulatory role of promoter downstream element

Due to the expansion of noncoding region of the genome in higher organisms, the regulation network is greatly complicated at transcriptional level(Blackwood & Kadonaga, 1998; Mordstein et al., 2020; Shlyueva et al., 2014). Besides the enhancer roles of noncoding

region, transcription of noncoding region also impacts transcription of adjacent protein-coding genes(Barski et al., 2010; Field & Adelman, 2020; Oler et al., 2010). Moreover, processing of noncoding region such as splicing is critical for transcription level and RNA export(Le Hir et al., 2003; Mordstein et al., 2020). In chapter three, I showed that the noncoding region downstream of core promoter determines the transcription level of genes with nonoptimal codon usage. It is possible that more regulatory elements remain to be uncovered. In addition, since we only used reporters rather than endogenous genes, the generality and conservation of the phenomenon needs further investigation. For example, our reporters contain no intron and poorly represents the higher organisms' genome, where the majority of protein coding genes contain various numbers of introns(Castillo-Davis et al., 2002; Le Hir et al., 2003). Interestingly, a recent report showed the effect of splicing enhance nuclear export of nonoptimal codon genes(C. C.-C. Wu et al., 2019a). Thus, the influence of large intronic sequences on codon-dependent transcription regulation needs further study.

Using promoter from Hsp70 as an example, I identified the regulatory element CDE in the downstream region of core promoter that is capable to enhance nonoptimal gene transcription. In other promoters such as *Myc* and *Per*, I also found they contain sequences that promote the expression of nonoptimal coding sequences. However, no similarity could be detected between the sequences from *Hsp70*, *Myc* and *Per*. Due to lack of enough sequences for motif enrichment analysis, the consensus motifs that enhances nonoptimal gene transcription are still unclear. Thus, insights from genome-wide investigations on endogenous genes or utilizing synthetic sequence library are needed for enrichment of the consensus motif that can promote the expression of nonoptimal codon sequences.

The effects of regulatory element downstream of core promoter and codon usage together determine the nucleosome density and H3K27ac level. Low nucleosome density and high H3K27ac level are both correlated with high transcription level(Clapier & Cairns, 2009; Roh et al., 2004). However, it is still unclear whether the epigenetic features lead to transcription change or they are the post-transcriptional markers reflecting transcriptional level. Thus, additional studies are needed to understand the role of regulatory element downstream of core promoter in affecting epigenetic regulatory elements.

BIBLIOGRAPHY

This is an entry that is preformatted to have a hanging indent that will bring the second line of the bibliographic reference in about .5 inch. Apply the Reference Item style to produce this format.

Adalat, R., Saleem, F., Bashir, A., Ahmad, M., Zulfiqar, S., & Shakoori, A. R. (2018, Sep 22). Multiple upstream start codons (AUG) in 5' untranslated region enhance translation efficiency of cry2Ac11 without helper protein. *J Cell Biochem*. https://doi.org/10.1002/jcb.27534

Andersen, P. R., Domanski, M., Kristiansen, M. S., Storvall, H., Ntini, E., Verheggen, C., Schein, A., Bunkenborg, J., Poser, I., Hallais, M., Sandberg, R., Hyman, A., LaCava, J., Rout, M. P., Andersen, J. S., Bertrand, E., & Jensen, T. H. (2013, Dec). The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat Struct Mol Biol, 20*(12), 1367-1376. https://doi.org/10.1038/nsmb.2703

Arango, D., Sturgill, D., Alhusaini, N., Dillman, A. A., Sweet, T. J., Hanson, G., Hosogane, M., Sinclair, W. R., Nanan, K. K., Mandler, M. D., Fox, S. D., Zengeya, T. T., Andresson, T., Meier, J. L., Coller, J., & Oberdoerffer, S. (2018, Nov 12). Acetylation of Cytidine in mRNA Promotes Translation Efficiency. *Cell*. https://doi.org/10.1016/j.cell.2018.10.030

Arigo, J. T., Eyler, D. E., Carroll, K. L., & Corden, J. L. (2006, Sep 15). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol Cell, 23*(6), 841-851. https://doi.org/10.1016/j.molcel.2006.07.024

Banerji, J., Rusconi, S., & Schaffner, W. (1981, Dec). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell, 27*(2 Pt 1), 299-308. https://doi.org/10.1016/0092-8674(81)90413-x

Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., & Kouzarides, T. (2001, Mar 1). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature, 410*(6824), 120-124. https://doi.org/10.1038/35065138

Barski, A., Chepelev, I., Liko, D., Cuddapah, S., Fleming, A. B., Birch, J., Cui, K., White, R. J., & Zhao, K. (2010, 2010/05/01). Pol II and its associated epigenetic marks are present at Pol III–transcribed noncoding RNA genes. *Nature Structural & Molecular Biology, 17*(5), 629-634. https://doi.org/10.1038/nsmb.1806

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., & Zhao, K. (2007, May 18). High-resolution profiling of histone methylations in the human genome. *Cell, 129*(4), 823-837. https://doi.org/10.1016/j.cell.2007.05.009

Bazzini, A. A., Del Viso, F., Moreno-Mateos, M. A., Johnstone, T. G., Vejnar, C. E., Qin, Y., Yao, J., Khokha, M. K., & Giraldez, A. J. (2016, Oct 4). Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *Embo j, 35*(19), 2087-2103. https://doi.org/10.15252/embj.201694699

Becker, T., Armache, J. P., Jarasch, A., Anger, A. M., Villa, E., Sieber, H., Motaal, B. A., Mielke, T., Berninghausen, O., & Beckmann, R. (2011, Jun). Structure of the no-go mRNA decay complex Dom34-Hbs1 bound to a stalled 80S ribosome. *Nat Struct Mol Biol, 18*(6), 715-720. https://doi.org/10.1038/nsmb.2057

Bennetzen, J. L., & Hall, B. D. (1982, Mar 25). Codon selection in yeast. *J Biol Chem, 257*(6), 3026-3031.

Bertram, G., Bell, H. A., Ritchie, D. W., Fullerton, G., & Stansfield, I. (2000, Sep). Terminating eukaryote translation: domain 1 of release factor eRF1 functions in stop codon recognition. *Rna, 6*(9), 1236-1247.

Blackwood, E. M., & Kadonaga, J. T. (1998, Jul 3). Going the distance: a current view of enhancer action. *Science, 281*(5373), 60-63. https://doi.org/10.1126/science.281.5373.60

Boel, G., Letso, R., Neely, H., Price, W. N., Wong, K. H., Su, M., Luff, J., Valecha, M., Everett, J. K., Acton, T. B., Xiao, R., Montelione, G. T., Aalberts, D. P., & Hunt, J. F. (2016, Jan 21). Codon influence on protein expression in E. coli correlates with mRNA levels. *Nature, 529*(7586), 358-363. https://doi.org/10.1038/nature16509

Bonekamp, F., & Jensen, K. F. (1988, Apr 11). The AGG codon is translated slowly in E. coli even at very low expression levels. *Nucleic Acids Res, 16*(7), 3013-3024. https://doi.org/10.1093/nar/16.7.3013

Breathnach, R., & Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Annu Rev Biochem, 50*, 349-383. https://doi.org/10.1146/annurev.bi.50.070181.002025

Brockmann, R., Beyer, A., Heinisch, J. J., & Wilhelm, T. (2007, Mar 23). Posttranscriptional expression regulation: what determines translation rates? *PLoS Comput Biol, 3*(3), e57. https://doi.org/10.1371/journal.pcbi.0030057

Brown, A., Shao, S., Murray, J., Hegde, R. S., & Ramakrishnan, V. (2015, Aug 27). Structural basis for stop codon recognition in eukaryotes. *Nature, 524*(7566), 493-496. https://doi.org/10.1038/nature14896

Bulmer, M. (1987, Feb 19-25). Coevolution of codon usage and transfer RNA abundance. *Nature, 325*(6106), 728-730. https://doi.org/10.1038/325728a0

Burgess-Brown, N. A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U., & Gileadi, O. (2008, May). Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study. *Protein Expr Purif, 59*(1), 94-102. https://doi.org/10.1016/j.pep.2008.01.008

Burke, T. W., & Kadonaga, J. T. (1997). The downstream core promoter element, DPE, is conserved fromDrosophila to humans and is recognized by TAFII60 of Drosophila. *Genes & development, 11*(22), 3020-3031.

Buschauer, R., Matsuo, Y., Sugiyama, T., Chen, Y. H., Alhusaini, N., Sweet, T., Ikeuchi, K., Cheng, J., Matsuki, Y., Nobuta, R., Gilmozzi, A., Berninghausen, O., Tesina, P., Becker, T., Coller, J., Inada, T., & Beckmann, R. (2020, Apr 17). The Ccr4-Not complex monitors the translating ribosome for codon optimality. *Science, 368*(6488). https://doi.org/10.1126/science.aay6912

Butler, J. E., & Kadonaga, J. T. (2001, Oct 1). Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev, 15*(19), 2515-2519. https://doi.org/10.1101/gad.924301

Cairns, B. R. (2005, Apr). Chromatin remodeling complexes: strength in diversity, precision through specialization. *Curr Opin Genet Dev, 15*(2), 185-190. https://doi.org/10.1016/j.gde.2005.01.003

Cambray, G., Guimaraes, J. C., & Arkin, A. P. (2018, 2018/11/01). Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in Escherichia coli. *Nat Biotechnol, 36*(10), 1005-1015. https://doi.org/10.1038/nbt.4238

Cannarozzi, G., Schraudolph, N. N., Faty, M., von Rohr, P., Friberg, M. T., Roth, A. C., Gonnet, P., Gonnet, G., & Barral, Y. (2010, Apr 16). A role for codon order in translation dynamics. *Cell, 141*(2), 355-367. https://doi.org/10.1016/j.cell.2010.02.036

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S., & Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science, 298*(5595), 1039-1043.

Castillo-Davis, C. I., Mekhedov, S. L., Hartl, D. L., Koonin, E. V., & Kondrashov, F. A. (2002, 2002/08/01). Selection for short introns in highly expressed genes. *Nature Genetics, 31*(4), 415-418. https://doi.org/10.1038/ng940

Charneski, C. A., & Hurst, L. D. (2013). Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol, 11*(3), e1001508.

Chavatte, L., Frolova, L., Laugaa, P., Kisselev, L., & Favre, A. (2003, Aug 22). Stop codons and UGG promote efficient binding of the polypeptide release factor eRF1 to the ribosomal A site. *J Mol Biol, 331*(4), 745-758.

Cheng, P., He, Q., He, Q., Wang, L., & Liu, Y. (2005, Jan 15). Regulation of the Neurospora circadian clock by an RNA helicase. *Genes Dev, 19*(2), 234-241. https://doi.org/10.1101/gad.1266805

Chiabudini, M., Tais, A., Zhang, Y., Hayashi, S., Wölfle, T., Fitzke, E., & Rospert, S. (2014, Nov). Release factor eRF3 mediates premature translation termination on polylysine-stalled ribosomes in Saccharomyces cerevisiae. *Mol Cell Biol, 34*(21), 4062-4076. https://doi.org/10.1128/mcb.00799-14

Chu, D., Barnes, D. J., & von der Haar, T. (2011, Aug). The role of tRNA and ribosome competition in coupling the expression of different mRNAs in Saccharomyces cerevisiae. *Nucleic Acids Res, 39*(15), 6705-6714. https://doi.org/10.1093/nar/gkr300

Chu, D., Kazana, E., Bellanger, N., Singh, T., Tuite, M. F., & von der Haar, T. (2014, Jan 7). Translation elongation can control translation initiation on eukaryotic mRNAs. *Embo j, 33*(1), 21-34. https://doi.org/10.1002/embj.201385651

Clapier, C. R., & Cairns, B. R. (2009). The biology of chromatin remodeling complexes. *Annu Rev Biochem, 78*, 273-304. https://doi.org/10.1146/annurev.biochem.77.062706.153223

Crick, F. H., Barnett, L., Brenner, S., & Watts-Tobin, R. J. (1961, Dec 30). General nature of the genetic code for proteins. *Nature, 192*, 1227-1232. https://doi.org/10.1038/1921227a0

Curran, J. F. (1995, Feb 25). Decoding with the A:I wobble pair is inefficient. *Nucleic Acids Res, 23*(4), 683-688.

Dana, A., & Tuller, T. (2014a, Aug). The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res, 42*(14), 9171-9181. https://doi.org/10.1093/nar/gku646

Dana, A., & Tuller, T. (2014b). Properties and determinants of codon decoding time distributions. *BMC Genomics, 15 Suppl 6*(Suppl 6), S13. https://doi.org/10.1186/1471-2164-15-s6-s13

Defenouillere, Q., Yao, Y., Mouaikel, J., Namane, A., Galopier, A., Decourty, L., Doyen, A., Malabat, C., Saveanu, C., Jacquier, A., & Fromont-Racine, M. (2013, Mar 26). Cdc48-associated complex bound to 60S particles is required for the clearance of aberrant translation products. *Proc Natl Acad Sci U S A, 110*(13), 5046-5051. https://doi.org/10.1073/pnas.1221724110

Dimitrova, L. N., Kuroha, K., Tatematsu, T., & Inada, T. (2009). Nascent peptide-dependent translation arrest leads to Not4p-mediated protein degradation by the proteasome. *Journal of Biological Chemistry, 284*(16), 10343-10352.

Doma, M. K., & Parker, R. (2006, Mar 23). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature, 440*(7083), 561-564. https://doi.org/10.1038/nature04530

dos Reis, M., Savva, R., & Wernisch, L. (2004). Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res, 32*(17), 5036-5044. https://doi.org/10.1093/nar/gkh834

Duret, L. (2000, Jul). tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet, 16*(7), 287-289.

Duret, L. (2002, Dec). Evolution of synonymous codon usage in metazoans. *Curr Opin Genet Dev, 12*(6), 640-649. https://doi.org/10.1016/s0959-437x(02)00353-2

Duret, L., & Mouchiroud, D. (1999, Apr 13). Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. *Proc Natl Acad Sci U S A, 96*(8), 4482-4487. https://doi.org/10.1073/pnas.96.8.4482

Eyler, D. E., Wehner, K. A., & Green, R. (2013, Oct 11). Eukaryotic release factor 3 is required for multiple turnovers of peptide release catalysis by eukaryotic release factor 1. *J Biol Chem, 288*(41), 29530-29538. https://doi.org/10.1074/jbc.M113.487090

Falk, S., Finogenova, K., Melko, M., Benda, C., Lykke-Andersen, S., Jensen, T. H., & Conti, E. (2016, Dec 1). Structure of the RBM7-ZCCHC8 core of the NEXT complex reveals connections to splicing factors. *Nat Commun, 7*, 13573. https://doi.org/10.1038/ncomms13573

Falk, S., Finogenova, K., Melko, M., Benda, C., Lykke-Andersen, S., Jensen, T. H., & Conti, E. (2016, 2016/12/01). Structure of the RBM7–ZCCHC8 core of the NEXT complex reveals connections to splicing factors. *Nature Communications, 7*(1), 13573. https://doi.org/10.1038/ncomms13573

Fan, J., Kuai, B., Wu, G., Wu, X., Chi, B., Wang, L., Wang, K., Shi, Z., Zhang, H., Chen, S., He, Z., Wang, S., Zhou, Z., Li, G., & Cheng, H. (2017, Oct 2). Exosome cofactor hMTR4 competes with export adaptor ALYREF to ensure balanced nuclear RNA pools for degradation and export. *Embo j, 36*(19), 2870-2886. https://doi.org/10.15252/embj.201696139

Fan-Minogue, H., Du, M., Pisarev, A. V., Kallmeyer, A. K., Salas-Marco, J., Keeling, K. M., Thompson, S. R., Pestova, T. V., & Bedwell, D. M. (2008, Jun 6). Distinct eRF3 requirements suggest alternate eRF1 conformations mediate peptide release during eukaryotic translation termination. *Mol Cell, 30*(5), 599-609. https://doi.org/10.1016/j.molcel.2008.03.020

Fang, P., Spevak, C. C., Wu, C., & Sachs, M. S. (2004, Mar 23). A nascent polypeptide domain that can regulate translation elongation. *Proc Natl Acad Sci U S A, 101*(12), 4059-4064. https://doi.org/10.1073/pnas.0400554101

Fang, P., Wang, Z., & Sachs, M. S. (2000, Sep 1). Evolutionarily conserved features of the arginine attenuator peptide provide the necessary requirements for its function in translational regulation. *J Biol Chem, 275*(35), 26710-26719. https://doi.org/10.1074/jbc.M003175200

Ferrin, M. A., & Subramaniam, A. R. (2017, May 12). Kinetic modeling predicts a stimulatory role for ribosome collisions at elongation stall sites in bacteria. *Elife, 6*. https://doi.org/10.7554/eLife.23629

Ferrin, M. A., & Subramaniam, A. R. (2017). Kinetic modeling predicts a stimulatory role for ribosome collisions at elongation stall sites in bacteria. *Elife, 6*, e23629.

Field, A., & Adelman, K. (2020). Evaluating Enhancer Function and Transcription. *Annual Review of Biochemistry, 89*(1), 213-234. https://doi.org/10.1146/annurev-biochem-011420-095916

Frolova, L., Le Goff, X., Rasmussen, H. H., Cheperegin, S., Drugeon, G., Kress, M., Arman, I., Haenni, A. L., Celis, J. E., Philippe, M., & et al. (1994, Dec 15). A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. *Nature, 372*(6507), 701-703. https://doi.org/10.1038/372701a0

Frolova, L., Seit-Nebi, A., & Kisselev, L. (2002, Feb). Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1. *Rna, 8*(2), 129-136.

Frolova, L. Y., Tsivkovskii, R. Y., Sivolobova, G. F., Oparina, N. Y., Serpinsky, O. I., Blinov, V. M., Tatkov, S. I., & Kisselev, L. L. (1999, Aug). Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *Rna, 5*(8), 1014-1020.

Fu, J., Dang, Y., Counter, C., & Liu, Y. (2018, Nov 16). Codon usage regulates human KRAS expression at both transcriptional and translational levels. *J Biol Chem, 293*(46), 17929-17940. https://doi.org/10.1074/jbc.RA118.004908

Fu, J., Murphy, K. A., Zhou, M., Li, Y. H., Lam, V. H., Tabuloc, C. A., Chiu, J. C., & Liu, Y. (2016, Aug 1). Codon usage affects the structure and function of the Drosophila circadian clock protein PERIOD. *Genes Dev, 30*(15), 1761-1775. https://doi.org/10.1101/gad.281030.116

Gamble, C. E., Brule, C. E., Dean, K. M., Fields, S., & Grayhack, E. J. (2016, Jul 28). Adjacent Codons Act in Concert to Modulate Translation Efficiency in Yeast. *Cell, 166*(3), 679-690. https://doi.org/10.1016/j.cell.2016.05.070

Gardin, J., Yeasmin, R., Yurovsky, A., Cai, Y., Skiena, S., & Futcher, B. (2014, Oct 27). Measurement of average decoding rates of the 61 sense codons in vivo. *Elife, 3*. https://doi.org/10.7554/eLife.03735

Garzia, A., Jafarnejad, S. M., Meyer, C., Chapat, C., Gogakos, T., Morozov, P., Amiri, M., Shapiro, M., Molina, H., & Tuschl, T. (2017). The E3 ubiquitin ligase and RNA-binding protein ZNF598 orchestrates ribosome quality control of premature polyadenylated mRNAs. *Nature Communications, 8*(1), 1-10.

Gerashchenko, M. V., & Gladyshev, V. N. (2014). Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic Acids Res, 42*(17), e134. https://doi.org/10.1093/nar/gku671

Giacometti, S., Benbahouche, N. E. H., Domanski, M., Robert, M.-C., Meola, N., Lubas, M., Bukenborg, J., Andersen, J. S., Schulze, W. M., Verheggen, C., Kudla, G., Jensen, T. H., & Bertrand, E. (2017, 2017/03/14/). Mutually Exclusive CBC-Containing Complexes Contribute to RNA Fate. *Cell Reports, 18*(11), 2635-2650. https://doi.org/https://doi.org/10.1016/j.celrep.2017.02.046

Giacometti, S., Benbahouche, N. E. H., Domanski, M., Robert, M. C., Meola, N., Lubas, M., Bukenborg, J., Andersen, J. S., Schulze, W. M., Verheggen, C., Kudla, G., Jensen, T. H., & Bertrand, E. (2017, Mar 14). Mutually Exclusive CBC-

Containing Complexes Contribute to RNA Fate. *Cell Rep, 18*(11), 2635-2650. https://doi.org/10.1016/j.celrep.2017.02.046

Gingold, H., & Pilpel, Y. (2011, Apr 12). Determinants of translation efficiency and accuracy. *Mol Syst Biol, 7*, 481. https://doi.org/10.1038/msb.2011.14

Gobet, C., Weger, B. D., Marquis, J., Martin, E., Neelagandan, N., Gachon, F., & Naef, F. (2020). Robust landscapes of ribosome dwell times and aminoacyl-tRNAs in response to nutrient stress in liver. *Proceedings of the National Academy of Sciences, 117*(17), 9630-9641. https://doi.org/10.1073/pnas.1918145117

Goel, N. S., Rao, G. S., Ycas, M., Bremermann, H. J., & King, L. (1972, Jun). A method for calculating codon frequencies in DNA. *J Theor Biol, 35*(3), 399-457. https://doi.org/10.1016/0022-5193(72)90143-9

González, M., Martín-Ruíz, I., Jiménez, S., Pirone, L., Barrio, R., & Sutherland, J. D. (2011). Generation of stable Drosophila cell lines using multicistronic vectors. *Sci Rep, 1*, 75. https://doi.org/10.1038/srep00075

Gouy, M., & Gautier, C. (1982, Nov 25). Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res, 10*(22), 7055-7074. https://doi.org/10.1093/nar/10.22.7055

Graille, M., Chaillet, M., & van Tilbeurgh, H. (2008, Mar 14). Structure of yeast Dom34: a protein related to translation termination factor Erf1 and involved in No-Go decay. *J Biol Chem, 283*(11), 7145-7154. https://doi.org/10.1074/jbc.M708224200

Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., & Mercier, R. (1981, Jan 10). Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res, 9*(1), r43-74. https://doi.org/10.1093/nar/9.1.213-b

Grewal, S. I., & Jia, S. (2007, Jan). Heterochromatin revisited. *Nat Rev Genet, 8*(1), 35-46. https://doi.org/10.1038/nrg2008

Gromadzka, A. M., Steckelberg, A. L., Singh, K. K., Hofmann, K., & Gehring, N. H. (2016, Mar 18). A short conserved motif in ALYREF directs cap- and EJC-dependent assembly of export complexes on spliced mRNAs. *Nucleic Acids Res, 44*(5), 2348-2361. https://doi.org/10.1093/nar/gkw009

Gruber, J. J., Olejniczak, S. H., Yong, J., La Rocca, G., Dreyfuss, G., & Thompson, C. B. (2012, Jan 13). Ars2 promotes proper replication-dependent histone mRNA 3' end formation. *Mol Cell, 45*(1), 87-98. https://doi.org/10.1016/j.molcel.2011.12.020

Gruber, J. J., Zatechka, D. S., Sabin, L. R., Yong, J., Lum, J. J., Kong, M., Zong, W. X., Zhang, Z., Lau, C. K., Rawlings, J., Cherry, S., Ihle, J. N., Dreyfuss, G., & Thompson, C. B. (2009, Jul 23). Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. *Cell, 138*(2), 328-339. https://doi.org/10.1016/j.cell.2009.04.046

Guimaraes, J. C., Rocha, M., & Arkin, A. P. (2014, Apr). Transcript level and sequence determinants of protein abundance and noise in Escherichia coli. *Nucleic Acids Res, 42*(8), 4791-4799. https://doi.org/10.1093/nar/gku126

Gustafsson, C., Govindarajan, S., & Minshull, J. (2004, Jul). Codon bias and heterologous protein expression. *Trends Biotechnol, 22*(7), 346-353. https://doi.org/10.1016/j.tibtech.2004.04.006

Gutierrez, E., Shin, B.-S., Woolstenhulme, C. J., Kim, J.-R., Saini, P., Buskirk, A. R., & Dever, T. E. (2013). eIF5A promotes translation of polyproline motifs. *Molecular Cell, 51*(1), 35-45.

Guydosh, N. R., & Green, R. (2014, Feb 27). Dom34 rescues ribosomes in 3' untranslated regions. *Cell, 156*(5), 950-962. https://doi.org/10.1016/j.cell.2014.02.006

Hallais, M., Pontvianne, F., Andersen, P. R., Clerici, M., Lener, D., Benbahouche, N. E. H., Gostan, T., Vandermoere, F., Robert, M.-C., Cusack, S., Verheggen, C., Jensen, T. H., & Bertrand, E. (2013, 2013/12/01). CBC–ARS2 stimulates 3′-end maturation of multiple RNA families and favors cap-proximal processing. *Nature Structural & Molecular Biology, 20*(12), 1358-1366. https://doi.org/10.1038/nsmb.2720

Hallais, M., Pontvianne, F., Andersen, P. R., Clerici, M., Lener, D., Benbahouche Nel, H., Gostan, T., Vandermoere, F., Robert, M. C., Cusack, S., Verheggen, C., Jensen, T. H., & Bertrand, E. (2013, Dec). CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. *Nat Struct Mol Biol, 20*(12), 1358-1366. https://doi.org/10.1038/nsmb.2720

Hampsey, M. (1998, Jun). Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev, 62*(2), 465-503.

Hershberg, R., & Petrov, D. A. (2008). Selection on codon bias. *Annu Rev Genet, 42*, 287-299. https://doi.org/10.1146/annurev.genet.42.110807.091442

Hilal, T., Yamamoto, H., Loerke, J., Burger, J., Mielke, T., & Spahn, C. M. (2016, Dec 20). Structural insights into ribosomal rescue by Dom34 and Hbs1 at near-atomic resolution. *Nat Commun, 7*, 13521. https://doi.org/10.1038/ncomms13521

Hrossova, D., Sikorsky, T., Potesil, D., Bartosovic, M., Pasulka, J., Zdrahal, Z., Stefl, R., & Vanacova, S. (2015, Apr 30). RBM7 subunit of the NEXT complex binds U-rich sequences and targets 3'-end extended forms of snRNAs. *Nucleic Acids Res, 43*(8), 4236-4248. https://doi.org/10.1093/nar/gkv240

Hrossova, D., Sikorsky, T., Potesil, D., Bartosovic, M., Pasulka, J., Zdrahal, Z., Stefl, R., & Vanacova, S. (2015). RBM7 subunit of the NEXT complex binds U-rich sequences and targets 3′-end extended forms of snRNAs. *Nucleic Acids Research, 43*(8), 4236-4248.

Huisinga, K. L., Brower-Toland, B., & Elgin, S. C. (2006). The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma, 115*(2), 110-122.

Hussmann, J. A., Patchett, S., Johnson, A., Sawyer, S., & Press, W. H. (2015, Dec). Understanding Biases in Ribosome Profiling Experiments Reveals Signatures of Translation Dynamics in Yeast. *PLoS Genet, 11*(12), e1005732. https://doi.org/10.1371/journal.pgen.1005732

Huter, P., Arenz, S., Bock, L. V., Graf, M., Frister, J. O., Heuer, A., Peil, L., Starosta, A. L., Wohlgemuth, I., Peske, F., Nováček, J., Berninghausen, O., Grubmüller, H., Tenson, T., Beckmann, R., Rodnina, M. V., Vaiana, A. C., & Wilson, D. N. (2017, Nov 2). Structural Basis for Polyproline-Mediated Ribosome Stalling and Rescue by the Translation Elongation Factor EF-P. *Mol Cell, 68*(3), 515-527.e516. https://doi.org/10.1016/j.molcel.2017.10.014

Iasillo, C., Schmid, M., Yahia, Y., Maqbool, M. A., Descostes, N., Karadoulama, E., Bertrand, E., Andrau, J.-C., & Jensen, T. H. (2017). ARS2 is a general suppressor of pervasive transcription. *Nucleic Acids Research, 45*(17), 10229-10241. https://doi.org/10.1093/nar/gkx647

Iasillo, C., Schmid, M., Yahia, Y., Maqbool, M. A., Descostes, N., Karadoulama, E., Bertrand, E., Andrau, J. C., & Jensen, T. H. (2017, Sep 29). ARS2 is a general suppressor of pervasive transcription. *Nucleic Acids Res, 45*(17), 10229-10241. https://doi.org/10.1093/nar/gkx647

Ikemura, T. (1981, Sep 25). Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. *J Mol Biol, 151*(3), 389-409.

Ikemura, T. (1985, Jan). Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol, 2*(1), 13-34. https://doi.org/10.1093/oxfordjournals.molbev.a040335

Ikeuchi, K., Tesina, P., Matsuo, Y., Sugiyama, T., Cheng, J., Saeki, Y., Tanaka, K., Becker, T., Beckmann, R., & Inada, T. (2019, Mar 1). Collided ribosomes form a unique structural interface to induce Hel2-driven quality control pathways. *Embo j, 38*(5). https://doi.org/10.15252/embj.2018100276

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., & Weissman, J. S. (2009, Apr 10). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science, 324*(5924), 218-223. https://doi.org/10.1126/science.1168978

Ingolia, N. T., Lareau, L. F., & Weissman, J. S. (2011, Nov 11). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell, 147*(4), 789-802. https://doi.org/10.1016/j.cell.2011.10.002

Ishimura, R., Nagy, G., Dotu, I., Zhou, H., Yang, X.-L., Schimmel, P., Senju, S., Nishimura, Y., Chuang, J. H., & Ackerman, S. L. (2014). Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. *Science, 345*(6195), 455-459.

Jackson, R. J., Hellen, C. U., & Pestova, T. V. (2010, Feb). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol, 11*(2), 113-127. https://doi.org/10.1038/nrm2838

Jeacock, L., Faria, J., & Horn, D. (2018, Mar 15). Codon usage bias controls mRNA and protein abundance in trypanosomatids. *Elife, 7*. https://doi.org/10.7554/eLife.32496

Juszkiewicz, S., Chandrasekaran, V., Lin, Z., Kraatz, S., Ramakrishnan, V., & Hegde, R. S. (2018). ZNF598 is a quality control sensor of collided ribosomes. *Molecular Cell, 72*(3), 469-481. e467.

Juszkiewicz, S., & Hegde, R. S. (2017). Initiation of quality control during poly (A) translation requires site-specific ribosome ubiquitination. *Molecular Cell, 65*(4), 743-750. e744.

Juszkiewicz, S., Slodkowicz, G., Lin, Z., Freire-Pritchett, P., Peak-Chew, S.-Y., & Hegde, R. S. (2020, 2020/07/13). Ribosome collisions trigger cis-acting feedback inhibition of translation initiation. *Elife, 9*, e60038. https://doi.org/10.7554/eLife.60038

Kirchner, S., Cai, Z., Rauscher, R., Kastelic, N., Anding, M., Czech, A., Kleizen, B., Ostedgaard, L. S., Braakman, I., Sheppard, D. N., & Ignatova, Z. (2017, May). Alteration of protein function by a silent polymorphism linked to tRNA

abundance. *PLoS Biol, 15*(5), e2000779. https://doi.org/10.1371/journal.pbio.2000779

Konigsberg, W., & Godson, G. N. (1983, Feb). Evidence for use of rare codons in the dnaG gene and other regulatory genes of Escherichia coli. *Proc Natl Acad Sci U S A, 80*(3), 687-691. https://doi.org/10.1073/pnas.80.3.687

Kramer, E. B., & Farabaugh, P. J. (2007, Jan). The frequency of translational misreading errors in E. coli is largely determined by tRNA competition. *Rna, 13*(1), 87-96. https://doi.org/10.1261/rna.294907

Kryuchkova, P., Grishin, A., Eliseev, B., Karyagina, A., Frolova, L., & Alkalaeva, E. (2013, Apr). Two-step model of stop codon recognition by eukaryotic release factor eRF1. *Nucleic Acids Res, 41*(8), 4573-4586. https://doi.org/10.1093/nar/gkt113

Kudla, G., Lipinski, L., Caffin, F., Helwak, A., & Zylicz, M. (2006). High Guanine and Cytosine Content Increases mRNA Levels in Mammalian Cells. *PLOS Biology, 4*(6), e180. https://doi.org/10.1371/journal.pbio.0040180

Kurland, C. G. (1991, Jul 22). Codon bias and gene expression. *FEBS Lett, 285*(2), 165-169. https://doi.org/10.1016/0014-5793(91)80797-7

Kutach, A. K., & Kadonaga, J. T. (2000, Jul). The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. *Mol Cell Biol, 20*(13), 4754-4764. https://doi.org/10.1128/mcb.20.13.4754-4764.2000

Kwon, K. C., Chan, H. T., Leon, I. R., Williams-Carrier, R., Barkan, A., & Daniell, H. (2016, Sep). Codon Optimization to Enhance Expression Yields Insights into Chloroplast Translation. *Plant Physiol, 172*(1), 62-77. https://doi.org/10.1104/pp.16.00981

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., & Jenuwein, T. (2001, Mar 1). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature, 410*(6824), 116-120. https://doi.org/10.1038/35065132

Lampson, B. L., Pershing, N. L., Prinz, J. A., Lacsina, J. R., Marzluff, W. F., Nicchitta, C. V., MacAlpine, D. M., & Counter, C. M. (2013, Jan 7). Rare codons regulate KRas oncogenesis. *Curr Biol, 23*(1), 70-75. https://doi.org/10.1016/j.cub.2012.11.031

LaRiviere, F. J., Cole, S. E., Ferullo, D. J., & Moore, M. J. (2006). A late-acting quality control process for mature eukaryotic rRNAs. *Molecular Cell, 24*(4), 619-626.

Lavner, Y., & Kotlar, D. (2005, Jan 17). Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene, 345*(1), 127-138. https://doi.org/10.1016/j.gene.2004.11.035

Le Hir, H., Nott, A., & Moore, M. J. (2003, Apr). How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci, 28*(4), 215-220. https://doi.org/10.1016/s0968-0004(03)00052-5

Lee, H. H., Kim, Y. S., Kim, K. H., Heo, I., Kim, S. K., Kim, O., Kim, H. K., Yoon, J. Y., Kim, H. S., Kim, D. J., Lee, S. J., Yoon, H. J., Kim, S. J., Lee, B. G., Song, H. K., Kim, V. N., Park, C. M., & Suh, S. W. (2007, Sep 21). Structural and functional insights into Dom34, a key component of no-go mRNA decay. *Mol Cell, 27*(6), 938-950. https://doi.org/10.1016/j.molcel.2007.07.019

Lemaire, S., Fontrodona, N., Aubé, F., Claude, J.-B., Polvèche, H., Modolo, L., Bourgeois, C. F., Mortreux, F., & Auboeuf, D. (2019, 2019/11/29). Characterizing the interplay between gene nucleotide composition bias and splicing. *Genome Biology, 20*(1), 259. https://doi.org/10.1186/s13059-019-1869-y

Lemon, B., & Tjian, R. (2000, Oct 15). Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev, 14*(20), 2551-2569. https://doi.org/10.1101/gad.831000

Letzring, D. P., Dean, K. M., & Grayhack, E. J. (2010, Dec). Control of translation efficiency in yeast by codon-anticodon interactions. *Rna, 16*(12), 2516-2528. https://doi.org/10.1261/rna.2411710

Letzring, D. P., Wolf, A. S., Brule, C. E., & Grayhack, E. J. (2013, Sep). Translation of CGA codon repeats in yeast involves quality control components and ribosomal protein L1. *Rna, 19*(9), 1208-1217. https://doi.org/10.1261/rna.039446.113

Li, G. W., Burkhardt, D., Gross, C., & Weissman, J. S. (2014, Apr 24). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell, 157*(3), 624-635. https://doi.org/10.1016/j.cell.2014.02.033

Lin, Y., May, G. E., Kready, H., Nazzaro, L., Mao, M., Spealman, P., Creeger, Y., & McManus, C. J. (2019, Sep 26). Impacts of uORF codon identity and position on translation regulation. *Nucleic Acids Res, 47*(17), 9358-9367. https://doi.org/10.1093/nar/gkz681

Lindhout, P., Neeleman, L., Van Tol, H., & Van Vloten-Doting, L. (1985, Nov 4). Ribosomes are stalled during in vitro translation of alfalfa mosaic virus RNA 1. *Eur J Biochem, 152*(3), 625-631.

Lithwick, G., & Margalit, H. (2003, Dec). Hierarchy of sequence-dependent features associated with prokaryotic translation. *Genome Res, 13*(12), 2665-2673. https://doi.org/10.1101/gr.1485203

Louder, R. K., He, Y., López-Blanco, J. R., Fang, J., Chacón, P., & Nogales, E. (2016). Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature, 531*(7596), 604-609.

Lu, J., & Deutsch, C. (2008, Dec 5). Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J Mol Biol, 384*(1), 73-86. https://doi.org/10.1016/j.jmb.2008.08.089

Lubas, M., Andersen, P. R., Schein, A., Dziembowski, A., Kudla, G., & Jensen, T. H. (2015, Jan 13). The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep, 10*(2), 178-192. https://doi.org/10.1016/j.celrep.2014.12.026

Lubas, M., Christensen, M. S., Kristiansen, M. S., Domanski, M., Falkenby, L. G., Lykke-Andersen, S., Andersen, J. S., Dziembowski, A., & Jensen, T. H. (2011, Aug 19). Interaction profiling identifies the human nuclear exosome targeting complex. *Mol Cell, 43*(4), 624-637. https://doi.org/10.1016/j.molcel.2011.06.028

Lyu, X., Yang, Q., Li, L., Dang, Y., Zhou, Z., Chen, S., & Liu, Y. (2020, Jun). Adaptation of codon usage to tRNA I34 modification controls translation kinetics and proteome landscape. *PLoS Genet, 16*(6), e1008836. https://doi.org/10.1371/journal.pgen.1008836

Lyumkis, D., Oliveira dos Passos, D., Tahara, E. B., Webb, K., Bennett, E. J., Vinterbo, S., Potter, C. S., Carragher, B., & Joazeiro, C. A. (2014, Nov 11). Structural basis for translational surveillance by the large ribosomal subunit-associated protein quality control complex. *Proc Natl Acad Sci U S A, 111*(45), 15981-15986. https://doi.org/10.1073/pnas.1413882111

Man, O., & Pilpel, Y. (2007, Mar). Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species. *Nat Genet, 39*(3), 415-421. https://doi.org/10.1038/ng1967

Mao, Y., Liu, H., Liu, Y., & Tao, S. (2014, Apr). Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in Saccharomyces cerevisiae. *Nucleic Acids Res, 42*(8), 4813-4822. https://doi.org/10.1093/nar/gku159

Margueron, R., & Reinberg, D. (2011, Jan 20). The Polycomb complex PRC2 and its mark in life. *Nature, 469*(7330), 343-349. https://doi.org/10.1038/nature09784

Matsuo, Y., Ikeuchi, K., Saeki, Y., Iwasaki, S., Schmidt, C., Udagawa, T., Sato, F., Tsuchiya, H., Becker, T., & Tanaka, K. (2017). Ubiquitination of stalled ribosome triggers ribosome-associated quality control. *Nature Communications, 8*(1), 1-14.

Meola, N., Domanski, M., Karadoulama, E., Chen, Y., Gentil, C., Pultz, D., Vitting-Seerup, K., Lykke-Andersen, S., Andersen, J. S., & Sandelin, A. (2016). Identification of a nuclear exosome decay pathway for processed transcripts. *Molecular Cell, 64*(3), 520-533.

Meola, N., Domanski, M., Karadoulama, E., Chen, Y., Gentil, C., Pultz, D., Vitting-Seerup, K., Lykke-Andersen, S., Andersen, J. S., Sandelin, A., & Jensen, T. H. (2016, Nov 3). Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. *Mol Cell, 64*(3), 520-533. https://doi.org/10.1016/j.molcel.2016.09.025

Meydan, S., & Guydosh, N. R. (2020, 2020/08/20/). Disome and Trisome Profiling Reveal Genome-wide Targets of Ribosome Quality Control. *Molecular Cell, 79*(4), 588-602.e586. https://doi.org/https://doi.org/10.1016/j.molcel.2020.06.010

Mitarai, N., Sneppen, K., & Pedersen, S. (2008, Sep 26). Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization. *J Mol Biol, 382*(1), 236-245. https://doi.org/10.1016/j.jmb.2008.06.068

Mordstein, C., Savisaar, R., Young, R. S., Bazile, J., Talmane, L., Luft, J., Liss, M., Taylor, M. S., Hurst, L. D., & Kudla, G. (2020, 2020/04/22/). Codon Usage and Splicing Jointly Influence mRNA Localization. *Cell Systems, 10*(4), 351-362.e358. https://doi.org/https://doi.org/10.1016/j.cels.2020.03.001

Morey, L., & Helin, K. (2010, Jun). Polycomb group protein-mediated repression of transcription. *Trends Biochem Sci, 35*(6), 323-332. https://doi.org/10.1016/j.tibs.2010.02.009

Moriyama, E. N., & Powell, J. R. (1997, Nov). Codon usage bias and tRNA abundance in Drosophila. *J Mol Evol, 45*(5), 514-523.

Motamedi, M. R., Hong, E. J., Li, X., Gerber, S., Denison, C., Gygi, S., & Moazed, D. (2008, Dec 26). HP1 proteins form distinct complexes and mediate heterochromatic gene silencing by nonoverlapping mechanisms. *Mol Cell, 32*(6), 778-790. https://doi.org/10.1016/j.molcel.2008.10.026

Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D., & Grewal, S. I. (2001, Apr 6). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science, 292*(5514), 110-113. https://doi.org/10.1126/science.1060118

Nestorov, P., Tardat, M., & Peters, A. H. (2013). H3K9/HP1 and Polycomb: two key epigenetic silencing pathways for gene regulation and embryo development. *Curr Top Dev Biol, 104*, 243-291. https://doi.org/10.1016/b978-0-12-416027-9.00008-5

Newman, Z. R., Young, J. M., Ingolia, N. T., & Barton, G. M. (2016). Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9. *Proceedings of the National Academy of Sciences, 113*(10), E1362-E1371. https://doi.org/10.1073/pnas.1518976113

Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F., & O'Neal, C. (1965, May). RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proc Natl Acad Sci U S A, 53*(5), 1161-1168. https://doi.org/10.1073/pnas.53.5.1161

O'Sullivan, C., Christie, J., Pienaar, M., Gambling, J., Nickerson, P. E., Alford, S. C., Chow, R. L., & Howard, P. L. (2015, Nov). Mutagenesis of ARS2 Domains To Assess Possible Roles in Cell Cycle Progression and MicroRNA and Replication-Dependent Histone mRNA Biogenesis. *Mol Cell Biol, 35*(21), 3753-3767. https://doi.org/10.1128/mcb.00272-15

Ohno, M., Segref, A., Bachi, A., Wilm, M., & Mattaj, I. W. (2000, Apr 14). PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell, 101*(2), 187-198. https://doi.org/10.1016/s0092-8674(00)80829-6

Oler, A. J., Alla, R. K., Roberts, D. N., Wong, A., Hollenhorst, P. C., Chandler, K. J., Cassiday, P. A., Nelson, C. A., Hagedorn, C. H., Graves, B. J., & Cairns, B. R. (2010, 2010/05/01). Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. *Nature Structural & Molecular Biology, 17*(5), 620-628. https://doi.org/10.1038/nsmb.1801

Pavlov, M. Y., Watts, R. E., Tan, Z., Cornish, V. W., Ehrenberg, M., & Forster, A. C. (2009, Jan 6). Slow peptide bond formation by proline and other N-alkylamino acids in translation. *Proc Natl Acad Sci U S A, 106*(1), 50-54. https://doi.org/10.1073/pnas.0809211106

Pechmann, S., Chartron, J. W., & Frydman, J. (2014, Dec). Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. *Nat Struct Mol Biol, 21*(12), 1100-1105. https://doi.org/10.1038/nsmb.2919

Pedersen, S. (1984, Dec 1). Escherichia coli ribosomes translate in vivo with variable rate. *Embo j, 3*(12), 2895-2898.

Pisarev, A. V., Skabkin, M. A., Pisareva, V. P., Skabkina, O. V., Rakotondrafara, A. M., Hentze, M. W., Hellen, C. U., & Pestova, T. V. (2010, Jan 29). The role of ABCE1 in eukaryotic posttermination ribosomal recycling. *Mol Cell, 37*(2), 196-210. https://doi.org/10.1016/j.molcel.2009.12.034

Pisareva, V. P., Skabkin, M. A., Hellen, C. U., Pestova, T. V., & Pisarev, A. V. (2011, May 4). Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. *Embo j, 30*(9), 1804-1817. https://doi.org/10.1038/emboj.2011.93

Plotkin, J. B., & Kudla, G. (2011, 2011/01/01). Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics, 12*(1), 32-42. https://doi.org/10.1038/nrg2899

Pop, C., Rouskin, S., Ingolia, N. T., Han, L., Phizicky, E. M., Weissman, J. S., & Koller, D. (2014, Dec 23). Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol Syst Biol, 10*(12), 770. https://doi.org/10.15252/msb.20145524

Pop, C., Rouskin, S., Ingolia, N. T., Han, L., Phizicky, E. M., Weissman, J. S., & Koller, D. (2014). Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol Syst Biol, 10*(12), 770.

Pouyet, F., Mouchiroud, D., Duret, L., & Sémon, M. (2017, 2017/08/15). Recombination, meiotic expression and human codon usage. *Elife, 6*, e27344. https://doi.org/10.7554/eLife.27344

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M. S., Mapendano, C. K., Schierup, M. H., & Jensen, T. H. (2008). RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters. *Science, 322*(5909), 1851-1854. https://doi.org/10.1126/science.1164096

Presnyak, V., Alhusaini, N., Chen, Y. H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K. E., Graveley, B. R., & Coller, J. (2015, Mar 12). Codon optimality is a major determinant of mRNA stability. *Cell, 160*(6), 1111-1124. https://doi.org/10.1016/j.cell.2015.02.029

Qian, W., Yang, J. R., Pearson, N. M., Maclean, C., & Zhang, J. (2012). Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet, 8*(3), e1002603. https://doi.org/10.1371/journal.pgen.1002603

Radhakrishnan, A., Chen, Y. H., Martin, S., Alhusaini, N., Green, R., & Coller, J. (2016, Sep 22). The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation

by Monitoring Codon Optimality. *Cell, 167*(1), 122-132.e129. https://doi.org/10.1016/j.cell.2016.08.053

Roh, T. Y., Ngau, W. C., Cui, K., Landsman, D., & Zhao, K. (2004, Aug). High-resolution genome-wide mapping of histone modifications. *Nat Biotechnol, 22*(8), 1013-1016. https://doi.org/10.1038/nbt990

Sabi, R., & Tuller, T. (2015). A comparative genomics study on the effect of individual amino acids on ribosome stalling. *BMC Genomics, 16 Suppl 10*, S5. https://doi.org/10.1186/1471-2164-16-s10-s5

Sabin, L. R., Zhou, R., Gruber, J. J., Lukinova, N., Bambina, S., Berman, A., Lau, C. K., Thompson, C. B., & Cherry, S. (2009, Jul 23). Ars2 regulates both miRNA- and siRNA- dependent silencing and suppresses RNA virus infection in Drosophila. *Cell, 138*(2), 340-351. https://doi.org/10.1016/j.cell.2009.04.045

Santos, D. A., Shi, L., Tu, B. P., & Weissman, J. S. (2019). Cycloheximide can distort measurements of mRNA levels and translation efficiency. *Nucleic Acids Research, 47*(10), 4974-4985. https://doi.org/10.1093/nar/gkz205

Sauna, Z. E., & Kimchi-Sarfaty, C. (2011, Aug 31). Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet, 12*(10), 683-691. https://doi.org/10.1038/nrg3051

Schübeler, D., MacAlpine, D. M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D. E., O'Neill, L. P., Turner, B. M., Delrow, J., Bell, S. P., & Groudine, M. (2004, Jun 1). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev, 18*(11), 1263-1271. https://doi.org/10.1101/gad.1198204

Schulze, W. M., & Cusack, S. (2017, Nov 3). Structural basis for mutually exclusive co-transcriptional nuclear cap-binding complexes with either NELF-E or ARS2. *Nat Commun, 8*(1), 1302. https://doi.org/10.1038/s41467-017-01402-w

Schulze, W. M., Stein, F., Rettel, M., Nanao, M., & Cusack, S. (2018, Apr 27). Structural analysis of human ARS2 as a platform for co-transcriptional RNA sorting. *Nat Commun, 9*(1), 1701. https://doi.org/10.1038/s41467-018-04142-7

Shah, K., Cheng, Y., Hahn, B., Bridges, R., Bradbury, N. A., & Mueller, D. M. (2015, Mar 27). Synonymous codon usage affects the expression of wild type and F508del CFTR. *J Mol Biol, 427*(6 Pt B), 1464-1479. https://doi.org/10.1016/j.jmb.2015.02.003

Shah, P., Ding, Y., Niemczyk, M., Kudla, G., & Plotkin, J. B. (2013, Jun 20). Rate-limiting steps in yeast protein translation. *Cell, 153*(7), 1589-1601. https://doi.org/10.1016/j.cell.2013.05.049

Shao, S., Brown, A., Santhanam, B., & Hegde, R. S. (2015, Feb 5). Structure and assembly pathway of the ribosome quality control complex. *Mol Cell, 57*(3), 433-444. https://doi.org/10.1016/j.molcel.2014.12.015

Shao, S., & Hegde, R. S. (2014, Sep 18). Reconstitution of a minimal ribosome-associated ubiquitination pathway with purified factors. *Mol Cell, 55*(6), 880-890. https://doi.org/10.1016/j.molcel.2014.07.006

Shao, S., von der Malsburg, K., & Hegde, R. S. (2013, Jun 6). Listerin-dependent nascent protein ubiquitination relies on ribosome subunit dissociation. *Mol Cell, 50*(5), 637-648. https://doi.org/10.1016/j.molcel.2013.04.015

Sharp, P. M., & Li, W.-H. (1987). The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research, 15*(3), 1281-1295.

Sharp, P. M., Tuohy, T. M., & Mosurski, K. R. (1986, Jul 11). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res, 14*(13), 5125-5143. https://doi.org/10.1093/nar/14.13.5125

Shcherbik, N., Chernova, T. A., Chernoff, Y. O., & Pestov, D. G. (2016, Aug 19). Distinct types of translation termination generate substrates for ribosome-associated quality control. *Nucleic Acids Res, 44*(14), 6840-6852. https://doi.org/10.1093/nar/gkw566

Shlyueva, D., Stampfel, G., & Stark, A. (2014, Apr). Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet, 15*(4), 272-286. https://doi.org/10.1038/nrg3682

Shoemaker, C. J., Eyler, D. E., & Green, R. (2010, Oct 15). Dom34:Hbs1 promotes subunit dissociation and peptidyl-tRNA drop-off to initiate no-go decay. *Science, 330*(6002), 369-372. https://doi.org/10.1126/science.1192430

Shoemaker, C. J., & Green, R. (2011, Dec 20). Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. *Proc Natl Acad Sci U S A, 108*(51), E1392-1398. https://doi.org/10.1073/pnas.1113956108

Silla, T., Karadoulama, E., Mąkosa, D., Lubas, M., & Jensen, T. H. (2018, May 15). The RNA Exosome Adaptor ZFC3H1 Functionally Competes with Nuclear Export

Activity to Retain Target Transcripts. *Cell Rep, 23*(7), 2199-2210. https://doi.org/10.1016/j.celrep.2018.04.061

Simms, C. L., Hudson, B. H., Mosior, J. W., Rangwala, A. S., & Zaher, H. S. (2014). An active role for the ribosome in determining the fate of oxidized mRNA. *Cell Reports, 9*(4), 1256-1264.

Simms, C. L., Yan, L. L., & Zaher, H. S. (2017, Oct 19). Ribosome Collision Is Critical for Quality Control during No-Go Decay. *Mol Cell, 68*(2), 361-373.e365. https://doi.org/10.1016/j.molcel.2017.08.019

Sinha, N. K., Ordureau, A., Best, K., Saba, J. A., Zinshteyn, B., Sundaramoorthy, E., Fulzele, A., Garshott, D. M., Denk, T., Thoms, M., Paulo, J. A., Harper, J. W., Bennett, E. J., Beckmann, R., & Green, R. (2020, 2020/08/03). EDF1 coordinates cellular responses to ribosome collisions. *Elife, 9*, e58828. https://doi.org/10.7554/eLife.58828

Skabkin, M. A., Skabkina, O. V., Dhote, V., Komar, A. A., Hellen, C. U., & Pestova, T. V. (2010, Aug 15). Activities of Ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling. *Genes Dev, 24*(16), 1787-1801. https://doi.org/10.1101/gad.1957510

Skabkin, M. A., Skabkina, O. V., Hellen, C. U., & Pestova, T. V. (2013, Jul 25). Reinitiation and other unconventional posttermination events during eukaryotic translation. *Mol Cell, 51*(2), 249-264. https://doi.org/10.1016/j.molcel.2013.05.026

Smale, S. T., & Baltimore, D. (1989, Apr 7). The "initiator" as a transcription control element. *Cell, 57*(1), 103-113. https://doi.org/10.1016/0092-8674(89)90176-1

Smale, S. T., & Kadonaga, J. T. (2003). The RNA Polymerase II Core Promoter. *Annual Review of Biochemistry, 72*(1), 449-479. https://doi.org/10.1146/annurev.biochem.72.121801.161520

Sonenberg, N., & Hinnebusch, A. G. (2009, Feb 20). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell, 136*(4), 731-745. https://doi.org/10.1016/j.cell.2009.01.042

Song, H., Mugnier, P., Das, A. K., Webb, H. M., Evans, D. R., Tuite, M. F., Hemmings, B. A., & Barford, D. (2000, Feb 4). The crystal structure of human eukaryotic release factor eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell, 100*(3), 311-321.

Sørensen, M. A., Kurland, C. G., & Pedersen, S. (1989, May 20). Codon usage determines translation rate in Escherichia coli. *J Mol Biol, 207*(2), 365-377. https://doi.org/10.1016/0022-2836(89)90260-x

Spencer, P. S., Siller, E., Anderson, J. F., & Barral, J. M. (2012, Sep 21). Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J Mol Biol, 422*(3), 328-335. https://doi.org/10.1016/j.jmb.2012.06.010

Speth, C., Szabo, E. X., Martinho, C., Collani, S., zur Oven-Krockhaus, S., Richter, S., Droste-Borel, I., Macek, B., Stierhof, Y.-D., Schmid, M., Liu, C., & Laubinger, S. (2018, 2018/08/28). Arabidopsis RNA processing factor SERRATE regulates the transcription of intronless genes. *Elife, 7*, e37078. https://doi.org/10.7554/eLife.37078

Stock, J. K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A. G., & Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nature cell biology, 9*(12), 1428-1435.

Subramaniam, A. R., Zid, B. M., & O'Shea, E. K. (2014, Nov 20). An integrated approach reveals regulatory controls on bacterial translation elongation. *Cell, 159*(5), 1200-1211. https://doi.org/10.1016/j.cell.2014.10.043

Sundaramoorthy, E., Leonard, M., Mak, R., Liao, J., Fulzele, A., & Bennett, E. J. (2017). ZNF598 and RACK1 regulate mammalian ribosome-associated quality control function by mediating regulatory 40S ribosomal ubiquitylation. *Molecular Cell, 65*(4), 751-760. e754.

Svidritskiy, E., Demo, G., & Korostelev, A. A. (2018, Aug 10). Mechanism of premature translation termination on a sense codon. *J Biol Chem, 293*(32), 12472-12479. https://doi.org/10.1074/jbc.AW118.003232

Tanner, D. R., Cariello, D. A., Woolstenhulme, C. J., Broadbent, M. A., & Buskirk, A. R. (2009, Dec 11). Genetic identification of nascent peptides that induce ribosome stalling. *J Biol Chem, 284*(50), 34809-34818. https://doi.org/10.1074/jbc.M109.039040

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., & Libri, D. (2006, Sep 15). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol Cell, 23*(6), 853-864. https://doi.org/10.1016/j.molcel.2006.07.029

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., & Pilpel, Y. (2010, Apr 16). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell, 141*(2), 344-354. https://doi.org/10.1016/j.cell.2010.03.031

Tuller, T., Waldman, Y. Y., Kupiec, M., & Ruppin, E. (2010). Translation efficiency is determined by both codon bias and folding energy. *Proceedings of the National Academy of Sciences, 107*(8), 3645-3650. https://doi.org/10.1073/pnas.0909910107

Ude, S., Lassak, J., Starosta, A. L., Kraxenberger, T., Wilson, D. N., & Jung, K. (2013). Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science, 339*(6115), 82-85.

Varenne, S., Buc, J., Lloubes, R., & Lazdunski, C. (1984, Dec 15). Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J Mol Biol, 180*(3), 549-576. https://doi.org/10.1016/0022-2836(84)90027-5

Vasiljeva, L., & Buratowski, S. (2006, Jan 20). Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell, 21*(2), 239-248. https://doi.org/10.1016/j.molcel.2005.11.028

Verma, R., Reichermeier, K. M., Burroughs, A. M., Oania, R. S., Reitsma, J. M., Aravind, L., & Deshaies, R. J. (2018, May). Vms1 and ANKZF1 peptidyl-tRNA hydrolases release nascent chains from stalled ribosomes. *Nature, 557*(7705), 446-451. https://doi.org/10.1038/s41586-018-0022-5

Vivanco-Dominguez, S., Bueno-Martinez, J., Leon-Avila, G., Iwakura, N., Kaji, A., Kaji, H., & Guarneros, G. (2012, Apr 13). Protein synthesis factors (RF1, RF2, RF3, RRF, and tmRNA) and peptidyl-tRNA hydrolase rescue stalled ribosomes at sense codons. *J Mol Biol, 417*(5), 425-439. https://doi.org/10.1016/j.jmb.2012.02.008

Vo Ngoc, L., Cassidy, C. J., Huang, C. Y., Duttke, S. H., & Kadonaga, J. T. (2017, Jan 1). The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. *Genes Dev, 31*(1), 6-11. https://doi.org/10.1101/gad.293837.116

Wada, M., & Ito, K. (2019, Feb). Misdecoding of rare CGA codon by translation termination factors, eRF1/eRF3, suggests novel class of ribosome rescue pathway in S. cerevisiae. *Febs j, 286*(4), 788-802. https://doi.org/10.1111/febs.14709

Wang, C., Han, B., Zhou, R., & Zhuang, X. (2016, May 5). Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells. *Cell, 165*(4), 990-1001. https://doi.org/10.1016/j.cell.2016.04.040

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., & Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature, 431*(7010), 873-878.

WASYLYK, B., & CHAMBON, P. (1979). Transcription by eukaryotic RNA polymerases A and B of chromatin assembled in vitro. *Eur J Biochem, 98*(2), 317-327.

Wasylyk, B., Thevenin, G., Oudet, P., & Chambon, P. (1979). Transcription of in vitro assembled chromatin by Escherichia coli RNA polymerase. *J Mol Biol, 128*(3), 411-440.

Weick, E.-M., Puno, M. R., Januszyk, K., Zinder, J. C., DiMattia, M. A., & Lima, C. D. (2018). Helicase-dependent RNA decay illuminated by a cryo-EM structure of a human nuclear RNA exosome-MTR4 complex. *Cell, 173*(7), 1663-1677. e1621.

Weinberg, D. E., Shah, P., Eichhorn, S. W., Hussmann, J. A., Plotkin, J. B., & Bartel, D. P. (2016, Feb 23). Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. *Cell Rep, 14*(7), 1787-1799. https://doi.org/10.1016/j.celrep.2016.01.043

Wolin, S. L., & Walter, P. (1988, Nov). Ribosome pausing and stacking during translation of a eukaryotic mRNA. *Embo j, 7*(11), 3559-3569.

Wong, L. E., Li, Y., Pillay, S., Frolova, L., & Pervushin, K. (2012, Jul). Selectivity of stop codon recognition in translation termination is modulated by multiple conformations of GTS loop in eRF1. *Nucleic Acids Res, 40*(12), 5751-5765. https://doi.org/10.1093/nar/gks192

Wu, C., Dasgupta, A., Shen, L., Bell-Pedersen, D., & Sachs, M. S. (2018, Mar 15). The cell free protein synthesis system from the model filamentous fungus Neurospora crassa. *Methods, 137*, 11-19. https://doi.org/10.1016/j.ymeth.2017.12.003

Wu, C. C.-C., Peterson, A., Zinshteyn, B., Regot, S., & Green, R. (2020). Ribosome collisions trigger general stress responses to regulate cell fate. *Cell, 182*(2), 404-416. e414.

Wu, C. C.-C., Zinshteyn, B., Wehner, K. A., & Green, R. (2019a). High-resolution ribosome profiling defines discrete ribosome elongation states and translational regulation during cellular stress. *Molecular Cell, 73*(5), 959-970. e955.

Wu, C. C.-C., Zinshteyn, B., Wehner, K. A., & Green, R. (2019b, 2019/03/07/). High-Resolution Ribosome Profiling Defines Discrete Ribosome Elongation States and Translational Regulation during Cellular Stress. *Molecular Cell, 73*(5), 959-970.e955. https://doi.org/https://doi.org/10.1016/j.molcel.2018.12.009

Wu, Q., Medina, S. G., Kushawah, G., DeVore, M. L., Castellano, L. A., Hand, J. M., Wright, M., & Bazzini, A. A. (2019, 2019/04/23). Translation affects mRNA stability in a codon-dependent manner in human cells. *Elife, 8*, e45396. https://doi.org/10.7554/eLife.45396

Xu, Y., Ma, P., Shah, P., Rokas, A., Liu, Y., & Johnson, C. H. (2013, 2013/03/01). Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. *Nature, 495*(7439), 116-120. https://doi.org/10.1038/nature11942

Yan, X., Hoek, T. A., Vale, R. D., & Tanenbaum, M. E. (2016, May 5). Dynamics of Translation of Single mRNA Molecules In Vivo. *Cell, 165*(4), 976-989. https://doi.org/10.1016/j.cell.2016.04.034

Yu, C. H., Dang, Y., Zhou, Z., Wu, C., Zhao, F., Sachs, M. S., & Liu, Y. (2015, Sep 3). Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Mol Cell, 59*(5), 744-754. https://doi.org/10.1016/j.molcel.2015.07.018

Zabidi, M. A., & Stark, A. (2016, Dec). Regulatory Enhancer-Core-Promoter Communication via Transcription Factors and Cofactors. *Trends Genet, 32*(12), 801-814. https://doi.org/10.1016/j.tig.2016.10.003

Zhang, S. P., Zubay, G., & Goldman, E. (1991, Aug 30). Low-usage codons in Escherichia coli, yeast, fruit fly and primates. *Gene, 105*(1), 61-72. https://doi.org/10.1016/0378-1119(91)90514-c

Zhao, F., Yu, C. H., & Liu, Y. (2017, Aug 21). Codon usage regulates protein structure and function by affecting translation elongation speed in Drosophila cells. *Nucleic Acids Res, 45*(14), 8484-8492. https://doi.org/10.1093/nar/gkx501

Zhou, M., Guo, J., Cha, J., Chae, M., Chen, S., Barral, J. M., Sachs, M. S., & Liu, Y. (2013, Mar 7). Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature, 495*(7439), 111-115. https://doi.org/10.1038/nature11833

Zhou, Z., Dang, Y., Zhou, M., Li, L., Yu, C. H., Fu, J., Chen, S., & Liu, Y. (2016, Oct 11). Codon usage is an important determinant of gene expression levels largely

through its effects on transcription. *Proc Natl Acad Sci U S A, 113*(41), E6117-e6125. https://doi.org/10.1073/pnas.1606724113

Zhou, Z., Dang, Y., Zhou, M., Yuan, H., & Liu, Y. (2018, Mar 16). Codon usage biases co-evolve with transcription termination machinery to suppress premature cleavage and polyadenylation. *Elife, 7*. https://doi.org/10.7554/eLife.33569

Zurita Rendon, O., Fredrickson, E. K., Howard, C. J., Van Vranken, J., Fogarty, S., Tolley, N. D., Kalia, R., Osuna, B. A., Shen, P. S., Hill, C. P., Frost, A., & Rutter, J. (2018, Jun 6). Vms1p is a release factor for the ribosome-associated quality control complex. *Nat Commun, 9*(1), 2197. https://doi.org/10.1038/s41467-018-04564-3