Title

Genome-wide role of codon usage on transcription and identification of potential regulators

Permalink

https://escholarship.org/uc/item/0t2365r3

Journal

Proceedings of the National Academy of Sciences of the United States of America, 118(6)

ISSN

0027-8424

Authors

Zhao, Fangzhou
Zhou, Zhipeng
Dang, Yunkun
et al.

Publication Date

2021-02-09

DOI

10.1073/pnas.2022590118

Peer reviewed

Genome-wide role of codon usage on transcription and identification of potential regulators

Fangzhou Zhao[a], Zhipeng Zhou[a,b], Yunkun Dang[c], Hyunsoo Na[d], Catherine Adam[d], Anna Lipzen[d], Vivian Ng[d], Igor V. Grigoriev[d,e], and Yi Liu[a,1]

[a]Department of Physiology, University of Texas Southwestern Medical Center, Dallas, TX 75390; [b]State Key Laboratory of Agricultural Microbiology, College of Plant Science and Technology, Huazhong Agricultural University, 430070 Wuhan, Hubei, China; [c]State Key Laboratory for Conservation and Utilization of Bio-Resources and Center for Life Science, School of Life Sciences, Yunnan University, Kunming, Yunnan 650091, China; [d]US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; and [e]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720

Edited by Jay C. Dunlap, Geisel School of Medicine at Dartmouth, Hanover, NH, and approved December 18, 2020 (received for review November 12, 2020)

Codon usage bias is a fundamental feature of all genomes and plays an important role in determining gene expression levels. The codon usage was thought to influence gene expression mainly due to its impact on translation. Recently, however, codon usage was shown to affect transcription of fungal and mammalian genes, indicating the existence of a gene regulatory phenomenon with unknown mechanism. In *Neurospora*, codon usage biases strongly correlate with mRNA levels genome-wide, and here we show that the correlation between codon usage and RNA levels is maintained in the nucleus. In addition, codon optimality is tightly correlated with both total and nuclear RNA levels, suggesting that codon usage broadly influences mRNA levels through transcription in a translation-independent manner. A large-scale RNA sequencing-based genetic screen in *Neurospora* identified 18 candidate factors that when deleted decreased the genome-wide correlation between codon usage and RNA levels and reduced the codon usage effect on gene expression. Most of these factors, such as the H3K36 methyltransferase, are chromatin regulators or transcription factors. Together, our results suggest that the transcriptional effect of codon usage is mediated by multiple transcriptional regulatory mechanisms.

codon usage | transcription | *Neurospora* | translation | H3K36 methyltransferase

Codon usage bias, the preference for certain synonymous codons for almost all amino acids, is a feature of all eukaryotic and prokaryotic genomes analyzed (1–5). Studies in both eukaryotes and prokaryotes indicate that codon usage is an important determinant of gene expression levels (6–10), and codon optimization frequently enhances heterologous gene expression (10–12). The role of codon usage in regulating gene expression was thought to be due mostly to its impact on translation (4, 13–15). Supporting this notion, codon usage has been shown to play an important role in determining translation elongation rate in both fungi and animal cells (16–18). Rare codons cause ribosome stalling, resulting in premature translation termination and reduced translation efficiency (9, 19, 20). In addition to the regulation of translation kinetics, codon usage can also affect mRNA stability (17, 21–23). Thus, codon usage can impact gene expression at both translational and posttranscriptional levels.

The genome of the filamentous fungus *Neurospora crassa* has a strong codon usage bias for C/G at the wobble positions and has been used as a model system to study codon usage bias (24, 25). We previously showed that in *Neurospora* preferred codons increase the rate of translation elongation, whereas rare codons slow translation elongation, resulting in ribosome stalling and premature termination (16, 19). The codon usage in *Neurospora* genes strongly correlates with both protein and mRNA levels genome-wide, however (10), indicating that the effect of codon usage on protein expression is mostly due to its effect on mRNA levels. Using several *Neurospora* genes as reporters, we showed, surprisingly, that the impact of codon usage on gene expression is mainly due to its effects on transcription and is largely independent of translation (10). Furthermore, we showed that rare codons can also impact mRNA levels by causing premature transcription termination (26). A role of codon usage in transcription is further supported by studies in mammalian cells that show that codon usage or GC content (which correlates with codon usage) within gene coding regions influence mRNA synthesis efficiency (27–29). Together, these results suggest that codon usage biases are results of adaptation of protein coding sequences to both transcription and translation machineries. The conserved and robust codon usage-mediated transcriptional effect is surprising given that transcriptional control is thought to be mostly mediated by gene promoters. Thus, such an effect of codon usage represents a gene regulatory phenomenon with its mechanism unknown. In addition, although the transcriptional effects of codon usage were demonstrated for some genes in fungi and mammalian cells, it is not clear whether this is a phenomenon that broadly affects gene expression.

In this study, we examined the mechanism of how codon usage regulates gene transcription. By sequencing nuclear RNA, we showed that there is a strong genome-wide correlation between codon usage and RNA levels in the nucleus, suggesting that codon usage bias has a broad effect on gene transcription independent of translation. We carried out a large-scale RNA-sequencing (RNA-seq)-based genetic screen and identified 18 genes involved in mediating the codon usage effect on gene

Significance

Codon usage bias, the preference for certain synonymous codons, is a feature of all genomes and plays an important role in determining gene expression levels. Surprisingly, in addition to its role in mRNA translation, codon usage was recently shown to regulate gene expression at the transcriptional level. In this study, our results suggest that codon usage influences transcription genome-wide independently of its effect on translation. By carrying out a large-scale genetic screen, we identified 18 factors participating in multiple transcriptional regulatory mechanisms involved in the codon usage-mediated transcriptional effects. Together, our study establishes a foundation for future mechanistic understanding of this gene regulatory phenomenon.

Author contributions: F.Z. and Y.L. designed research; F.Z., Z.Z., Y.D., H.N., C.A., A.L., V.N., and I.V.G. performed research; F.Z. and Z.Z. contributed new reagents/analytic tools; F.Z., Y.D., A.L., and Y.L. analyzed data; and F.Z. and Y.L. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: Yi.Liu@UTsouthwestern.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2022590118/-/DCSupplemental.

Published February 1, 2021.

expression. Given the identities of these factors, it appears that the effect of codon usage on transcription is mediated by multiple chromatin and transcriptional regulators.

Results

Codon Usage Correlates with mRNA Transcription Genome-Wide. A strong positive correlation was previously observed between codon usage bias and mRNA levels genome-wide from mRNA-seq results in *Neurospora* (10). Although experiments using reporter genes in fungi and mammalian cells suggest that codon usage plays an important role in regulating gene expression at the transcriptional level (10, 20), the genome-wide correlation between codon usage bias and mRNA levels could be indirect, due to the role of codon usage on translation-dependent processes such as mRNA degradation (21, 23). To determine whether the role of codon usage in regulating transcription is a global phenomenon independent of translation, we performed nuclear RNA sequencing using a wild-type *Neurospora* strain, which eliminated potential translation-dependent effects and better reflects genome-wide transcription than total mRNA-seq. The purity of our nuclear preparation was confirmed by the absence of cytosolic protein tubulin and enrichment of histone H3 (*SI Appendix*, Fig. S1). As expected, the nuclear transcripts showed significant intron retention while the total mRNA-seq transcripts were almost entirely devoid of intron-containing transcripts. We used two different methods to sequence nuclear RNA: normal RNA-seq and poly(A)-tail-primed sequencing (2P-seq), which sequences the 3′ ends of poly(A)-containing RNAs. Consistent with the previous study, both sequencing methods for total RNA revealed strong positive correlations between RNA levels and gene codon bias index (CBI) (Fig. 1 *A* and *B*, *Left*). A CBI of 1 indicates that a gene has extreme optimal codon bias, and a value of 0 indicates completely random codon usage (7, 30). For nuclear RNA, similar positive correlation coefficients were observed for both sequencing methods (Fig. 1 *A* and *B*, *Right*). These results suggest that the genome-wide positive correlation between codon usage bias and

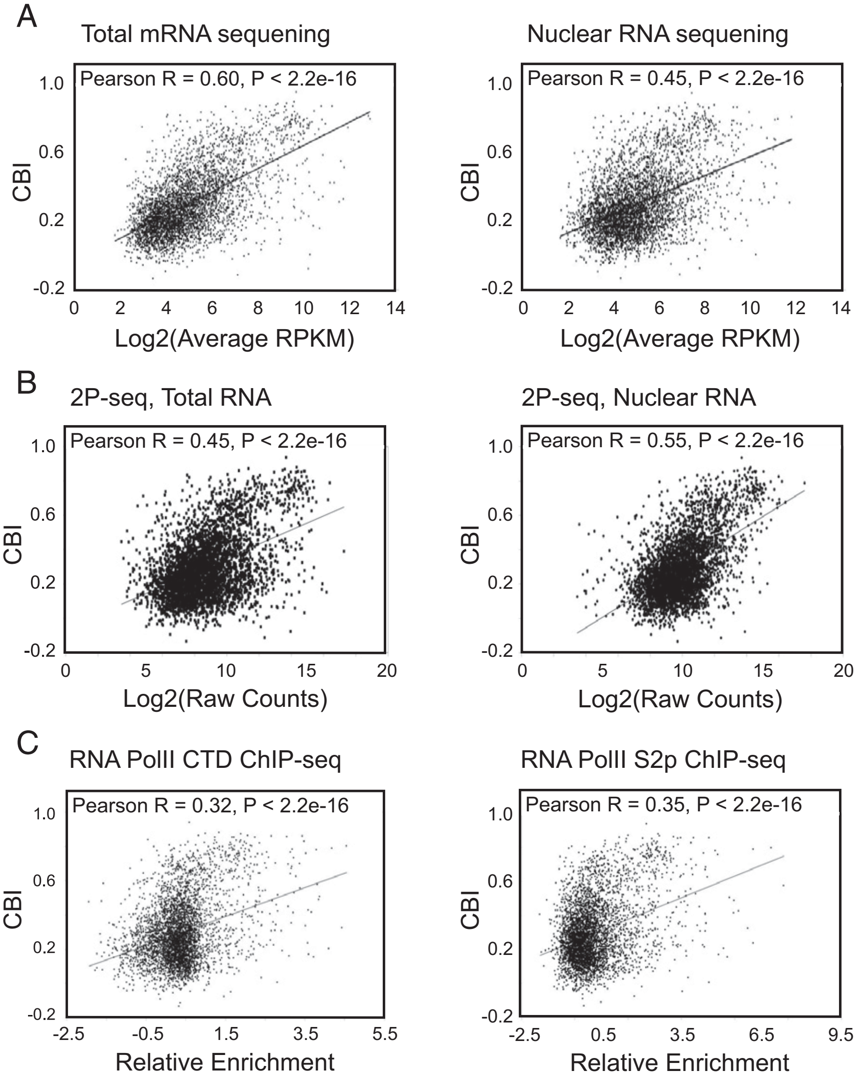


Fig. 1. Genome-wide correlation between codon usage bias and nuclear RNA levels. (*A*) Pearson correlation between CBI and total RNA levels (*Left*) or nuclear RNA levels (*Right*). Correlation coefficients and *P* values are given at the *Top*. RNA levels were measured by RNA sequencing of the wild-type strain FGSC4200. Log$_2$(average RPKM) values given are averages for each gene from two replicate samples. (*B*) Pearson correlation between CBI and poly(A)-selected total (*Left*) and nuclear (*Right*) RNA levels determined using 2P-seq of a wild-type strain 87-3 (*bd, a*) (52). Log$_2$(raw counts) from representative analyses of two samples are shown. (*C*) Pearson correlations between CBI and enrichments from ChIP experiments on wild-type strain (87-3) using an antibody to the nonphosphorylated Pol II CTD (*Left*) and to the Ser-2-phosphorylated CTD (*Right*).

mRNA levels is largely due to an effect of codon usage bias on transcription rather than translation.

To further confirm the correlation between codon usage and transcription, we analyzed data from chromatin immunoprecipitation sequencing (ChIP-seq) experiments using RNA polymerase II (Pol II) antibodies that recognize either the nonphosphorylated or the Ser-2-phosphorylated carboxyl-terminal domain (CTD). The Pol II CTD ChIP reflects the Pol II recruitment at an individual gene locus while the Pol II Ser-2-phosphorylated form reflects an elongating Pol II level (31). The enrichment level of either form of Pol II positively correlates with gene transcription level. As expected, a positive correlation between CBI and Pol II enrichment was seen in both ChIP-seq experiments, suggesting that genes enriched for preferred codons are associated with higher transcription levels (Fig. 1*C*). The reduced correlation coefficients compared to the RNA-seq experiments are likely due to sensitivity of the ChIP experiments which are limited by the specificity and pull-down efficiency of the Pol II antibodies.

We also determined the genome-wide Pearson correlations in *Saccharomyces cerevisiae* between codon usage biases and previously computed mRNA synthesis rates from various mRNA half-life and global run-on sequencing (GRO-seq) experiments (32). Consistent with the previous study (32), a modest-to-strong positive correlation was found between mRNA synthesis rates and gene tRNA adaptation index (tAI_g) (*Materials and Methods*) in these experimental data (*SI Appendix*, Fig. S2), indicating that the effect of codon usage on genome-wide transcription is conserved in fungi. Importantly, the correlation between mRNA synthesis rates and tAI_g is much stronger than that between mRNA half-life and tAI_g in all except for one of these results (*SI Appendix*, Fig. S2), suggesting that the general impact of codon usage on transcription is stronger than its impact on mRNA half-life in *S. cerevisiae*.

Codon Optimality Tightly Associates with Total and Nuclear RNA Levels. To determine how codon optimality is associated with genome-wide mRNA levels, we determined the Pearson correlation coefficients between individual codon frequencies (except for stop codons) in all genes and their mRNA levels obtained from mRNA-seq of total RNA. Individual codons were separated into optimal (most preferred), intermediate, and rare (least preferred) groups based on their genome-wide usage frequencies within corresponding codon families. Remarkably, we found that all optimal codons showed positive correlations with genome-wide mRNA levels, whereas all rare codons exhibited negative correlations (Fig. 2*A*). In addition, certain optimal codons (e.g., GTC) and rare codons (e.g., GAA) showed much stronger positive or negative correlations, respectively, than others. Furthermore, some synonymous codons are on opposite ends of positive or negative correlation rankings. For example, of the two codons for lysine, AAG is the codon with the second highest positive correlation ($R = 0.36$), and AAA is the codon with one of the highest negative correlations ($R = -0.39$). This indicates that the correlation between codon usage and mRNA levels is due to codon optimality and not amino acid usage.

When the same analysis was carried out for the nuclear RNA-seq results, again almost all optimal codons exhibited positive correlations with nuclear RNA levels, and all rare codons exhibited negative correlations (Fig. 2*B*). In addition, the Pearson correlation coefficient values of individual codons were extremely highly correlated ($R = 0.96$) between total mRNA and nuclear RNA-seq analyses (Fig. 2*C*). For both total and nuclear RNA, the optimal GTC codon had the highest positive correlations with RNA levels; rare codons GAA and AAA are among those with the highest negative correlations. These results suggest that codon optimality within open reading frames plays an important role in determining gene transcription levels independent of translation.

An RNA-Seq-Based Screen Identifies Factors that Mediate the Codon Usage Effect on RNA Levels. mRNA-seq of the wild-type *Neurospora* strain FGSC4200 in 12 experiments showed that the Pearson correlation coefficient values between CBI and mRNA levels only varied within a small range (*SI Appendix*, Fig. S3), suggesting that the genome-wide correlation value can be used to reflect the impact of codon usage on mRNA levels. To understand the mechanism of how codon usage influences transcription and mRNA levels, we carried out a genetic screen by performing RNA-seq of 206 *Neurospora* gene knockout (KO) strains to identify strains with significantly impaired correlations (Dataset S1). The genes evaluated include predicted transcription factors, chromatin remodeling factors, histone modifiers, histone variants, nucleosome regulators, cell cycle regulators, protein phosphatases, RNA binding proteins, and proteins likely involved in RNA decay pathways (Fig. 3*A* and Dataset S1). RNA-seq was performed in triplicate for each strain (Dataset S2). For each strain, we compared the correlation coefficient between mRNA level and CBI to that of the wild-type (WT) strain (FGSC4200). Of the 206 strains evaluated, 18 had significantly reduced correlation coefficients between CBI and mRNA levels (Fig. 3 *B* and *C* and *SI Appendix*, Fig. S4). These 18 genes include 10 predicted transcription factors (*fkh1*, *ada-2*, *ada-3*, *ada-6*, *fl*, *col-24*, *vad-3*, *kal-1*, *msn-1*, and NCU03043), three potential chromatin remodeling factors (NCU07975, NCU04445, and NCU04424), two histone deacetylases (*nst5* and *nst7*), two histone methyltransferases (*set-1* and *set-2*), and one RNA binding protein (*scp160*). Thus, the vast majority (17/18) of these factors are known or predicted chromatin/transcription regulatory factors.

In budding yeast, Dhh1 and Dbp2 influence RNA stability in a codon usage and translation-dependent manner (33, 34). The *Neurospora dhh1* and *dbp2* knockout strains, however, did not show any reduction in the correlation between codon usage and mRNA levels (Fig. 3*D*), indicating that either these two genes do not play a significant role in determining the codon usage effect on mRNA levels in *Neurospora* or that their functions on regulating mRNA levels in a codon usage-dependent manner is not conserved in *Neurospora*. This is also consistent with our findings that the correlation between mRNA level and CBI is largely independent of translation in *Neurospora*.

Because the knockout of many of the genes resulted in growth defects, we examined the relationship between the CBI versus mRNA level correlation and strain growth rate in 40 selected knockout strains. The growth rates of these strains were determined by race tube assays. Despite the wide variation of growth rates in these strains, there was no clear correlation between growth rates and codon-mRNA correlations (Fig. 3*E*), suggesting that the impaired CBI versus mRNA level correlations in the candidate strains were not simply caused by growth defects.

Codon Usage-Dependent mRNA Level Changes Confirmed in Mutants. If a candidate gene expresses a protein involved in mediating the codon usage effect on mRNA levels, its knockout should result in mRNA changes in a codon usage-dependent manner. To confirm this, we identified the up- or down-regulated genes in each of the 18 candidate knockout strains compared to the wild-type strain. In all strains, the down-regulated genes were significantly enriched for genes with higher CBI and the up-regulated were significantly enriched for genes with lower CBI (Fig. 4*A* and *SI Appendix*, Fig. S5). In contrast, although there were many differentially regulated genes in the $dhh1^{KO}$ and $hda\text{-}2^{KO}$ strains relative to the wild-type strain, the differentially regulated genes had no differences in CBI compared to the unchanged genes (Fig. 4*B*).

The $set\text{-}2^{KO}$ strain had one of the largest decreases in correlation between CBI and RNA levels compared to the wild-type strain of the strains examined: The Pearson *R* value was ∼0.42, whereas that of the wild-type strain was ∼0.58. *set-2* encodes the histone methyltransferase that methylates histone H3 lysine 36 in

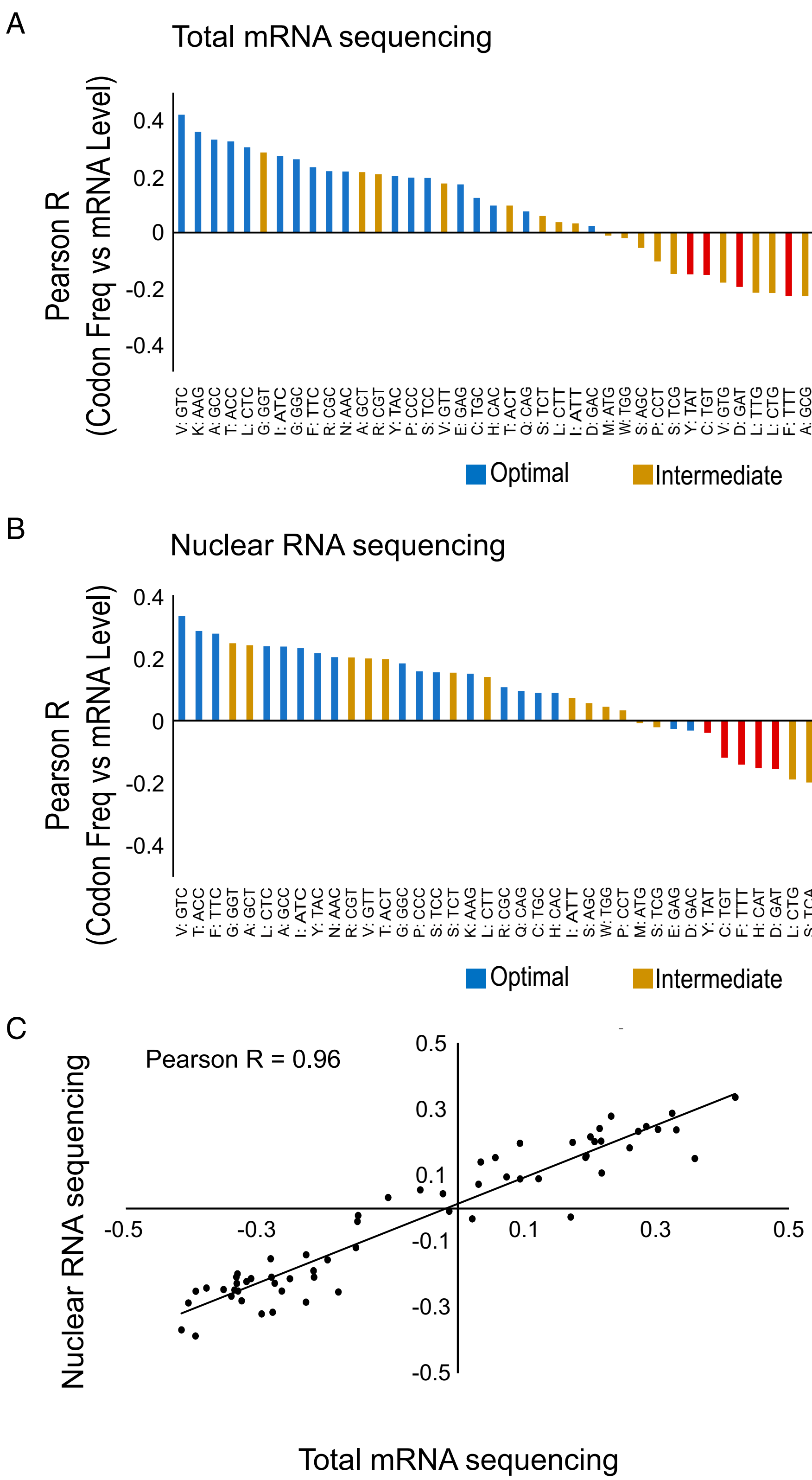
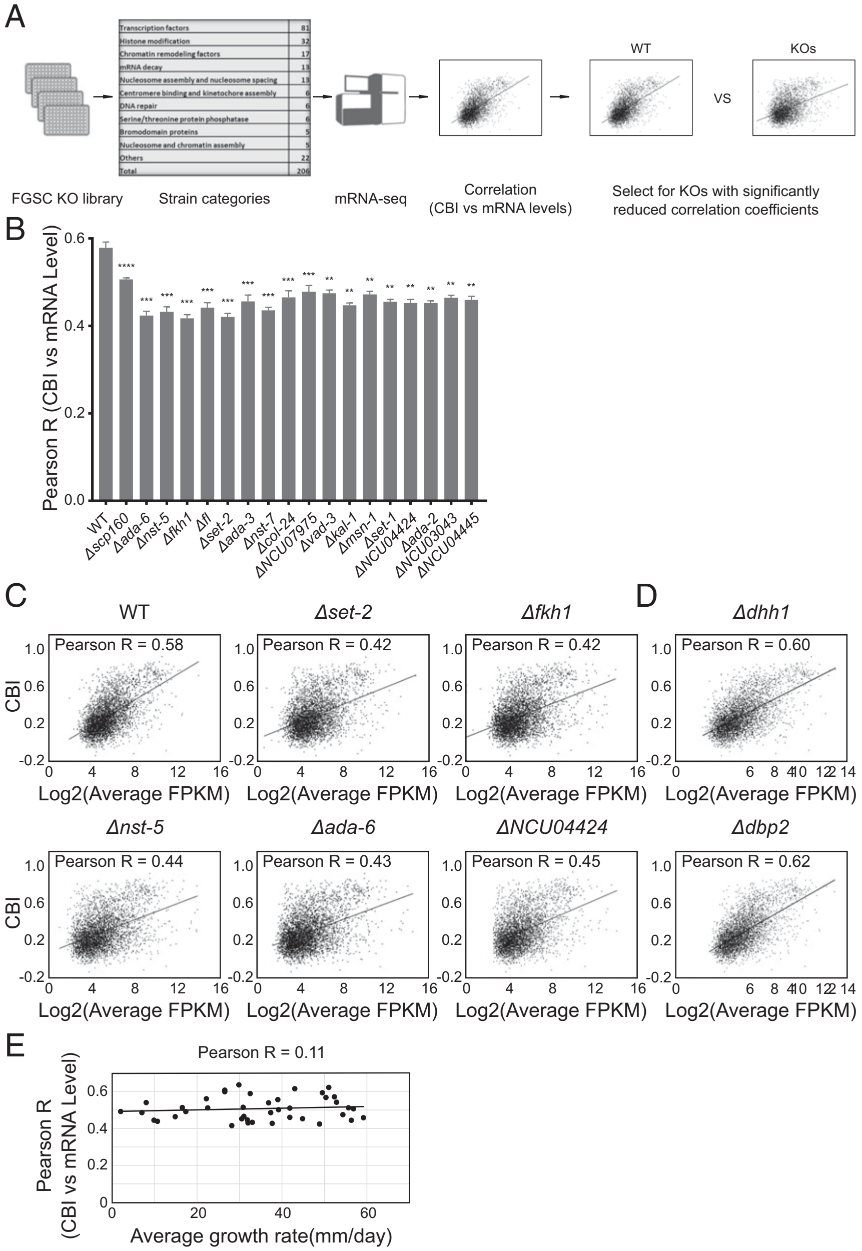


Fig. 2. Codon usage frequency correlates with total and with nuclear RNA levels. (*A* and *B*) Pearson correlation coefficients between individual codon occurrence and mRNA level plotted for 61 codons calculated based on *A* total RNA-seq data (averaged over two replicate samples) and *B* nuclear RNA-seq data (averaged over two replicate samples); experiments were performed with the wild-type strain FGSC4200. Codon frequencies in each gene were calculated as the ratio between the number of any specific codon and the total number of codons in that gene. Optimal codons, those most frequently used in the *N. crassa* genome within each synonymous codon family, are indicated by blue bars. The least-preferred (rare) codons are indicated by red. The intermediate codons are indicated by yellow bars. (*C*) Correlation between the correlation coefficients for total and nuclear codon occurrence versus RNA level.



Transcription factors	81
Histone modification	32
Chromatin remodeling factors	17
mRNA decay	13
Nucleosome assembly and nucleosome spacing	13
Centromere binding and kinetochore assembly	6
DNA repair	6
Serine/threonine protein phosphatase	6
Bromodomain proteins	5
Nucleosome and chromatin assembly	5
Others	22
Total	206

Fig. 3. Screen for factors mediating the effects of codon usage on mRNA level. (*A*) Schematic of the RNA-seq-based genetic screening process used to identify knockout strains with significantly lower CBI versus mRNA level correlations than the wild-type (FGSC4200) strain. Correlation coefficients were averaged over three replicate samples. (*B*) Comparison of correlations between CBI and mRNA levels in the wild-type and 18 identified knockout strains with significantly lower correlations. $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. (*C*) Correlations between CBI and mRNA levels in the wild-type and representative identified knockout strains. Correlations for wild-type strain were obtained from three replicates of samples in a representative batch. (*D*) Correlations between CBI and mRNA levels in the *dhh1*^{KO} and *dbp2*^{KO} strains. All correlations in *C* and *D* are statistically significant with *P* value <2.2e-16. (*E*) Correlations between CBI and RNA levels plotted versus growth rates for 40 selected single-gene knockout strains.

Neurospora (35). To examine whether loss of SET-2 causes correlation changes in only certain codons, we determined the single codon correlations with mRNA levels in the *set-2*^{KO} mutant compared to the wild-type strain. The correlations for most codons were decreased in the *set-2*^{KO} mutant but the codon optimality order among synonymous codons observed for the wild-type strain was largely maintained in the mutant (Fig. 4*C*). This result suggests that SET-2 has a general impact on the codon usage effect on mRNA levels rather than a specific effect on certain codons. Similar decreases of correlations for most codons were also found in other mutants (*SI Appendix*, Fig. S6). The lack of effect on specific codons for some sequence-specific transcription factors could be attributed to several reasons. First, their recognition motifs may be enriched in multiple codons on different open reading frames. Second, codon usage bias of genes shares the same bias for many codon families and genes with strong codon bias typically having similar biases for all preferred codons and nonpreferred codons. As a result, the specific codon effect can be masked. Third, their codon usage-dependent role on gene expression might be due to an indirect effect.

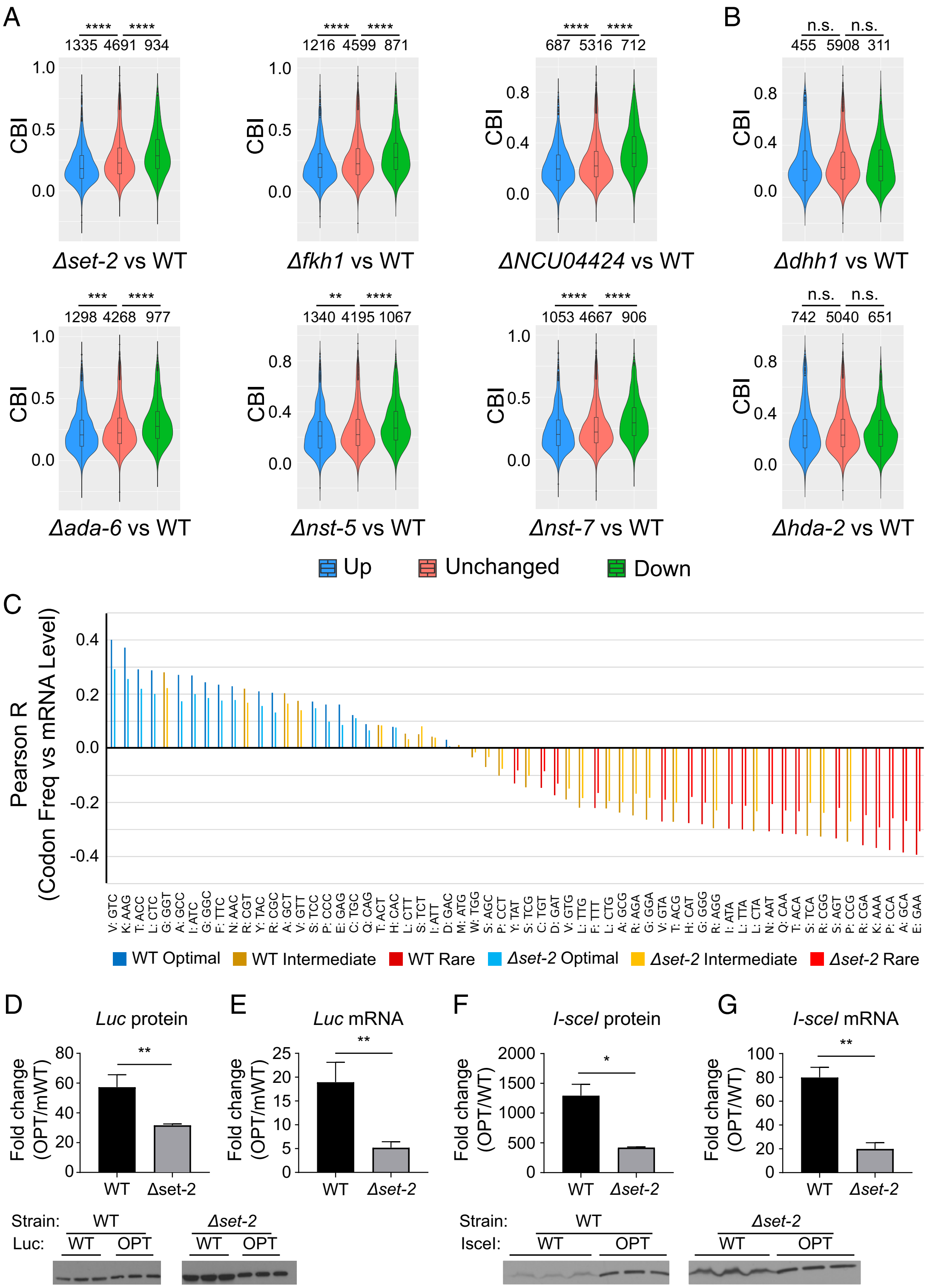


Fig. 4. Candidate factors regulate mRNA levels in a codon usage-dependent manner. (*A*) Violin plots of CBI distribution for differentially regulated genes in six representative candidate strains. $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$. (*B*) Violin plots of CBI distribution for differentially regulated genes in *dhh1*^{KO} and *hda-2*^{KO} strains. n.s.: not significant. (*C*) Pearson correlation coefficients between codon frequencies and mRNA levels for 61 codons in the wild-type (FGSC4200) and *set-2*^{KO} strains. mRNA levels were averaged over three replicate samples. Optimal, intermediate, and rare codons are indicated by red, yellow, and blue bars, respectively. (*D*) Quantification of fold differences of *luciferase* between codon-optimized (OPT) and middle-region WT protein levels in the wild-type and *set-2*^{KO} strains. Protein fold differences were calculated as the ratio between OPT protein levels and the average of WT protein levels in the indicated strains. $**P < 0.01$. Western blot is shown below the graph. OPT protein samples were diluted 50 times. (*E*) Quantification of fold differences in *luciferase* mRNA level between the OPT and WT *luciferase* expressed in the WT strain and *set-2*^{KO} strain; mRNA levels were quantified by RT-qPCR. Fold differences were calculated as described in *D*. $**P < 0.01$. (*F*) The same analysis as in *D* using *I-sceI* as the reporter gene. $*P < 0.05$. Western blot is shown below the graph. OPT protein samples were diluted 500 times. (*G*) Same analysis as in *E* using *I-sceI* as the reporter gene. $**P < 0.01$. *P* values in *D*–*G* were obtained from Student's *t* test with three independent replicates.

To further confirm the codon usage-dependent effect in the *set-2*^{KO} mutant, we introduced a codon-optimized and a middle-region WT *luciferase* expression construct separately into the wild-type and *set-2*^{KO} strains and compared the fold changes of luciferase protein and mRNA. It was previously shown that codon optimization of *luciferase* results in a dramatic increase in luciferase protein and RNA levels (10). The fold changes in amounts of luciferase protein and mRNA upon codon optimization were both significantly reduced in the *set-2*^{KO} mutant compared to the wild-type strain (Fig. 4 *D* and *E*). The reduction of fold change is mainly due to the increase of the WT luciferase protein or mRNA levels in the *set-2*^{KO} mutant (*SI Appendix*, Fig. S7 *A* and *B*). Similar results using another reporter gene *I-sceI* further confirmed the impaired codon usage effect in the *set-2*^{KO} mutant (Fig. 4 *F* and *G* and *SI Appendix*, Fig. S7 *C* and *D*). Together, these results suggest that SET-2 is involved in mediating the codon usage effect on mRNA levels.

Multiple Pathways Influence the Codon Usage-Mediated mRNA Effect. The identification of 18 candidate knockout strains prompted us to examine whether these genes function in distinct pathways to regulate the codon usage-mediated mRNA effect. To address this question, we performed hierarchical clustering of these strains based on differentially expressed genes. The results are shown in Fig. 5*A*. These strains were clustered into three different groups, with *Set-2*^{KO} and *scp160*^{KO} being most distant from most of the other strains. *set-2*^{KO} is one of the strains with the most severely impaired codon-mRNA correlations. *Scp160* is predicted to encode a multi-K homology domain RNA binding protein and is the only nonchromatin regulatory factor in the set of 18 candidate genes. Interestingly, its homolog in yeast, Scp160p, was previously shown to regulate the efficiency of translation of codon-optimized mRNAs (36), but whether it alters mRNA levels in a codon usage-dependent manner is not known. NCU07975, NCU04445, and NCU04424 are all predicted to encode chromatin remodeling factors and are clustered together, suggesting that they function in the same pathway. The transcription factors *ada-6* and *fluffy* (*fl*) are clustered together. The transcription factor encoded by *ada-6* was previously shown to regulate *fl* expression (37), suggesting that some of these factors mediate their codon-mRNA effect indirectly. In total, the expression of ∼5,200 genes are differentially regulated in these 18 strains, about 70% of the detectable genes in *Neurospora*. Together, these results indicate that the codon usage effect on mRNA levels is mediated by multiple independent pathways in *Neurospora* and most of the pathways involve mediators of chromatin structure or transcription factors.

We also performed principal component analysis based on mRNA expression profiles for all the screened strains. The result showed that our identified candidate strains are largely separated from the noncandidate strains (Fig. 5*B*), suggesting the existence of a common effect on gene regulation in the candidate strains. Importantly, when we performed the principal component analysis by using only 20% of genes with the highest CBI and 20% of genes with the lowest CBI, there was a clearer separation between the candidate strains and the noncandidate strains (Fig. 5*C*). It is interesting to note that *scp160*, which was implicated in regulating the translation efficiency of codon-optimized mRNAs, was separated from the rest of the identified genes. These results further suggest the roles of the candidate genes in codon usage-dependent effect on gene expression.

SET-2 and FKH1 Affect Transcription in a Codon-Dependent Manner. To address the mechanisms involved in the effect of codon usage on mRNA levels, we examined the two strains with the most severe impairment of CBI versus mRNA level correlation, *set-2*^{KO}

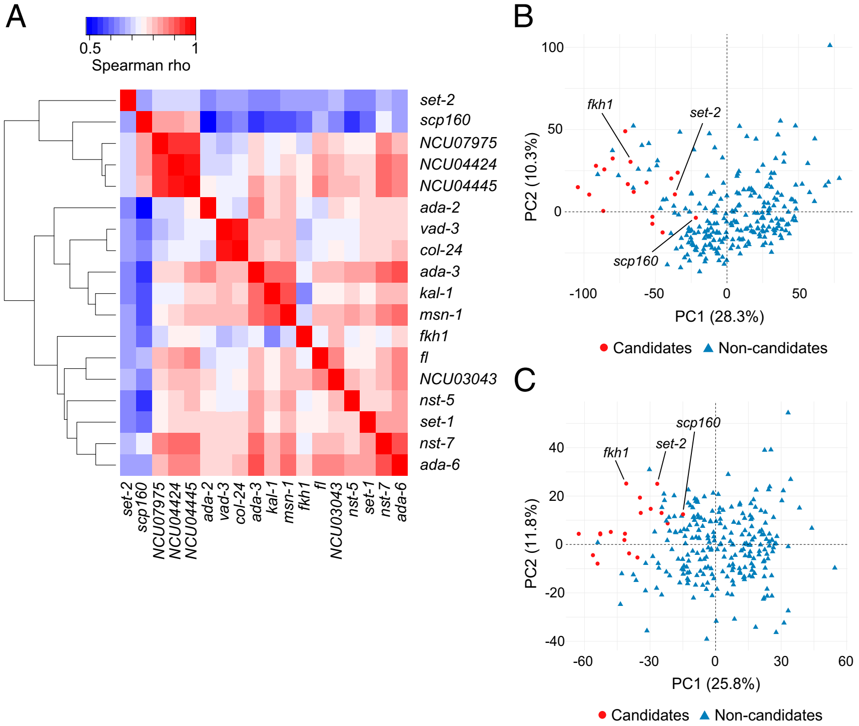


Fig. 5. Clustering analyses for the candidate strains. (*A*) Hierarchical clustering of the candidate strains based on their differentially expressed genes. For each pair of strains, the differentially expressed genes were combined and the Spearman correlations of Log$_2$ (fold change) were used to calculate distance matrix for agglomerative hierarchical clustering. (*B*) Principal component analysis for all the strains screened based on their expression profile. Candidate and noncandidate strains are indicated by red and blue symbols, respectively. (*C*) Principal component analysis for the same set of strains using only 20% of genes with the highest CBI and 20% of genes with the lowest CBI.

and *fkh1*^{KO}, by nuclear RNA sequencing (Dataset S3). The methyltransferase SET-2 influences the efficiency of transcription elongation, impacts chromatin structure, suppresses cryptic transcription, and is required for normal development in *Neurospora* (35, 38, 39). FKH1 is a conserved eukaryotic forkhead family protein. In addition to regulating the expression of mitotic regulators during cell cycle progression, its yeast homolog is known to be involved in chromatin remodeling, and it also associates with the coding regions of active genes and regulates transcription elongation and termination by affecting the phosphorylation of the Pol II CTD (40–42). As expected, the correlations between CBI and nuclear RNA levels were significantly decreased in the *set-2*^{KO} and *fkh1*^{KO} strains compared to the wild-type strain (Fig. 6*A*). In the *set-2*^{KO} strain, the up-regulated nuclear RNAs are enriched for genes with low CBIs and the down-regulated nuclear RNAs are enriched for genes with high CBIs (Fig. 6 *B*, *Right*). In the *fkh1*^{KO} strain, the down-regulated nuclear RNAs are significantly enriched for genes with high CBIs (Fig. 6 *B*, *Left*). Furthermore, most of the differentially regulated genes in the *set-2*^{KO} and *fkh1*^{KO} strains do not overlap, suggesting that SET-2 and FKH1 regulate the codon usage effect on transcription by different mechanisms.

We also performed total RNA-seq on samples of the same tissues used for nuclear RNA-seq. The comparison of the total and nuclear RNA-seq results revealed that most of the up- and down-regulated genes in the total RNA-seq also showed the same regulation in nuclear RNA-seq (Fig. 6*C*). These results suggest that the reduced correlations observed in the *set-2*^{KO} and *fkh1*^{KO} strains are mostly due to transcriptional effects of these two factors. To exclude the possibility that the impaired codon usage-mRNA correlations in these mutants are just due to these mutants sharing particular sets of differentially regulated transcripts with altered codon biases, we identified the overlapping differentially regulated genes between the *set-2* and *fkh-1* mutants. As shown in *SI Appendix*, Fig. S8, these mutants still exhibited a marked reduction of the genome-wide correlation between CBI and mRNA/nuclear RNA levels when these overlapping differentially regulated genes were removed from the analyses.

To further confirm that the nuclear RNA level changes in the *set-2*^{KO} strain are due to changes at the transcriptional level, we performed ChIP-qPCR experiments using antibodies to the nonphosphorylated Pol II CTD or to the Ser-2-phosphorylated CTD and examined the enrichment of Pol II over 11 differentially expressed genes (CBIs of the genes are provided in Dataset S4). As expected, binding of Pol II was increased over the five up-regulated genes and decreased over the six down-regulated genes in the *set-2*^{KO} strain compared to the wild-type strain (Fig. 7). Together, our results suggest that the candidate factors we identified in our study are involved in regulating mRNA transcription in a codon usage-dependent manner.

Discussion

Codon usage plays a major role in controlling gene expression levels at both transcriptional and translational levels. Although the role of codon usage in regulating transcription is conserved in fungal and mammalian cells, the mechanism is unknown. Our study intended to determine the extent and mechanism of this phenomenon. Using *Neurospora* as our model system, we showed that there are strong genome-wide correlations between codon usage and RNA levels in the nucleus. In addition, correlations between codon usage bias and RNA levels were almost the same for both total and nuclear RNA. These results suggest that the genome-wide correlations between codon usage and RNA levels are largely determined by effects on transcription rather than translation-dependent effects.

To determine how codon usage influences transcription, we carried out a large-scale RNA-seq-based screen utilizing available *Neurospora* knockout strains. This brute-force screen led to the identification of 18 candidate genes that when deleted result in reduced correlations between CBI and RNA levels. Several lines of evidence suggest that these genes are involved in mediating the role of codon usage in transcription. First, all except one of these factors are predicted to be transcription regulators: transcription factors, chromatin remodelers, histone methyltransferases, and histone deacetylases. Second, deletion of these genes results in differential gene expression in a codon usage-dependent manner: down-regulated genes are enriched for genes with high CBI

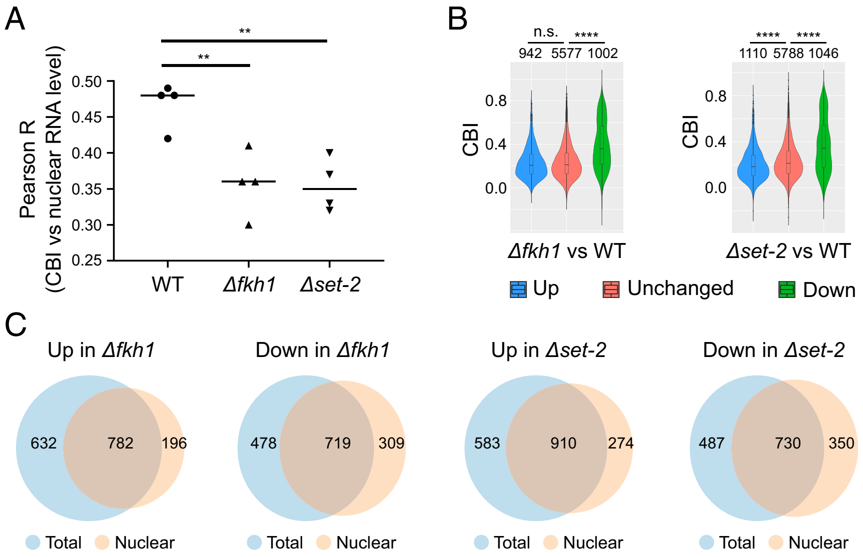


Fig. 6. Differentially regulated genes largely overlap in the *set-2*^{KO} and *fkh1*^{KO} strains. (*A*) Pearson correlation coefficients between CBI and nuclear RNA levels in the wild-type, *set-2*^{KO}, and *fkh1*^{KO} strains. Data for four replicate samples for each strain are plotted; the medians are indicated by the horizontal lines. $^{**}P < 0.01$ (measured by Student's *t* test with four replicates). (*B*) Violin plots of CBI distributions for differentially regulated genes in the nuclei in the *set-2*^{KO} and *fkh1*^{KO} strains. Differentially regulated genes were identified from two replicate samples of nuclear RNA-seq experiments. $^{****}P < 0.0001$, n.s.: not significant. (*C*) Pie charts showing the overlaps of up- and down-regulated genes between total RNA-seq and nuclear RNA-seq data in the *set-2*^{KO} and *fkh1*^{KO} strains.

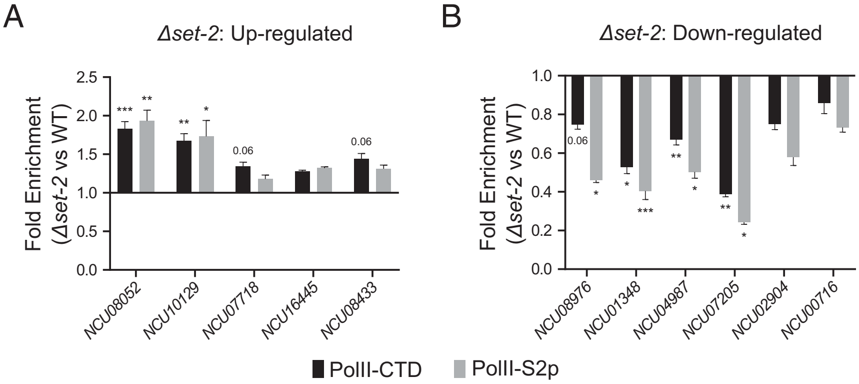


Fig. 7. Pol II enrichment correlates with differential gene expression in the *set-2*^{KO} strain. Fold enrichment in ChIP assays using an antibody to non-phosphorylated Pol II CTD and an antibody to the Ser-2-phosphorylated CTD at the (*A*) up-regulated and (*B*) down-regulated gene loci in the *set-2*^{KO} strain. The differentially regulated genes were selected from the RNA-seq results. Fold enrichments were calculated as the ratio of tubulin-normalized target DNA levels in the *set-2*^{KO} and wild-type strains. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. *P* values were obtained from Student's *t* test with three replicates.

and up-regulated genes are enriched for genes with low CBIs. Third, sequencing of nuclear RNA of two strains, the *set-2*^{KO} and the *fkh1*^{KO} strains, revealed reduced CBI versus RNA level correlations in the nucleus and codon usage-dependent differential gene expression. Fourth, the large overlap between genes differentially expressed in the total and nuclear RNA samples suggests that the reduced correlations are largely due to transcriptional effects. Finally, we showed that the codon usage effect observed for two reporter genes in the wild-type strain was significantly reduced in the *set-2*^{KO} strain. Together, these results suggest that the 18 identified genes encode factors that mediate the codon usage effects on transcription. It is important to note that although deletion of these candidate genes results in impaired genome-wide codon-RNA correlation and codon usage-dependent differential gene expression, the diversity of these transcriptional regulators and our finding that deletion of any of these single genes only results in differential gene expression of a subset of genes suggest that multiple independent pathways mediate the influence of codon bias on transcription.

Interestingly, the deletion of the *Neurospora* homologs of yeast Dhh1 and Dbp2, which were previously shown to regulate mRNA levels in codon usage- and translation-dependent manners (33, 34), did not have significant effects on the genome-wide correlation between codon usage and mRNA levels in *Neurospora*. Consistent with the fact that most of the identified factors are transcriptional regulators, these results also suggest that transcription plays a more important role in determining the codon-RNA correlation than translation, at least in *Neurospora*.

Although the identification of these factors laid the foundation for the mechanistic understanding of the role of codon usage in regulating transcription, how they regulate gene transcription in a codon usage-dependent manner is still not clear. Among the mutants identified, *set-2*^{KO} and *fkh1*^{KO} strains exhibited the most severe impairment in the genome-wide correlation between CBI and RNA levels. SET-2 methylates H3K36 cotranscriptionally and regulates transcription elongation and chromatin structure (35, 38, 39). The yeast homolog of the conserved eukaryotic forkhead family transcription factor FKH1 is known to be associated with gene coding regions to regulate transcription elongation and termination by affecting the Pol II CTD phosphorylation status (40–42). Thus, both factors are directly involved in regulating chromatin structure of coding regions, which may also spread into surrounding regions such as gene promoters. Since codon usage bias also determines the nucleotide sequence of coding regions, it is possible that regions with optimal codon usage profiles have different affinities for these factors than do regions with nonoptimal codons, resulting in chromatin structure changes that influence transcription in a codon usage-dependent but translation-independent manner. Consistent with a role for codon usage in chromatin structure, NCU07975, NCU04445, and NCU04424, which encode potential chromatin remodelers, were also shown to have impaired genome-wide correlations between CBI and RNA levels. In addition, it is also possible that these identified factors may mediate the effect of codon usage on transcription indirectly by regulating the expression of factors that are directly involved. Since the transcriptional effect of codon usage bias can be independent of translation, the DNA sequence of the coding region may be recognized by the transcriptional regulators in the form of DNA elements resulting in chromatin structures that suppress or activate transcription. Future mechanistic studies of these factors should reveal how they regulate gene expression in a codon usage-dependent manner.

Materials and Methods

Strains and Culturing Conditions. The *Neurospora* strains used in this study are listed in *SI Appendix*, Table S1, including the wild-type strain, the identified candidate strains, control strains, and strains transformed with reporter genes. All strains used for the RNA-seq screen are listed in Dataset S1. The *Neurospora* knockout strains were obtained from the Fungal Genetics Stock Center (43) and were inoculated onto minimal slants [3% sucrose, 1× Vogel's salt solution (44), and 1.5% agar]. After 8 to 21 d of growth on slants, the conidia suspension in liquid medium was inoculated into 50 mL liquid medium (2% glucose and 1× Vogel's salt solution) in a 150-mm Petri dish. After incubating the Petri dishes at room temperature for 30 to 48 h to allow the formation of a mycelial mat, mycelium disks were cut and inoculated into 50 mL of liquid medium in flasks. The mycelium disks were then grown on a shaker at room temperature for 24 h under constant light. For total mRNA samples sequenced by the Joint Genome Institute, 2% sucrose was used instead of glucose in liquid medium and the tissue samples were grown at 30 °C in flasks. The tissue samples were then harvested, dried and snap frozen in liquid nitrogen and stored at −80 °C for subsequent experiments. The identity of the knockout strains was verified by the RNA-seq data with FPKM (fragments per kilobase of transcript per million mapped reads) of target genes to be 0 or close to 0.

Constructs and Plasmid Transformation in *Neurospora*. Codon-optimized (OPT)/middle-region WT *luciferase* and OPT/WT *I-sceI* were obtained from our previous study (16) and the constructs were inserted into a modified vector with *csr-1* and *bar* double selection markers.

Plasmids containing wild-type or codon-optimized reporter genes were transformed into *Neurospora* by homologous recombination. The cassettes containing the reporter genes together with a *bar* gene were integrated at the *csr-1* locus, which results in resistance to both cyclosporin A and glufosinate. Approximately 1 to 2 μg plasmids were transformed into freshly isolated conidia by electroporation (45) and the resulting transformants were first selected in nitrogen-free bottom agar [1× nitrogen-free Vogel's salt solution with 0.5% proline in place of NH_4NO_3, 1.5% agar, 1× FGS (0.05% glucose, 0.05% fructose and 2% sorbose) (46)] with 250 μg/mL glufosinate for 2 d. Afterward, the transformants were further selected by overlay of top agar (1× Vogel's salt solution, 1.5% agar, 3% sucrose) with 5 μg/mL cyclosporin A for approximately 1 to 2 d. The selection was performed in 15-cm Petri dishes at 30 °C. For each transformation assay, at least three independent transformants were obtained and used in our experiments. The transformants were verified by their expression of the reporter genes.

RNA Sequencing and Data Analyses. Methods for nuclear RNA extraction, RNA sequencing, and data analysis are described in *SI Appendix*.

2P-Seq and ChIP-Seq Data Processing. The same set of 4,136 genes used for RNA-seq were used in our ChIP-seq and 2P-seq analyses. For 2P-seq data, we used raw counts to perform correlation analyses since each read represents a single transcript in 2P-seq. For ChIP-seq data, reads per million mapped reads (RPM) of each gene were normalized to the RPM obtained from input samples, and the ratio was used to quantify the relative enrichment of Pol II.

Calculation of CBI and tAI_g. The CBIs for all *N. crassa* transcripts were calculated using CodonW 1.4.2 (http://codonw.sourceforge.net/) (47). For genes with multiple variants, the CBI of the T0 variant in the *Neurospora* transcripts reference file was used to represent the CBI of the gene. The tAI is a measure of codon usage bias based on tRNA gene copy number and codon-anticodon interaction (48). The tAI_g for yeast genes was obtained together with mRNA synthesis rate and half-life data from the previous study (32).

Growth-Rate Measurement. A race tube assay was used to measure the growth rate of a selected set of stains (49). Each strain was inoculated at one end of a tube, and tubes were incubated at room temperature under constant light. The growth distances were measured every 24 h, and the average growth rates are reported in millimeters per day.

Clustering Analysis and Principal Component Analysis of the Knockout Strains Based on Gene Expression Profiles. A distance matrix was generated between the 18 candidate strains with impaired correlations between CBI and RNA levels compared to the wild-type strain. Genes with average FPKM less than 1 in any strain were excluded from the analyses. For every pair of the strains, their differentially regulated genes were combined and the distances (or dissimilarity) were calculated as (1-Spearman correlations of Log_2 [fold change]). The distance matrix was used for agglomerative hierarchical clustering by complete linkage method to generate the heatmap in R (version 3.6.1) with heatmap.2 function.

For principal component analysis, the average FPKM of the 4,136 gene sets from three replicates of RNA-seq results of each strain were used in the analysis in Fig. 5*B*. Principal component analysis was performed in R (version 3.6.1) with function "prcomp" and the results were plotted by function "fviz_pca_ind" in the package "factoextra." In the analysis in Fig. 5*C*, the 4,136 genes were ranked by their CBI values and the top 20% and bottom 20% of the genes were selected for the analysis.

Data Sources. 2P-seq and ChIP-seq data were obtained from previous studies in Y.L.'s laboratory (10, 26). RNA synthesis rates and half-lives data in *S. cerevisiae* were obtained from a previous study (32).

Data Availability. *SI Appendix*, Table S1 includes a complete list of the strains in the genetic screen. Dataset S2 contains FPKM/RPKM (reads per kilobase of transcript per million mapped reads) data for total RNA sequencing performed by the Joint Genome Institute. Dataset S3 contains the results for nuclear RNA sequencing (including corresponding total RNA sequencing controls). The raw sequencing data for the strains used in the genetic screen can be accessed at the Genome Portal of the Joint Genome Institute under proposal ID 503459 (50) and project ID 1185261 under proposal ID 982 (51). Customized python and R codes can be accessed in Github at https://github.com/zfz1991/PNAS-codon-transcription.

ACKNOWLEDGMENTS. We thank members of Y.L.'s laboratory for technical assistance and discussion. We thank Drs. Zachary Lewis and Michael Freitag for assistance during the genetic screen and Dr. Tao Wang for suggestions on statistical analyses. This work was supported by grants from the NIH (R35GM118118), the Cancer Prevention and Research Institute of Texas (RP160268), and the Welch Foundation (I-1560) to Y.L. The work conducted by the US Department of Energy (DOE) Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the DOE under Contract No. DE-AC02-05CH11231.

1. T. Ikemura, Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* **2**, 13–34 (1985).
2. P. M. Sharp, T. M. Tuohy, K. R. Mosurski, Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* **14**, 5125–5143 (1986).
3. J. M. Comeron, Selective and mutational patterns associated with gene expression in humans: Influences on synonymous composition and intron presence. *Genetics* **167**, 1293–1304 (2004).
4. J. B. Plotkin, G. Kudla, Synonymous but not the same: The causes and consequences of codon bias. *Nat. Rev. Genet.* **12**, 32–42 (2011).
5. R. Hershberg, D. A. Petrov, Selection on codon bias. *Annu. Rev. Genet.* **42**, 287–299 (2008).
6. Y. Xu *et al*., Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. *Nature* **495**, 116–120 (2013).
7. M. Zhou *et al*., Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* **495**, 111–115 (2013).
8. W. Hense *et al*., Experimentally increased codon bias in the Drosophila Adh gene leads to an increase in larval, but not adult, alcohol dehydrogenase activity. *Genetics* **184**, 547–555 (2010).
9. B. L. Lampson *et al*., Rare codons regulate KRas oncogenesis. *Curr. Biol.* **23**, 70–75 (2013).
10. Z. Zhou *et al*., Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E6117–E6125 (2016).
11. L. Jeacock, J. Faria, D. Horn, Codon usage bias controls mRNA and protein abundance in trypanosomatids. *eLife* **7**, e32496 (2018).
12. V. D. Gooch *et al*., Fully codon-optimized luciferase uncovers novel temperature characteristics of the Neurospora clock. *Eukaryot. Cell* **7**, 28–37 (2008).
13. H. Gingold, Y. Pilpel, Determinants of translation efficiency and accuracy. *Mol. Syst. Biol.* **7**, 481 (2011).
14. T. E. Quax, N. J. Claassens, D. Söll, J. van der Oost, Codon bias as a means to fine-tune gene expression. *Mol. Cell* **59**, 149–161 (2015).
15. C. E. Gamble, C. E. Brule, K. M. Dean, S. Fields, E. J. Grayhack, Adjacent codons act in concert to modulate translation efficiency in yeast. *Cell* **166**, 679–690 (2016).
16. C. H. Yu *et al*., Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol. Cell* **59**, 744–754 (2015).
17. F. Zhao, C. H. Yu, Y. Liu, Codon usage regulates protein structure and function by affecting translation elongation speed in Drosophila cells. *Nucleic Acids Res.* **45**, 8484–8492 (2017).
18. D. E. Weinberg *et al*., Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep.* **14**, 1787–1799 (2016).
19. Q. Yang *et al*., eRF1 mediates codon usage effects on mRNA translation efficiency through premature termination at rare codons. *Nucleic Acids Res.* **47**, 9243–9258 (2019).
20. J. Fu, Y. Dang, C. Counter, Y. Liu, Codon usage regulates human KRAS expression at both transcriptional and translational levels. *J. Biol. Chem.* **293**, 17929–17940 (2018).
21. V. Presnyak *et al*., Codon optimality is a major determinant of mRNA stability. *Cell* **160**, 1111–1124 (2015).
22. A. A. Bazzini *et al*., Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *EMBO J.* **35**, 2087–2103 (2016).
23. Q. Wu *et al*., Translation affects mRNA stability in a codon-dependent manner in human cells. *eLife* **8**, e45396 (2019).
24. A. Radford, J. H. Parish, The genome and genes of Neurospora crassa. *Fungal Genet. Biol.* **21**, 258–266 (1997).
25. M. Zhou, T. Wang, J. Fu, G. Xiao, Y. Liu, Nonoptimal codon usage influences protein structure in intrinsically disordered regions. *Mol. Microbiol.* **97**, 974–987 (2015).
26. Z. Zhou, Y. Dang, M. Zhou, H. Yuan, Y. Liu, Codon usage biases co-evolve with transcription termination machinery to suppress premature cleavage and polyadenylation. *eLife* **7**, e33569 (2018).
27. G. Kudla, L. Lipinski, F. Caffin, A. Helwak, M. Zylicz, High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biol.* **4**, e180 (2006).
28. S. Krinner *et al*., CpG domains downstream of TSSs promote high levels of gene expression. *Nucleic Acids Res.* **42**, 3551–3564 (2014).
29. Z. R. Newman, J. M. Young, N. T. Ingolia, G. M. Barton, Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E1362–E1371 (2016).
30. J. L. Bennetzen, B. D. Hall, Codon selection in yeast. *J. Biol. Chem.* **257**, 3026–3031 (1982).
31. P. Komarnitsky, E. J. Cho, S. Buratowski, Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* **14**, 2452–2460 (2000).
32. Y. Harigaya, R. Parker, Analysis of the association between codon optimality and mRNA stability in Schizosaccharomyces pombe. *BMC Genomics* **17**, 895 (2016).

33. A. Radhakrishnan *et al*., The DEAD-box protein Dhh1p couples mRNA decay and translation by monitoring codon optimality. *Cell* **167**, 122–132.e9 (2016).
34. L. Espinar, M. A. Schikora Tamarit, J. Domingo, L. B. Carey, Promoter architecture determines cotranslational regulation of mRNA. *Genome Res.* **28**, 509–518 (2018).
35. K. K. Adhvaryu, S. A. Morris, B. D. Strahl, E. U. Selker, Methylation of histone H3 lysine 36 is required for normal development in Neurospora crassa. *Eukaryot. Cell* **4**, 1455–1464 (2005).
36. W. D. Hirschmann *et al*., Scp160p is required for translational efficiency of codon-optimized mRNAs in yeast. *Nucleic Acids Res.* **42**, 4043–4055 (2014).
37. X. Sun *et al*., The Zn(II)2Cys6-type transcription factor ADA-6 regulates conidiation, sexual development, and oxidative stress response in *Neurospora crassa*. *Front. Microbiol.* **10**, 750 (2019).
38. M. J. Carrozza *et al*., Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**, 581–592 (2005).
39. E. J. Wagner, P. B. Carpenter, Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.* **13**, 115–126 (2012).
40. P. Jorgensen, M. Tyers, The fork'ed path to mitosis. *Genome Biol.* **1**, REVIEWS1022 (2000).
41. J. A. Sherriff, N. A. Kent, J. Mellor, The Isw2 chromatin-remodeling ATPase cooperates with the Fkh2 transcription factor to repress transcription of the B-type cyclin gene CLB2. *Mol. Cell. Biol.* **27**, 2848–2860 (2007).
42. A. Morillon, J. O'Sullivan, A. Azad, N. Proudfoot, J. Mellor, Regulation of elongating RNA polymerase II by forkhead transcription factors in yeast. *Science* **300**, 492–495 (2003).
43. H. V. Colot *et al*., A high-throughput gene knockout procedure for Neurospora reveals functions for multiple transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10352–10357 (2006).
44. H. J. Vogel, A convenient growth medium for *Neurospora crassa*. *Microbial Genetics Bulletin* **13**, 42–47 (1956).
45. B. S. Margolin, M. Freitag, E. U. Selker, Improved plasmids for gene targeting at the *his-3* locus of *Neurospora crassa* by electroporation. *Fungal Genet. Newsl.* **44**, 24–36 (1999).
46. H. E. Brockman, F. J. de Serres, "SORBOSE TOXICITY" IN NEUROSPORA. *Am. J. Bot.* **50**, 709–714 (1963).
47. J. F. Peden, "Analysis of codon usage," PhD thesis, University of Nottingham, Nottingham, UK (1999).
48. M. dos Reis, L. Wernisch, R. Savva, Unexpected correlations between gene expression and codon usage bias from microarray data for the whole Escherichia coli K-12 genome. *Nucleic Acids Res.* **31**, 6976–6985 (2003).
49. J. J. Loros, J. C. Dunlap, Genetic and molecular analysis of circadian rhythms in *Neurospora*. *Annu. Rev. Physiol.* **63**, 757–794 (2001).
50. Y. Liu, Determination of fungal chromatin regulatory network and its impact on gene expression. JGI Genome Portal. https://genome.jgi.doe.gov/portal/DetofExpression/DetofExpression.info.html. Deposited 17 April 2020.
51. N. L. Glass, Y. Liu, The fungal nutritional ENCODE project. JGI Genome Portal. https://genome.jgi.doe.gov/portal/TheFunENCproject/TheFunENCproject.info.html. Deposited 16 April 2018.
52. W. J. Belden *et al*., The band mutation in Neurospora crassa is a dominant allele of ras-1 implicating RAS signaling in circadian output. *Genes Dev.* **21**, 1494–505 (2007).