# Codon usage is an important determinant of gene expression levels largely through its effects on transcription

Zhipeng Zhou[a,1], Yunkun Dang[a,1], Mian Zhou[a,b], Lin Li[c], Chien-hung Yu[a], Jingjing Fu[a], She Chen[c], and Yi Liu[a,2]

[a]Department of Physiology, The University of Texas Southwestern Medical Center, Dallas, TX 75390; [b]State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai 200237, China; and [c]National Institute of Biological Sciences, Changping District, Beijing 102206, China

Edited by Jay C. Dunlap, Geisel School of Medicine at Dartmouth, Hanover, NH, and approved August 11, 2016 (received for review April 27, 2016)

**Codon usage biases are found in all eukaryotic and prokaryotic genomes, and preferred codons are more frequently used in highly expressed genes. The effects of codon usage on gene expression were previously thought to be mainly mediated by its impacts on translation. Here, we show that codon usage strongly correlates with both protein and mRNA levels genome-wide in the filamentous fungus *Neurospora*. Gene codon optimization also results in strong up-regulation of protein and RNA levels, suggesting that codon usage is an important determinant of gene expression. Surprisingly, we found that the impact of codon usage on gene expression results mainly from effects on transcription and is largely independent of mRNA translation and mRNA stability. Furthermore, we show that histone H3 lysine 9 trimethylation is one of the mechanisms responsible for the codon usage-mediated transcriptional silencing of some genes with nonoptimal codons. Together, these results uncovered an unexpected important role of codon usage in ORF sequences in determining transcription levels and suggest that codon biases are an adaptation of protein coding sequences to both transcription and translation machineries. Therefore, synonymous codons not only specify protein sequences and translation dynamics, but also help determine gene expression levels.**

*Neurospora* | codon usage | transcription

Gene expression is regulated by transcriptional and post-transcriptional mechanisms. Promoter strength and RNA stability are thought to be the major determinants of mRNA levels, whereas transcript levels and protein stability are proposed to be largely responsible for protein levels in cells. The use of synonymous codons in the gene coding regions are not random, and codon usage bias is an essential feature of most genomes (1–4). Selection for efficient and accurate translation is thought to be the major cause of codon usage bias (4–9). Recent experimental studies demonstrated that codon usage regulates translation elongation speed and cotranslational protein folding (10–13). This effect of codon biases on translation elongation speed and the correlations between codon usage and certain protein structural motifs suggested that codon usage regulates translation elongation rates to optimize cotranslational protein folding processes (11, 12, 14–16).

Codon optimization has long since been used to enhance protein expression for heterologous gene expression. In addition, the fact that highly expressed proteins are mostly encoded by genes with mostly optimal codons led to the hypothesis that codon usage impacts protein expression levels by affecting translation efficiency (7, 17–19). Recent studies, however, suggested that overall translation efficiency is mainly determined by the efficiency of translation initiation, a process that is mostly determined by RNA structure but not codon usage near the translational start site (20–22). Recently, codon usage was suggested to be an important determinant of mRNA stability in *Saccharomyces cerevisiae* and *Escherichia coli* through an effect on translation elongation (23, 24). It is important to note that *S. cerevisiae* prefers to have A or T at wobble positions, whereas mammals, *Drosophila*, and many other fungi prefer C or G. Because such differences may alter mRNA recognition by RNA decay pathways, it is not known whether a similar mechanism exists in other eukaryotic organisms.

The filamentous fungus *Neurospora crassa* exhibits a strong codon usage bias for C or G at wobble positions (16, 25). Codon optimization has been shown to enhance protein expression of heterologous genes in *Neurospora* (26, 27). We recently demonstrated that codon usage is important for the expression, function, and structure of the clock protein FRQ (13, 16). Codon usage affects local rates of translation elongation in *Neurospora* (preferred codons speed up elongation and rare codons slow it down) and also cotranslational protein folding (12). However, how codon usage regulates gene expression remains unclear.

## Results

**Codon Usage Biases Strongly Correlate with Protein and Transcript Levels Genome-Wide in *Neurospora*.** To determine the codon usage effect on protein expression genome-wide, we performed whole-proteome quantitative analyses of *Neurospora* whole-cell extract by mass spectrometry experiments. These analyses led to the identification and quantification of ∼4,000 *Neurospora* proteins based on their emPAI (exponentially modified protein abundance index) values (28), which are proportional to their relative abundances in a protein mixture. As shown in *SI Appendix*, Fig. S1, the results obtained from analyses of two independent replicate samples were highly consistent, indicating the reliability and sensitivity of the

### Significance

Codon usage bias is an essential feature of all genomes. The effects of codon usage biases on gene expression were previously thought to be mainly due to its impacts on translation. Here, we show that codon usage bias strongly correlates with protein and mRNA levels genome-wide in the filamentous fungus *Neurospora*, and codon usage is an important determinant of gene expression. Surprisingly, we found that the impacts of codon usage on gene expression are mainly due to effects on transcription and are largely independent of translation. Together, these results uncovered an unexpected role of codon biases in determining transcription levels by affecting chromatin structures and suggest that codon biases are results of genome adaptation to both transcription and translation machineries.

Author contributions: Z.Z., Y.D., and Y.L. designed research; Z.Z., Y.D., M.Z., L.L., and S.C. performed research; C.-h.Y. and J.F. contributed new reagents/analytic tools; Z.Z., Y.D., M.Z., S.C., and Y.L. analyzed data; and Z.Z., Y.D., and Y.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]Z.Z. and Y.D. contributed equally to this study.

[2]To whom correspondence should be addressed. Email: yi.liu@utsouthwestern.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1606724113/-/DCSupplemental.

method. In addition, RNA-sequencing (seq) analysis of the *Neurospora* mRNA was performed to determine correlations between mRNA levels with codon usage biases. To determine the codon usage bias of *Neurospora* genes, the codon bias index (CBI) for every protein-coding gene in the genome was calculated. CBI ranges from −1, indicating that all codons within a gene are nonpreferred, to +1, indicating that all codons are the most preferred, with a value of 0 indicative of random use (29). Because CBI estimates the codon bias for each gene rather than for individual codons, the relative codon biases of different genes can be compared.

For the ∼4,000 proteins detected by mass spectrometry, which account for more than 40% of the total predicted protein encoding genes of the *Neurospora* genome, there is a strong positive correlation (Pearson's product-moment correlation coefficient $r$ is 0.74) between relative protein abundances and mRNA levels (Fig. 1*A* and Dataset S1), suggesting that transcript levels largely determine protein levels. Importantly, we also observed a strong positive correlation ($r = 0.64$) between relative protein abundances and CBI values (Fig. 1*B*). Interestingly, a similarly strong positive correlation ($r = 0.62$) was seen between CBI and relative mRNA levels (Fig. 1*C*). Because codon usage was previously hypothesized to affect translation efficiency, we wondered whether mRNA levels could better predict protein levels if codon usage scores were taken into account. Surprisingly, compared with using mRNA alone, the two factors together did not markedly improve the correlation value with protein (Fig. 1*D*). These results suggest the possibility that codon usage is an important determinant of protein production genome-wide mainly through its role in affecting mRNA levels.

Based on phylogenetic distribution, *Neurospora* protein encoding genes can be classified into five mutually exclusive lineage specificity groups: eukaryote/prokaryote-core (conserved in nonfungal eukaryotes and/or prokaryotes), dikarya-core (conserved in Basidiomycota and Ascomycota species), Ascomycota-core, Pezizomycotina-specific, and *N. crassa*-specific genes (30). The median CBI value of each group decreases as lineage specificity (*SI Appendix*, Fig. S1*B*), with the eukaryote/prokaryote-core genes having the highest average CBI values and the *N. crassa*-specific genes having the lowest average values. Interestingly, the difference of median mRNA levels of each gene group correlate with that of the group median CBI values (*SI Appendix*, Fig. S1*C*). These results suggest that codon usage may regulate gene expression by enhancing that of highly conserved genes and/or limiting that of evolutionarily recent genes.

Transcript levels are thought to be mainly determined by promoter strength. Therefore, it is surprising that codon usage, an intrinsic feature within ORFs, could have such a strong correlation with mRNA levels. Unlike *S. cerevisiae, Neurospora* codons prefer C or G at the wobble positions, the observed effects could be due to an effect of gene GC contents. However, at the genome-wide level, gene GC contents [calculated from transcription start site (TSS) to transcription end site (TES)] showed no correlation with protein levels or a weak negative correlation with mRNA levels (Fig. 1 *E* and *F*). However, GC content at the third position of codons (GC3), which strongly correlates with

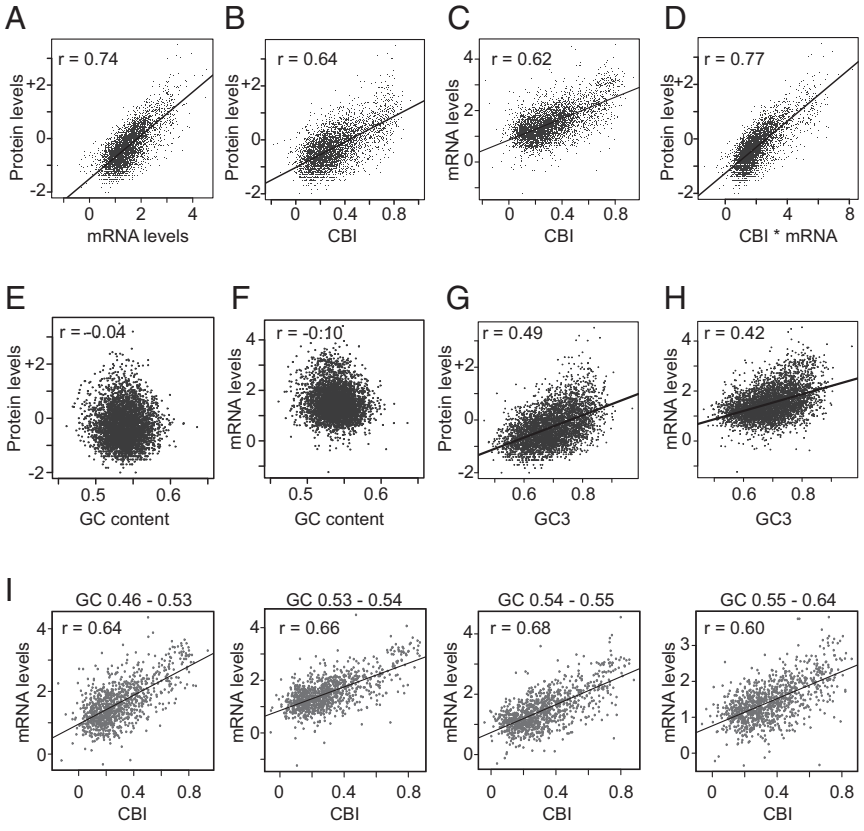


**Fig. 1.** Codon usage but not gene GC content correlates with protein and mRNA levels in *Neurospora*. (*A*) Scatter plot of protein levels ($\log_{10}$ emPAI) vs. mRNA levels ($\log_{10}$ RPKM). $P < 2.2 \times 10^{-16}$, $n = 4{,}013$. (*B*) Plot of protein levels vs. CBI. $P < 2.2 \times 10^{-16}$, $n = 4{,}013$. (*C*) Scatter analysis of mRNA levels vs. CBI. $P < 2.2 \times 10^{-16}$, $n = 4{,}013$. (*D*) Scatter plot of protein levels vs. CBI × mRNA. $P < 2.2 \times 10^{-16}$, $n = 4{,}013$. (*E* and *F*) Scatter plot of protein levels (*E*) or mRNA levels (*F*) vs. gene GC content, $n = 4{,}013$. (*G* and *H*) Scatter plot of protein levels (*G*) or mRNA levels (*H*) vs. GC3. $P < 2.2 \times 10^{-16}$, $n = 4{,}013$. (*I*) Plots of mRNA levels vs. CBI in four groups of genes with different gene GC content. *Neurospora* genes were ranked based on gene GC content, the outlier was removed, and the genes were divided into four groups with equal number of genes based on their gene GC contents. First group, gene GC content 0.46–0.53, $n = 987$; second, GC content 0.53–0.54, $n = 986$; third, GC content 0.54–0.55, $n = 987$; fourth, GC content 0.55–0.64, $n = 986$.

codon usage (*SI Appendix*, Fig. S1*D*), positively correlates with both protein and mRNA levels (Fig. 1 *G* and *H*). Furthermore, when *Neurospora* genes were split into four groups (each with the same number of genes) based on their GC contents, a similarly strong positive correlation between CBI and mRNA levels was seen in each group (Fig. 1*I*). Remarkably, the gene GC contents were almost the same for the two middle groups of genes. Together, these data suggest that codon usage but not just gene GC contents is an important determinant of mRNA expression levels in *Neurospora*.

**Codon Optimization Strongly Enhances Protein and mRNA Levels.** To directly determine the effect of codon usage on gene expression, we codon-optimized (opt) eight *Neurospora* genes and two heterologous reporter genes, firefly luciferase (*luc*) and *S. cerevisiae I-SceI*, based on the *Neurospora* codon usage table (*SI Appendix*, Table S1). These genes [wild-type (wt) or opt] were under the control of *Neurospora ccg-1* or *qa* promoters and were targeted to the *his-3* locus of *Neurospora* by homologous recombination. Homokaryotic transformants were obtained. Codon optimization resulted in marked increases in protein levels for each of the 10 genes (Fig. 2*A* and *SI Appendix*, Fig. S2*A*). For the eight *Neurospora* genes, codon optimization resulted in up to 25-fold increase of protein levels. For the two heterologous genes, luciferase and I-SceI proteins were barely detectable when encoded by wild-type codons; levels were 70- to more than 100-fold higher for the codon-optimized versions.

Comparison of mRNA levels in these strains showed that codon optimization also resulted in marked increases of the corresponding mRNA levels for each of the 10 genes; the fold increases in mRNA levels were comparable to those of the fold increases in protein levels (Fig. 2*B* and *SI Appendix*, Fig. S2*B*). For *luc* and *I-SceI* genes (under the control of the *ccg-1* promoter), the mRNA levels of the codon-optimized genes were more than 70-fold higher than those of the wild-type genes.

To determine whether the codon effect depended on the promoter used to drive the transgene, we used a construct (P*frq-luc*) in which the *luc* gene is under the control of the *frequency* (*frq*) promoter (a weak promoter in *Neurospora*) at the *his-3* locus. More than a 100-fold increase in LUC protein and *luc* mRNA levels due to codon optimization were also observed (Fig. 2*C*). Side-by-side comparison *luc* mRNA levels of the P*frq-luc* and P*ccg-1-luc* strains show that codon optimization had a much stronger effect on mRNA levels than by changing the promoter (Fig. 2*D*). Moreover, when the gene encoding for the septal pore-associated protein (SPA) 16 (*spa16*) was under the control of the *vvd* promoter, which is only activated after light induction (31), codon optimization led to more than 40-fold higher levels of both protein and mRNA levels after light treatment than the gene with the wild-type codon usage (Fig. 2*E*). These results indicate that the effect of codon optimization on gene expression is independent of the promoter used.

Furthermore, similar effects of codon optimization on mRNA levels were also seen when the P*ccg1-luc* transgene was targeted to the *csr-1* locus (*SI Appendix*, Fig. S2*C*), indicating that the effect of codon usage is independent on its genome locus. Together, these results suggest that gene codon usage has an important role in determining gene expression levels in *Neurospora*, an effect that is largely due to changes of transcript levels.

Because of codon biases for C at the wobble positions in *Neurospora*, codon optimization would result in increased GC content. Thus, it is possible that it is GC content rather than codon usage responsible for increased gene expression. To rule out this possibility, we created suboptimal *luc* and *I-SceI* genes, in which some of most preferred codons with C at the wobble position were replaced by the less preferred G. As a result, the GC contents of these suboptimal genes are the same as the fully optimized genes but with reduced codon usage scores. As shown in Fig. 2 *F* and *G*, both protein and mRNA levels of theses suboptimal genes are much lower than that of the fully optimized genes. These results suggest that codon usage, but not just GC content, is important for expression levels of these genes.

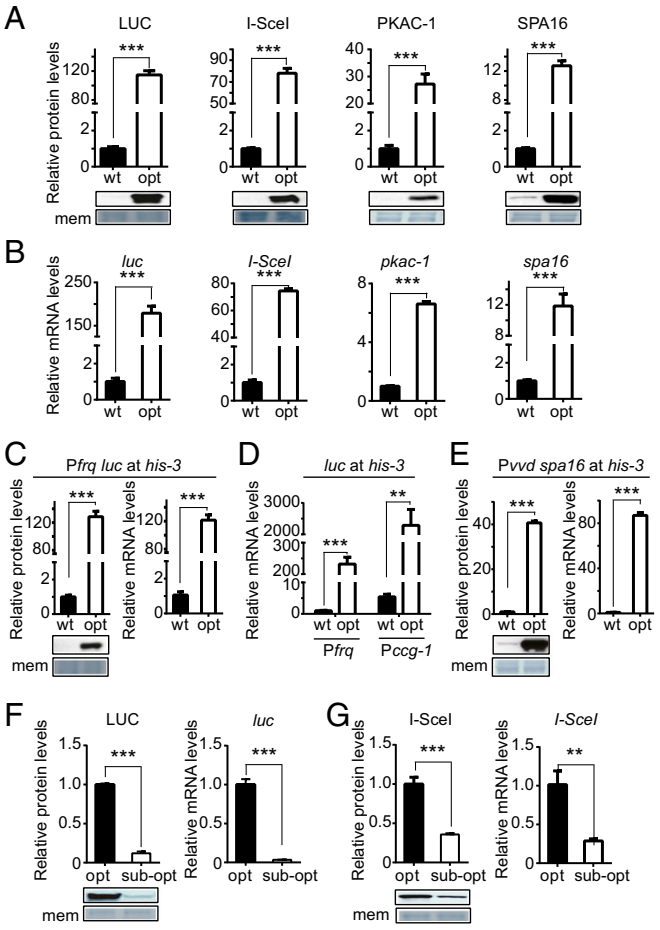


**Fig. 2.** Codon optimization results in increased levels of both protein and mRNA in *Neurospora*. (*A*) Western blot assays showing the effect of codon optimization on protein expression levels in the indicated *Neurospora* strains. An anti-luciferase antibody is used to detect for the protein levels of LUC, and anti-Myc antibody was used for other Myc-fusion proteins. (*A, Bottom*) A representative Western blot showing protein levels of wt or opt strains. (*A, Top*) Densitometric analyses of three independent samples. The *luc*, *I-SceI*, and *spa16* genes are driven by the *ccg-1* promoter, whereas *pkac-1* is under the control of the *qa-2* promoter. Membrane was stained and served as loading control. (*B*) Quantitative RT-PCR results showing the relative indicated mRNA levels of wt or optimized opt *luc*, *I-SceI*, *pkac-1*, and *spa16* strains. (*C*) The relative protein (*Left*) and mRNA (*Right*) levels of wt or optimized opt *luc* under the control of the *frq* promoter at the *his-3* locus. (*D*) The mRNA levels of wt or opt *luc* under the control of the *frq* promoter or the *ccg-1* promoter at the *his-3* locus. (*E*) The relative protein (*Left*) and mRNA (*Right*) levels of wt or opt *spa16* under the control of the *vvd* promoter at the *his-3* locus. The tissues were first cultured in constant darkness for 24 h, then transferred to light for 1 h, and the tissues were harvested. (*F*) The relative protein (*Left*) and mRNA levels of the opt or suboptimized (subopt) *luc* gene. (*G*) The relative protein (*Left*) and mRNA (*Right*) levels of the opt or subopt *I-SceI* gene. Error bars shown in all graphs are SDs of the means ($n = 3$). $^{**}P < 0.01$; $^{***}P < 0.001$.

**Codon Usage Does Not Consistently Influence mRNA Stability.** Non-optimal codons in an mRNA have been suggested to destabilize mRNA during protein translation in yeast and more recently in zebrafish (23, 32, 33). To determine whether the effect of codon usage in *Neurospora* is due to its effects on mRNA stability, we compared mRNA decay rates of seven wild-type and codon-optimized

*Neurospora* and heterologous genes after the addition of thiolutin, a transcription inhibitor in *Neurospora* (34). For five of these genes, codon optimization did not result in significant changes in mRNA decay rates (Fig. 3). The codon-optimized *luc* mRNA was more stable than the wild type; however, the modest change in mRNA stability cannot explain the more than 100-fold difference in mRNA levels observed (Fig. 2*B*). In contrast, the codon-optimized *NCU02621* mRNA is less stable than the wild-type gene. Therefore, unlike in yeast, codon usage does not appear to have a major or universal effect on mRNA stability in *Neurospora*.

**The Effect of Codon Usage Does Not Require Translation and Is Regulated at the Level of Transcription.** We have recently shown that codon usage influences the rate of translation elongation (12). Thus, it is possible that the effects of codon usage on gene expression are mediated by its role in mRNA translation. Three separate lines of evidence, however, indicate that the codon effect on RNA levels is not due to its role in translation. First, treatment of *Neurospora* cultures with the protein synthesis inhibitor cycloheximide (CHX) to block protein translation did not change the effect of codon optimization on *luc* or *I-SceI* mRNAs (Fig. 4*A*). Second, we created strains in which a stop codon was placed at the 14th amino acid position in the *luc* and *I-SceI* transgenes, which terminate translation for the rest of the mRNA and, thus, should eliminate potential translation-mediated codon effect. Although the introduction of the stop codon completely abolished the production of both wild-type and codon-optimized LUC and I-SceI proteins (*SI Appendix*, Fig. S3 *A* and *B*), it did not reverse the dramatic effect of codon optimization on *luc* and *I-SceI* mRNA levels in both quantitative RT-PCR (qRT-PCR) and Northern blot analysis (Fig. 4*B* and *SI Appendix*, Fig. S3 *C* and *D*).

Third, we introduced a stable stem loop, which has been shown to block translation initiation (35), into the 5′ UTR of the *I-SceI* transgene gene to block translation initiation. Although the introduction of the stem loop completely abolished the production of I-SceI protein (Fig. 4*C*) and did not significantly affect RNA stability (*SI Appendix*, Fig. S3*E*), it did not change the dramatic effect of codon optimization on *I-SceI* mRNA (Fig. 4*D*). Together, these results demonstrate that the codon usage effects on mRNA expression do not depend on translation.

The independence of the codon effect on translation prompted us to test whether codon usage affects transcription. We determined the levels of nuclear RNA, which better reflects gene transcription levels than total RNA. As expected, *frq* pre-mRNA was enriched in the nuclear RNA compared with the total RNA (*SI Appendix*, Fig. S3*F*). Codon optimization resulted in a dramatic increase in nuclear transcript levels of both *luc* and *I-SceI* (Fig. 4*E*), suggesting that the effect of codon usage on RNA level is mainly due to its effect on transcription.

To further confirm this conclusion, we introduced an intron (from the *pkac-1* gene) into the 5′ UTR of the *spa16* gene driven by the *vvd* promoter, so that the level of the intronic pre-mRNA could be determined. As shown in *SI Appendix*, Fig. S3*G*, the intron was efficiently spliced in the total RNA. As shown in Fig. 4*F*, the levels of mRNA and pre-mRNAs of the optimized *spa16* were both ∼10-fold higher than the wild-type *spa16*. Together, these data indicate that codon optimization results in increased transcription.

**Codon Optimality Promotes Enrichment of Active RNA Polymerase II.** To determine how transcription is affected, we created strains containing *frq* promoter driven *luc* gene (Fig. 5*A*, *Top*, the middle gray section of the *luc* ORF uses either wild-type or optimized codons). Under constant light, the transcription initiation of the *frq* promoter is mainly triggered by the binding of the WC complex at the pLRE box (36). Chromatin immunoprecipitation (ChIP) assays using WC-2 antibody showed that codon optimization of *luc* resulted in a significant increase of WC-2 binding at the pLRE locus (Fig. 5*A*). It was shown that the histone density affects the recruitment of the WC complex at the *frq* promoter (37), so we wondered whether histone density was affected after codon optimization. Indeed, ChIP assays showed that histone H3 enrichment levels on the opt-*luc* gene were significantly lower than that of the wt-*luc*, especially around the TSS region (Fig. 5*B*).

Nucleosome density is known to affect RNA polymerase II (Pol II)-mediated transcription (38, 39). To confirm the effect of codon usage on transcription, we performed the ChIP assay by using antibodies against nonphosphorylated and S2-phosphorylated Pol II C-terminal end (CTD) to compare their enrichment on wild-type and codon-optimized genes. The results in *SI Appendix*, Fig. S4 showed that ChIP using these two and other antibodies resulted in similar enrichment profiles at an endogenous gene locus as those in other organisms (38). As shown in Fig. 5 *C* and *D* and *SI Appendix*, Fig. S5 *A* and B, codon optimization resulted in a significant enrichment of Pol II CTD and S2P-CTD not only within the ORF region that was codon optimized, but also in adjacent regions of *luc*, *I-SceI*, and two *Neurospora* genes (*spa16* and *NCU02621*).

We also performed next generation sequencing of ChIP for both Ser-2 phosphorylated and nonphosphorylated Pol II and determined the relative enrichment of phosphorylated Pol II for

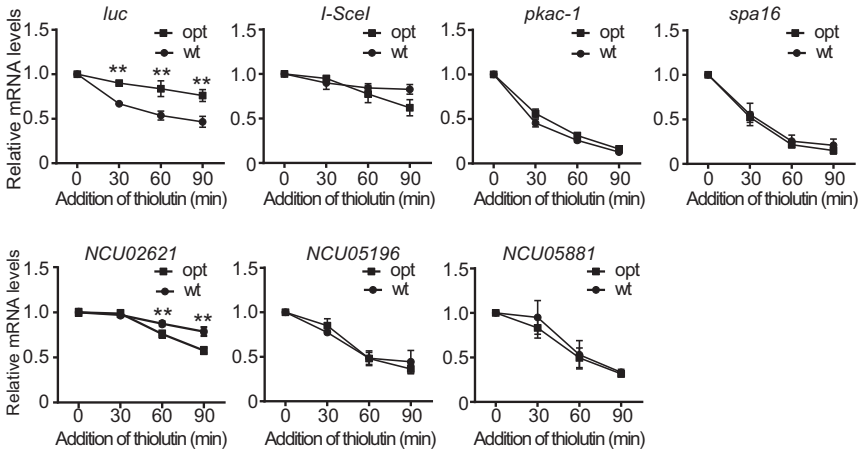


**Fig. 3.** Stability of mRNA is not consistently altered by codon optimization. The decay of the indicated mRNAs in the wt and opt strains is shown at the indicated time points after addition of thiolutin, a potent RNA synthesis inhibitor. The densitometric analyses of the results from three independent experiments are shown. $^{**}P < 0.01$.

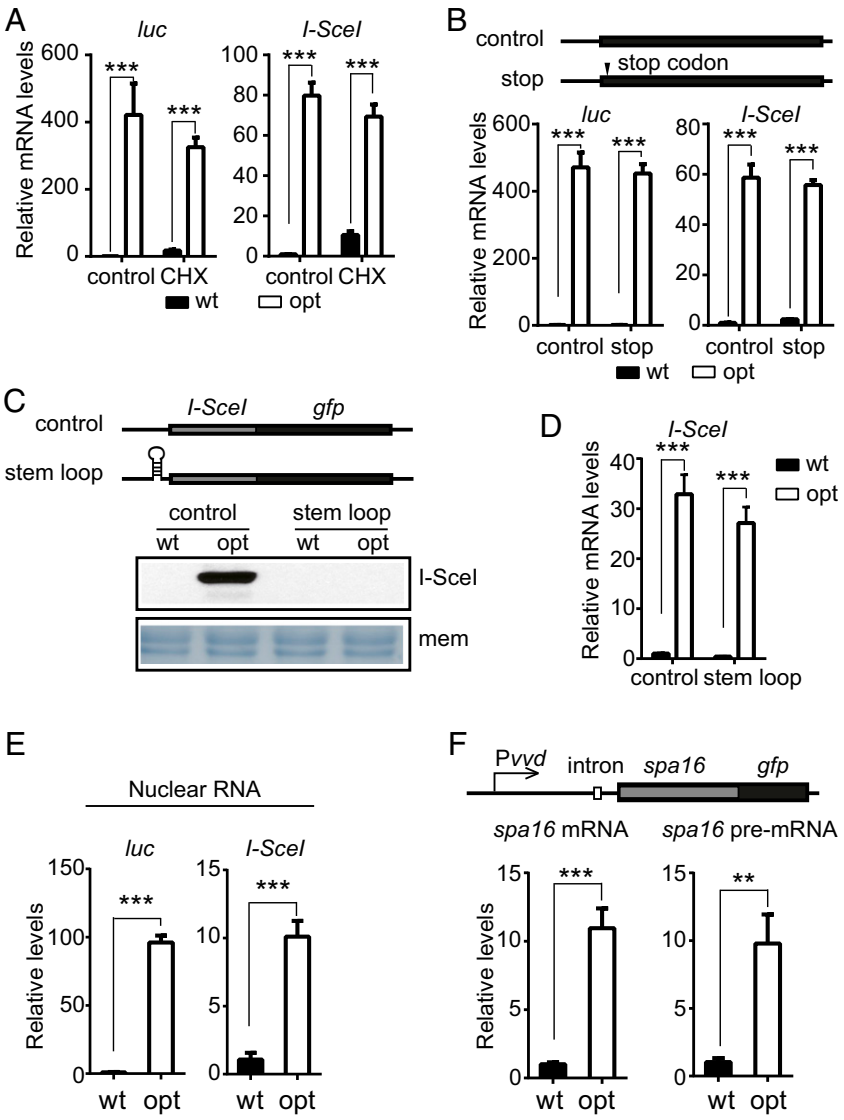


**Fig. 4.** The codon effects on gene expression does not require translation. (*A*) qRT-PCR experiments showing the relative levels of indicated mRNA in the wt and opt strains. CHX was added to block protein synthesis. (*B*) qRT-PCR experiments showing the relative levels of indicated mRNA in the wt and opt strains with or without a stop codon inserted at the 14th amino acid. (*C, Top*) Diagram depicting the *I-SceI* (wt or opt) constructs with/without a stem loop in the 5′ UTR. (*C, Bottom*) Western blot results using anti-Myc antibody showing that the introduction of the stem loop abolished the production of full-length I-SceI protein. (*D*) qRT-PCR results showing the relative levels of the wt and opt *I-SceI* mRNA levels with or without the stem loop in the 5′ UTR. (*E*) qRT-PCR results showing the relative levels of nuclear RNA (*luc* and *I-SceI*) in the indicated wt and opt *luc* and *I-SceI* strains. (*F*) *Top* depicts the constructs in which an intron was introduced into the 5′ UTR of the wt or opt *spa16*. (*F, Bottom*) Relative mRNA (*Left*) and pre-mRNA levels (*Right*) of the wt and opt *spa16* detected by qRT-PCR and strand-specific qRT-PCR, respectively. The tissues were cultured under constant light for 24 h before harvest. Error bars show the SDs of the means ($n = 3$). $^{**}P < 0.01$; $^{***}P < 0.001$.

all annotated *Neurospora* genes. As shown in *SI Appendix*, Fig. S5 *C* and *D*, the enrichment of Pol II positively correlates with mRNA levels genome-wide. Importantly, CBI values also positively correlate with enrichment of S2 phosphorylated (Fig. 5*E*) and nonphosphorylated (*SI Appendix*, Fig. S5*E*) CTD genome-wide, indicating that genes with more optimal codons are associated with higher levels of transcription. These results suggest that gene codon usage is an important determinant of gene transcription levels in *Neurospora*.

To determine whether the transcriptional effect by codon usage is due to changes in DNA sequences that may influence transcription efficiency or elongation (40), we compared the transcription efficiency of wild-type or optimized *luc* or *I-SceI* genes in a well-established *Neurospora* in vitro transcription system using linearized DNA as templates (41). Surprisingly, codon optimization of these genes had no effect on transcript abundance in this system (Fig. 5*F*). Together, these results indicate that the effect of codon usage on transcription does not depend on DNA sequences per se.

**H3K9me3 Is Responsible for the Codon Usage-Mediated Transcriptional Suppression of a Subset of Genes.** The fact that codon usage did not affect transcription efficiency in the in vitro transcription system raised the possibility that it may influence chromatin structure in vivo. After treating *Neurospora* cultures with trichostatin A (TSA), an inhibitor of histone deacetylases that inhibits the class I and II histone deacetylase (HDAC) families of enzymes but not class III HDACs, we found that the effects of codon optimization on luciferase mRNA and protein were mostly abolished (*SI Appendix*, Fig. S6 *A* and *B*). TSA treatment resulted in a dramatic up-regulation

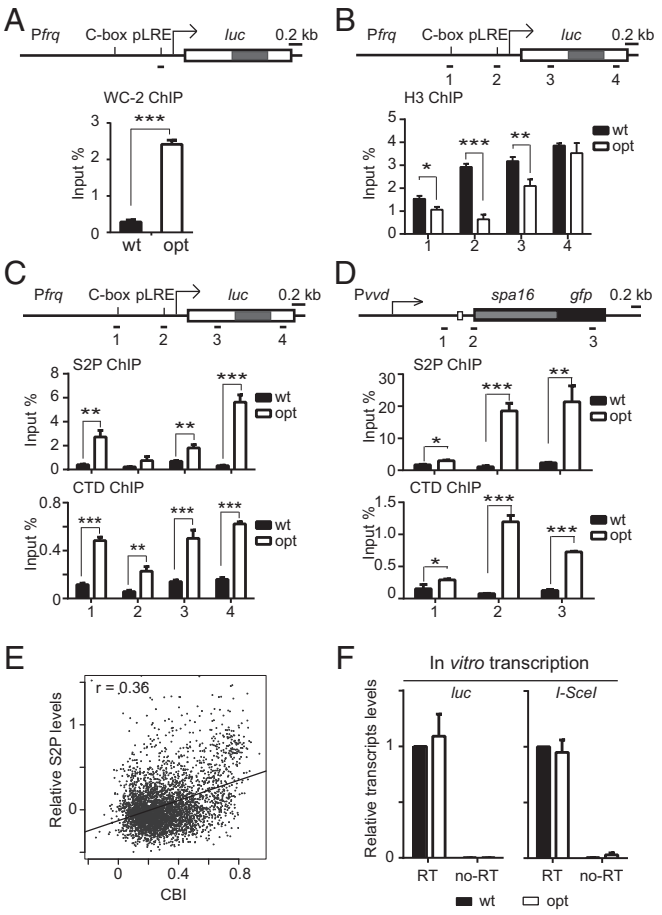


**Fig. 5.** Codon optimization resulted in enrichment of active Pol II. (*A*) WC-2 ChIP assay results showing the relative enrichment of the transcription activator WC-2 at the *frq* promoter in the wt and opt *luc* gene expressed in the *frq* knockout strains. The ChIP results were normalized by input DNA and represented as input%. (*B*) ChIP assay results showing the levels of histone H3 on different regions of the wt and opt *luc* gene. (*C* and *D*) ChIP assays showing the relative enrichment of phosphorylated Ser-2 (S2P) and non-phosphorylated Pol II CTD on the indicated regions. The top of the figure shows the indicated transgene at the *his-3* locus. Within each indicated ORF, the gray regions are encoded by optimized or wild-type codons in the opt or wt strains, respectively. The blank regions in *luc* were codon optimized. (*E*) Scatter analysis showing the correlation of S2P levels with CBI. Pearson's $r = 0.36$. $P < 2.2 \times 10^{-16}$, $n = 4{,}013$. (*F*) qRT-PCR results showing the relative RNA levels after in vitro transcription assays using linearized wt and opt *luc* and *I-SceI* plasmids as templates. Error bars show the SDs of the means ($n = 3$).

of luciferase protein expression in the wild-type *luc* strain but had little influence on protein expression in the optimized *luc* strain. However, TSA did not affect the expression of two endogenous genes (*SI Appendix*, Fig. S5 *C* and *D*). Because TSA was previously shown to cause the loss of DNA methylation, a process that requires the heterochromatin mark histone H3 lysine 9 methylation (H3K9me3) in *Neurospora* (42–44), we examined whether the *luc* transgene locus at this *his-3* locus is associated with H3K9me3 in the wild-type and optimized *luc* strains by ChIP assays (Fig. 6*A* and *SI Appendix*, Figs. S6 *E* and *F* and S7*A*). Only background signals were detected at the *luc* locus in the optimized *luc* strain, but high levels of H3K9me3, similar to those of known heterochromatin regions (*SI Appendix*, Fig. S6*E*), were seen in the strain with the *luc* of wild-type codons. H3K9me3 was not limited to the wild-type *luc* gene region and was also found in the promoter and at the 3′ end of the *luc* gene.

To determine the effect of H3K9me3 in regulating *luc* expression, we then introduced the wild-type and the optimized constructs into the *dim-5*$^{KO}$ strain in which H3K9me3 is completely abolished (43, 45). Remarkably, in *dim-5*$^{KO}$ strain, the effects of codon usage on luciferase protein and RNA were almost completely abolished (Fig. 6*B*), demonstrating that H3K9me3 is responsible for the codon usage effect of the *luc* gene expression.

H3K9me3 was also detected in the *I-SceI* locus in the strain containing the wild-type *I-SceI* transgene (Fig. 6*C* and *SI Appendix*, Fig. S6*G*). As expected, significant reduction of the effects of codon usage on *I-SceI* was observed in the *dim-5*$^{KO}$ strain (Fig. 6*D*). However, codon optimization still resulted in more than 20-fold up-regulation of the *I-SceI* mRNA and protein. These results indicate that, in addition to H3K9me3, additional unidentified mechanism(s) are also responsible for the codon usage effect on the transcription of the *I-SceI* gene. In *Neurospora*, there are two known types H3K9me3 loci. Most of the H3K9me3 sites are within transposon relics of repeat-induced point mutation (RIP) loci (44, 46, 47). In addition, convergent transcription can also trigger H3K9me3 and DNA methylation at certain loci (48). Neither the wild-type *luc* nor the *I-SceI* transgene locus resembles a typical RIP'd locus (*SI Appendix*, Fig. S7 *A–C*) (49) and neither has convergent transcription. Although the wild-type *luc* and *I-SceI* sequences have modestly lower GC content compared with the optimized sequence, other regions with similar levels of GC content around the transgene locus have no detectable H3K9me3 (*SI Appendix*, Fig. S7*A*). This observation is consistent with the bioinformatic results that codon usage, but not GC content, tightly correlates with gene expression levels (Fig. 1). Therefore, an additional mechanism is also responsible for the establishment of H3K9me3 triggered by nonoptimal codon usage.

## Discussion

By codon manipulation in vivo and by examining the genome-wide correlations between codon usage and protein and RNA

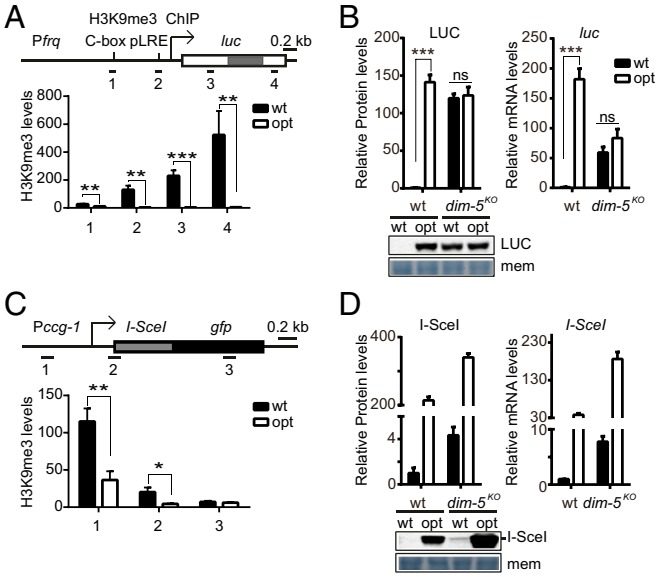


**Fig. 6.** H3K9me3 is responsible for the codon effects of *luc* and *I-SceI* transcription. (*A*) H3K9me3 ChIP assays using anti-H3K9me3 antibody (Active Motif, 39161) showing the relative H3K9me3 levels at the wt and opt *luc* transgene locus (*luc* driven by the *frq* promoter at *his-3* locus in the *frq*$^{KO}$ strain). The enrichment of H3K9me3 was normalized by *tubulin* and represented as relative H3K9me3 levels. (*B*) Comparison of the relative LUC protein and RNA levels between the wt and *dim-5*$^{KO}$ strains. (*C*) ChIP assays showing the relative H3K9me3 levels on the wt and opt *I-SceI* gene driven by the *ccg-1* promoter expressed from the *his-3* locus. (*D*) Comparison of the relative I-SceI protein and RNA levels between the wt and *dim-5*$^{KO}$ strains. Error bars show the SDs of the means ($n = 3$). *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

expression levels, we showed that codon usage is a major determinant of protein expression levels in *Neurospora* through its effects on mRNA levels. Surprisingly, such effects are mostly due to changes in transcription. Furthermore, we identified the chromatin modification H3K9me3 as one of the mechanisms that contributes to the effect of codon usage on transcription.

It was recently shown that codon usage is a major determinant of RNA stability in budding yeast through its effect on translation (23). Both this study and ours demonstrated that there are genome-wide effects of codon usage on mRNA levels. Therefore, in addition to embedded "codes" in protein elongation rates, codon usage biases may represent another "code" within ORF that determines transcript levels by affecting mRNA stability (budding yeast) or transcription efficiency (*Neurospora*). Therefore, codon usage is part of the transcriptional and post-transcriptional mechanisms that control the expression levels of individual genes. Unlike in *S. cerevisiae*, however, codon usage in *Neurospora* does not have consistent impacts on mRNA stability and its effect does not appear to require translation. This difference may be partly contributed by almost opposite codon usage preferences in the two organisms: *S. cerevisiae* prefers A or T at the wobble positions, whereas *Neurospora* strongly prefers C or G.

Codon usage does not have significant effects on mRNA stability for most tested *Neurospora* genes. Consistent with a transcriptional effect of codon usage, it was previously shown that mammalian genes with high GC contents, which means the use of more preferred codons, had higher expression levels than those with lower GC content; this observation was not a result of differences in mRNA degradation rates (50, 51). More recently, codon usage was shown to contribute to the balanced expression of Toll-like receptors by affecting transcription rather than translation in mammals (52).

Our results in *Neurospora* suggest that codon usage of an individual gene is due to coevolution of coding region sequences with transcription and translation machineries. The effect of codon usage on translation elongation and efficiency selected codons that are optimized for accurate and efficient translation and that enhance the cotranslation folding of proteins. However, the demand of optimal protein amount for each protein selected certain codons that are optimized for either activating/suppressing transcription or proper mRNA stability. As a result, codon usage is adapted to both translation and transcription processes; codon information is also read by the transcription machinery in forms of DNA elements, which are used to suppress or activate transcription. Although most known transcriptional regulatory elements reside in the promoter regions, our results demonstrate that the coding sequences can also play a major role in transcriptional regulation. Consistent with this conclusion, it was shown that a significant portion of transcription factor recognition sites reside within plant and human exonic regions, suggesting that the adaptation of coding region sequences to binding of transcription factors is an important evolutionary force that drives codon usage biases (53, 54).

Our results also suggest that codon usage impacts chromatin modifications and that this mechanism is primarily responsible for the codon usage effects we observed on transcription in *Neurospora*. H3K9me3 is one of the mechanisms that suppresses transcription of some endogenous genes with poor codon usage. How genes with poor codon usage result in H3K9me3 is unclear. In *Neurospora*, most known H3K9me3 sites are within transposon relics of repeat-induced point mutation loci (44, 46, 47). It was proposed that these RIP'd sequences recruit chromatin-modifying enzymes to result in de novo H3K9 trimethylation. Sequence analyses of the wild-type *luc* and *I-SceI* genes showed that they are different from typical RIP'd sequences (*SI Appendix*, Fig. S7). In addition, sequences nearby with similar GC contents do not result in H3K9me3. Therefore, it is likely that different mechanisms are involved in H3K9me3 establishment at these gene loci and at the RIP'd loci. Although H3K9me3 is almost completely responsible for the codon usage effects on expression from the *luciferase* gene, it only partially contributes to the codon usage effect of *I-SceI* and other genes and had no impact on some (Fig. 6 and *SI Appendix*, Fig. S6). Therefore, multiple mechanisms mediated by DNA elements specified by codon sequences regulate transcription levels.

## Materials and Methods

**Strains and Culture Conditions.** In this study, FGSC 4200 (*a*) was used as the wild-type strain for the proteomic, RNA-seq, and ChIP-seq analyses. The 301–15 (*bd, his-3, a*), 303–3 (*bd, frq10, his-3*) (55), *pkac-1*$^{KO}$ (*bd, his-3*) (56), and *dim-5*$^{KO}$ (*bd, his-3*) (57) strains were the host strain for *his-3* targeting constructs. A *bd ku70*$^{RIP}$ strain was used for the *csr-1* targeting transformation (58).

Culture conditions have been described (59). *Neurospora* mats were cut into discs and transferred to flasks with minimal medium [1× Vogel's, 2% (wt/wt) glucose]. After 24 h, the tissues were harvested. To induce the expression of *pkac-1*, liquid cultures were grown in ($10^{-5}$ M) quinic acid, pH 5.8, 1× Vogel's, 0.1% glucose, and 0.17% arginine. To induce the expression of *spa16*, discs were cultured in constant dark for 24 h and then transferred to light for 1 h before harvest (experiment in Fig. 2*E*); discs were cultured in constant light for 24 h before harvest (experiments in Figs. 4*F* and 5*D* and *SI Appendix*, Fig. S3*G*). For TSA treatment, $5 \times 10^6$ fresh conidia were directly inoculated into minimal medium with or without 2 μg/mL TSA (42). The tissues were harvested after 24 h, and protein and RNA analyses were performed as described below.

**Codon Optimization, Plasmid Constructs, and *Neurospora* Transformation.** Codon optimization was performed as described (13). Codons were optimized based on the *N. crassa* codon-usage frequency, and the codons in the optimized region were changed to the most preferred codon without changing amino acid sequences. For the optimized *luciferase* gene, all codons (550 codons) were most preferred codons (12). The middle region of the optimized *luc* gene (nucleotides 670–1292) was replaced with original firefly codons, and was used as wild-type *luc* in this study. The gene regions optimized are as follows: *I-SceI*, nucleotides 6–678 (of 678 nt in ORF); *pkac-1*, nucleotides 226–954 (of 1,787 nt in ORF); *spa16*, nucleotides 31–1794 (of 1,797 nt in ORF); *NCU02621*, nucleotides 31–756 and 856–1941 (of 2,127 nt in ORF); *NCU03855*, nucleotides 742–1509 (of 1,920 nt in ORF); *NCU05196*, nucleotides 34–564 and 1267–1569 (of 1,593 nt in ORF); *NCU05881*, nucleotides 31–465, 520–603, 814–1008, and 1219–2103 (of 2,103 in ORF); *spa1*, 31–858, and 1087–1272 (of 1,707 in ORF); *spa8*, 34–1788 (of 1,794 in ORF).

The pMF272.LUC-M-wt and pMF272.LUC-opt constructs, in which the *luc* gene was driven by the *ccg-1* promoter with a *his-3* targeting sequence, were generated (12). The PCR fragments containing the *ccg-1* promoter and wild-type or optimized *luc* ORF were inserted into pCSR1 (58) between NotI and EcoRI sites to generate the pCSR1.LUC-M-wt and pCSR1.LUC-opt constructs. The *frq* promoter was amplified and inserted into pBM61 (60) by using the NotI and XbaI sites to generate the pBM61.*frq* construct. The ORF of the wild-type or optimized *luc* was inserted into pBM61.*frq* between XbaI and SmaI sites to generate the pBM61.*frq*.LUC-M-wt and pBM61.*frq*.LUC-opt constructs. The suboptimal *luc* gene was synthesized by Genscript and inserted into pBM61.*frq* to create pBM61.*frq*.LUC-subopt construct. The construct pqa-5Myc-6His-PKAC-1 was generated (56). The optimized region of *pkac-1* was synthesized (Genscript) and used to replace the corresponding region of the pqa-5Myc-6His-PKAC-1 by using a homologous recombination-based cloning method (In-Fusion HD cloning kit; Clontech) to generate pqa-5Myc-6His-PKAC-1-opt. To create pMF272-Myc, a DNA fragment encoding five copies of the c-Myc peptide tag was added at the 3′ end of the GFP sequence in the plasmid pMF272 (61), which contains the *ccg-1* promoter and results in a GFP tag at the C-terminal end of the protein of interest. The pqa-5Myc-6His-I-SceI-wt and pqa-5Myc-6His-I-SceI-opt constructs were previously generated (62). PCR fragments containing I-SceI-wt or I-SceI-opt ORF were inserted into pMF272-Myc between XbaI and XmaI sites to generate the pMF272-Myc-I-SceI-wt and pMF272-Myc-I-SceI-opt constructs. The suboptimal *I-SceI* gene was synthesized by Genscript and inserted into pMF272-Myc to create pMF272-Myc-I-SceI-subopt construct. The cDNAs for *NCU02621*, *NCU03855*, *NCU05196*, *NCU05881*, *spa1*, *spa8*, and *spa16* were obtained by RT-PCR and inserted into the pMF272-Myc vector. Part or all of the wild-type ORFs of each of these seven genes were replaced by the synthesized fragments containing optimized codons (Genscript) using appropriate cutting sites. The *vvd* promoter was amplified and inserted into pBM61 (60) by using the NotI

and XbaI sites to generate the pBM61.*vvd* construct. The ORF of the wild-type or optimized *spa16* were inserted into pBM61.*vvd* between SpeI and EcoRI sites to generate the pBM61.*vvd.spa16*-wt and pBM61.*vvd.spa16*-opt constructs. The second intron of *pkac-1* ORF was amplified and inserted into the 5′ UTR of the *vvd* promoter of pBM61.*vvd.spa16*-wt and pBM61.*vvd.spa16*-opt constructs by using In-Fusion HD cloning kit (Clontech). The pMF272.LUC-M-wt-stop, pMF272.LUC-opt-stop, pMF272-Myc-I-SceI-wt-stop, and pMF272-Myc-I-SceI-opt-stop constructs were generated by site-directed mutagenesis. To generate pMF272-Myc-I-SceI-wt-stem loop, and pMF272-Myc-I-SceI-opt-stem loop constructs, the stem loop was inserted into the 5′ UTR of the *ccg-1* promoter as described (35). The resulting constructs were transformed into the host strains by electroporation as described (58, 63). Homokaryotic transformants were obtained by microconidia purification and confirmed by quantitative PCR or Southern blot analysis. The strains used in this study were listed in *SI Appendix*, Table S2.

**Protein and Proteomic Analyses.** Tissue harvest, protein extraction, and Western blot analysis were performed as described (64–66). For Western blot analyses, equal amounts of total protein (50 μg) were loaded in each lane. After electrophoresis, proteins were transferred onto PVDF membrane, and Western blot analysis was performed. Anti-luciferase antibody (L2164, Sigma) was used to detect LUC, and anti-Myc antibody (M4439, Sigma) was used to detect Myc-fusion proteins in this study.

For proteomic analyses, the wild-type strain FGSC4200 mats were cut into discs and cultured for 2 d in minimal medium at room temperature. Proteins were extracted by using RIPA buffer (50 mM Tris·HCl, pH 8, 150 mM NaCl, 1% Nonidet P-40, 0.5% sodium deoxycholate, and 0.1% SDS) and precipitated by the addition of trichloroacetic acid/acetone. The pellet was dissolved with urea lysis buffer (8 M urea, 75 mM NaCl, 50 mM Tris, pH 8.2) and digested by using sequencing-grade modified trypsin (Promega) at 37 °C overnight. Tryptic peptides were fractioned by strong cation exchange chromatography. Each fraction was subjected to LC-MS/MS analyses (13), and the emPAI were calculated for each protein (28). A detailed mass spectrometry protocol for *Neurospora* is available upon request from S.C.

**RNA and RNA-seq Analyses.** RNA extraction, qRT-PCR, strand-specific qRT-PCR, and Northern blot were performed as described (67). Total RNA was extracted by using TRIzol and then purified with 2.5 M LiCl (67). cDNA was obtained by reverse transcription using a High-Capacity cDNA Reverse Transcription Kit (ABI) in accordance with the manufacturer's instructions and subjected to real-time PCR analysis. β-*tubulin* was used as internal control. The primer sequences were listed in *SI Appendix*, Table S3.

The mRNA decay assay was performed as described (34). The tissues were harvested at different time points after the addition of thiolutin (final concentration 12 μg/mL), and Northern blot analyses were performed.

The nascent nuclear transcripts were isolated as described (67). Briefly, the nuclei were isolated, and nuclear RNA was extracted by using TRIzol. Contaminating DNA was removed by using TURBO DNase (Ambion), and the resulting RNA was used for qRT-PCR as described above.

The in vitro transcription assay was performed as described (41). Briefly, the transcription extract was prepared from freshly geminated conidia. The reaction mixtures (total volume 25 μL) contained 10 mM K Hepes, pH 7.9, 73 mM potassium acetate, 2.5 mM DTT, 400 mM each GTP, CTP, ATP, and UTP and 2 μL of transcription extract, 0.6 μL of creatine kinase (1 μg/μL), 0.1 μL of creatine phosphate (1 mM), and 0.5 μg of linearized plasmid as template. Reactions were incubated at room temperature for 30 min. The reactions were stopped by the addition of 100 μL of extraction buffer [100 mM sodium acetate, pH 5, 10 mM EDTA, 4% (wt/wt) SDS, and 4 mM urea] and extracted with 500 μL of hot acid phenol once. The supernatant was transferred to a new tube, and 3.5 volumes of pellet buffer [1 M ammonium acetate, 85% (wt/wt) ethanol] and 6 μL of glycoblue were added to precipitate the RNA. Contaminating DNA was removed by TURBO DNase (Ambion), and the resulting RNA was used for qRT-PCR as described above. Samples without reverse transcriptase were also included as negative controls.

For strand-specific mRNA-seq, total mRNA was extracted from the wild-type strain (FGSC 4200). The strand-specific mRNA sequencing library construction and sequencing were performed as described (62). Briefly, total mRNA were purified with NEBnext Oligo $d(T)_{25}$ (NEB), fragmented to an ∼200-bp length and cloned with NEBNext Ultra Directional RNA Library Prep Kit (NEB no. 7420S). Paired-end sequencing was performed by BGI, and short reads were mapped with tophat (version 2.1.1) to *N. crassa* genome (Broad Institute, version 10). Guided by the version 10 annotation, we analyzed the transcriptional level [fragments per kilobase of exon per million fragments mapped (FPKM)] and reannotated the TSS and TES (transcriptional start and transcriptional end sites) based on RNA sequencing result, with Cufflinks (version 2.0.9) as described (68). The GC content of each gene was calculated from TSS to TES, including introns. We also calculated the GC content at the third position of codons for each gene (i.e., total number of codons with G/C at third position divided by total number of codons). For genes with alternative splicing (∼5% of all annotated genes), the first annotated transcripts were used for analyses.

**Chromatin Immunoprecipitation Assay and ChIP-seq Analyses.** ChIP assays were performed as described with some modifications (67, 69). The tissues were cultured in 50 mL of minimal medium and were fixed by adding 1% formaldehyde for 15 min at room temperature. The tissues were harvested and grounded in the liquid nitrogen. One hundred to two hundred microliters of tissue powder was suspended in 300 μL of ChIP lysis buffer. Chromatin was sheared to approximately 500 bp by sonication. In each reaction, 500 μg of total lysate and 2 μL of antibody were added and incubated at 4 °C overnight. The antibodies used in this study were as follows: WC-2 (65), Pol II S2P (Abcam; ab5095), Pol II CTD (Abcam; ab26721), H3 (Abcam; ab1791), H3K9me3 (Active Motif; 39161), and Flag (Sigma; F3165, as negative control). Fifty micrograms of total lysate was saved as input. Twenty-five microliters of G protein coupled beads were added and incubated for 2 h. The beads washed at 4 °C for 5 min with the following buffers: ChIP lysis buffer, low salt buffer, high salt buffer, LNDET (0.25 M LiCl, 1% NP40, 1% deoxycholate, 1 mM EDTA) buffer, and twice with 10 mM Tris, 1 mM EDTA buffer. One hundred forty microliters of 10% chelex beads (Sigma) were added and heated at 96 °C for 20 min. After centrifugation, 100 μL of supernatant was transferred to a new tube. The input DNA was decross-linked and extracted by phenol. Immunoprecipitated DNA was quantified by using real-time quantitative PCR. For CTD, S2P, H3, and WC-2 ChIP, the results were normalized by input DNA and presented as input %. For H3K9me3 ChIP, the results were further normalized by the internal control *tubulin* and represented as relative H3K9me3 levels. The primers used were listed in *SI Appendix*, Table S3. Each experiment was performed independently three times.

For semiquantitative PCR, 1 μL of immunoprecipitated or input DNA was used as the template. Primer sets for both the gene of interest and *tubulin* were added to the same PCR. PCR condition was as follows: 4 min at 94 °C and 26 circles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s. PCR products were resolved by electrophoresis on 2% (wt/wt) argrose gels.

For ChIP-seq analyses, nuclear lysate was used. The nuclei were extracted as described above and resuspended in 300 μL of ChIP lysis buffer. Chromatin was sheared by sonication to ∼200- to 500-bp fragments. For each immunoprecipitation assay, total volume was 200 μL, and 50 μL of nuclear lysate and 2 μL of Pol II S2P or Pol II CTD were used for each reaction. The precipitated DNA was used to make the sequencing library in accordance with the manufacturer's protocol (iDeal Library Preparation Kit; Diagenode). Illumina sequencing was performed by BGI using an Illumina HiSEq 2000 platform.

**Data Analyses.** Raw reads from both ChIP-seq and RNA-seq were generated by using HiSeq 2000. The raw ChIP-seq data were aligned against *Neurospora crassa* genome (version 10, *Neurospora cassa* Database; Broad Institute) with bowtie (version 1.1.1). Two mismatches were allowed and the setting (-m 1–best–strata) was used to allow only one best hit for each read. PCR duplicates were then removed from alignment results with Samtools. The bedgraph files were generated with bedtools for visualization on IGV. To estimate the relative ChIP signal of each gene, we counted the number of reads within the span of annotated genes from the ChIP sample and normalized the counts based on the size of the library (i.e., RPM), which was then divided by the counts from control sample (input) processed in parallel. Raw mRNA-seq data were aligned to *N. crassa* genome with tophat (version 2.0.13), and then processed with Cufflinks (version 2.1.1) as described to obtain the gene expression data (68). The mRNA level (RPKM), protein level [exponentially modified protein abundance index (emPAI)], and relative ChIP level were $\log_{10}$-transformed.

**ACKNOWLEDGMENTS.** We thank Drs. Bing Li, Chen-Ming Chiang, and Shwu-Yuan Wu for helpful discussion; Dr. Qun He for providing the *dim-5* $^{KO}$ strains; and the members of our laboratory for technical assistance. This work is supported by National Institutes of Health Grants GM068496, GM084283, and 1R35GM118118 and Welch Foundation Grant I-1560 (to Y.L.).

1. Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 2(1):13–34.
2. Sharp PM, Tuohy TM, Mosurski KR (1986) Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res* 14(13):5125–5143.
3. Comeron JM (2004) Selective and mutational patterns associated with gene expression in humans: Influences on synonymous composition and intron presence. *Genetics* 167(3):1293–1304.
4. Plotkin JB, Kudla G (2011) Synonymous but not the same: The causes and consequences of codon bias. *Nat Rev Genet* 12(1):32–42.
5. Gingold H, Pilpel Y (2011) Determinants of translation efficiency and accuracy. *Mol Syst Biol* 7:481.
6. Akashi H (1994) Synonymous codon usage in Drosophila melanogaster: Natural selection and translational accuracy. *Genetics* 136(3):927–935.
7. Hershberg R, Petrov DA (2008) Selection on codon bias. *Annu Rev Genet* 42:287–299.
8. Drummond DA, Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134(2):341–352.
9. Qian W, Yang JR, Pearson NM, Maclean C, Zhang J (2012) Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet* 8(3):e1002603.
10. Spencer PS, Siller E, Anderson JF, Barral JM (2012) Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J Mol Biol* 422(3): 328–335.
11. Pechmann S, Chartron JW, Frydman J (2014) Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. *Nat Struct Mol Biol* 21(12):1100–1105.
12. Yu CH, et al. (2015) Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol Cell* 59(5):744–754.
13. Zhou M, et al. (2013) Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* 495(7439):111–115.
14. Zhou T, Weems M, Wilke CO (2009) Translationally optimal codons associate with structurally sensitive sites in proteins. *Mol Biol Evol* 26(7):1571–1580.
15. Pechmann S, Frydman J (2013) Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol* 20(2):237–243.
16. Zhou M, Wang T, Fu J, Xiao G, Liu Y (2015) Nonoptimal codon usage influences protein structure in intrinsically disordered regions. *Mol Microbiol* 97(5):974–987.
17. Quax TE, Claassens NJ, Söll D, van der Oost J (2015) Codon bias as a means to fine-tune gene expression. *Mol Cell* 59(2):149–161.
18. Duret L, Mouchiroud D (1999) Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. *Proc Natl Acad Sci USA* 96(8):4482–4487.
19. Hiraoka Y, Kawamata K, Haraguchi T, Chikashige Y (2009) Codon usage bias is correlated with gene expression levels in the fission yeast Schizosaccharomyces pombe. *Genes Cells* 14(4):499–509.
20. Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science* 324(5924):255–258.
21. Pop C, et al. (2014) Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol Syst Biol* 10:770.
22. Tuller T, et al. (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141(2):344–354.
23. Presnyak V, et al. (2015) Codon optimality is a major determinant of mRNA stability. *Cell* 160(6):1111–1124.
24. Boël G, et al. (2016) Codon influence on protein expression in E. coli correlates with mRNA levels. *Nature* 529(7586):358–363.
25. Radford A, Parish JH (1997) The genome and genes of Neurospora crassa. *Fungal Genet Biol* 21(3):258–266.
26. Gooch VD, et al. (2008) Fully codon-optimized luciferase uncovers novel temperature characteristics of the Neurospora clock. *Eukaryot Cell* 7(1):28–37.
27. Morgan LW, Greene AV, Bell-Pedersen D (2003) Circadian and light-induced expression of luciferase in Neurospora crassa. *Fungal Genet Biol* 38(3):327–332.
28. Ishihama Y, et al. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* 4(9):1265–1272.
29. Bennetzen JL, Hall BD (1982) Codon selection in yeast. *J Biol Chem* 257(6):3026–3031.
30. Kasuga T, Mannhaupt G, Glass NL (2009) Relationship between phylogenetic distribution and genomic features in Neurospora crassa. *PLoS One* 4(4):e5286.
31. Hurley JM, Chen CH, Loros JJ, Dunlap JC (2012) Light-inducible system for tunable protein expression in Neurospora crassa. *G3 (Bethesda)* 2(10):1207–1212.
32. Mishima Y, Tomari Y (2016) Codon usage and 3′ UTR length determine maternal mRNA stability in zebrafish. *Mol Cell* 61(6):874–885.
33. Bazzini AA, et al. (2016) Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *EMBO J,* e201694699.
34. Guo J, Cheng P, Yuan H, Liu Y (2009) The exosome regulates circadian gene expression in a posttranscriptional negative feedback loop. *Cell* 138(6):1236–1246.
35. Doma MK, Parker R (2006) Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* 440(7083):561–564.
36. Froehlich AC, Liu Y, Loros JJ, Dunlap JC (2002) White Collar-1, a circadian blue light photoreceptor, binding to the frequency promoter. *Science* 297(5582):815–819.
37. Cha J, Zhou M, Liu Y (2013) CATP is a critical component of the Neurospora circadian clock by regulating the nucleosome occupancy rhythm at the frequency locus. *EMBO Rep* 14(10):923–930.
38. Komarnitsky P, Cho EJ, Buratowski S (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* 14(19):2452–2460.
39. Hsin JP, Manley JL (2012) The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* 26(19):2119–2137.
40. Zamft B, Bintu L, Ishibashi T, Bustamante C (2012) Nascent RNA structure modulates the transcriptional dynamics of RNA polymerases. *Proc Natl Acad Sci USA* 109(23): 8948–8953.
41. Tyler BM, Giles NH (1985) Accurate transcription of cloned Neurospora RNA polymerase II-dependent genes in vitro by homologous soluble extracts. *Proc Natl Acad Sci USA* 82(16):5450–5454.
42. Selker EU (1998) Trichostatin A causes selective loss of DNA methylation in Neurospora. *Proc Natl Acad Sci USA* 95(16):9430–9435.
43. Tamaru H, Selker EU (2001) A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. *Nature* 414(6861):277–283.
44. Lewis ZA, et al. (2009) Relics of repeat-induced point mutation direct heterochromatin formation in Neurospora crassa. *Genome Res* 19(3):427–437.
45. Lewis ZA, Adhvaryu KK, Honda S, Shiver AL, Selker EU (2010) Identification of DIM-7, a protein required to target the DIM-5 H3 methyltransferase to chromatin. *Proc Natl Acad Sci USA* 107(18):8310–8315.
46. Rountree MR, Selker EU (2010) DNA methylation and the formation of heterochromatin in Neurospora crassa. *Heredity (Edinb)* 105(1):38–44.
47. Selker EU, Fritz DY, Singer MJ (1993) Dense nonsymmetrical DNA methylation resulting from repeat-induced point mutation in Neurospora. *Science* 262(5140): 1724–1728.
48. Dang Y, Li L, Guo W, Xue Z, Liu Y (2013) Convergent transcription induces dynamic DNA methylation at disiRNA loci. *PLoS Genet* 9(9):e1003761.
49. Margolin BS, et al. (1998) A methylated Neurospora 5S rRNA pseudogene contains a transposable element inactivated by repeat-induced point mutation. *Genetics* 149(4): 1787–1797.
50. Kudla G, Lipinski L, Caffin F, Helwak A, Zylicz M (2006) High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biol* 4(6):e180.
51. Krinner S, et al. (2014) CpG domains downstream of TSSs promote high levels of gene expression. *Nucleic Acids Res* 42(6):3551–3564.
52. Newman ZR, Young JM, Ingolia NT, Barton GM (2016) Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9. *Proc Natl Acad Sci USA* 113(10):E1362–E1371.
53. Stergachis AB, et al. (2013) Exonic transcription factor binding directs codon choice and affects protein evolution. *Science* 342(6164):1367–1372.
54. Sullivan AM, et al. (2014) Mapping and dynamics of regulatory DNA and transcription factor networks in A. thaliana. *Cell Reports* 8(6):2015–2030.
55. Cha J, Yuan H, Liu Y (2011) Regulation of the activity and cellular localization of the circadian clock protein FRQ. *J Biol Chem* 286(13):11469–11478.
56. Huang G, et al. (2007) Protein kinase A and casein kinases mediate sequential phosphorylation events in the circadian negative feedback loop. *Genes Dev* 21(24): 3283–3295.
57. Xu H, et al. (2010) DCAF26, an adaptor protein of Cul4-based E3, is essential for DNA methylation in Neurospora crassa. *PLoS Genet* 6(9):e1001132.
58. Bardiya N, Shiu PK (2007) Cyclosporin A-resistance based gene placement system for Neurospora crassa. *Fungal Genet Biol* 44(5):307–314.
59. Aronson BD, Johnson KA, Loros JJ, Dunlap JC (1994) Negative feedback defining a circadian clock: Autoregulation of the clock gene frequency. *Science* 263(5153): 1578–1584.
60. Honda S, Selker EU (2009) Tools for fungal proteomics: Multifunctional neurospora vectors for gene replacement, protein expression and protein purification. *Genetics* 182(1):11–23.
61. Freitag M, Hickey PC, Raju NB, Selker EU, Read ND (2004) GFP as a tool to analyze the organization, dynamics and function of nuclei and microtubules in Neurospora crassa. *Fungal Genet Biol* 41(10):897–910.
62. Yang Q, Ye QA, Liu Y (2015) Mechanism of siRNA production from repetitive DNA. *Genes Dev* 29(5):526–537.
63. Bell-Pedersen D, Dunlap JC, Loros JJ (1996) Distinct cis-acting elements mediate clock, light, and developmental regulation of the *Neurospora crassa eas* (*ccg-2*) gene. *Mol Cell Biol* 16(2):513–521.
64. Garceau NY, Liu Y, Loros JJ, Dunlap JC (1997) Alternative initiation of translation and time-specific phosphorylation yield multiple forms of the essential clock protein FREQUENCY. *Cell* 89(3):469–476.
65. Cheng P, Yang Y, Heintzen C, Liu Y (2001) Coiled-coil domain-mediated FRQ-FRQ interaction is essential for its circadian clock function in *Neurospora*. *EMBO J* 20(1–2): 101–108.
66. Zhou Z, Wang Y, Cai G, He Q (2012) Neurospora COP9 signalosome integrity plays major roles for hyphal growth, conidial development, and circadian function. *PLoS Genet* 8(5):e1002712.
67. Xue Z, et al. (2014) Transcriptional interference by antisense RNA is required for circadian clock function. *Nature* 514(7524):650–653.
68. Trapnell C, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7(3):562–578.
69. Zhou Z, et al. (2013) Suppression of WC-independent frequency transcription by RCO-1 is essential for Neurospora circadian clock. *Proc Natl Acad Sci USA* 110(50): E4867–E4874.