# An RNA structure-mediated, posttranscriptional model of human α-1-antitrypsin expression

Meredith Corley[a,b], Amanda Solem[a], Gabriela Phillips[a], Lela Lackey[a], Benjamin Ziehr[c,d], Heather A. Vincent[c,d], Anthony M. Mustoe[e], Silvia B. V. Ramos[f], Kevin M. Weeks[e], Nathaniel J. Moorman[c,d], and Alain Laederach[a,b,1]

[a]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; [b]Curriculum in Bioinformatics and Computational Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; [c]Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; [d]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; [e]Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; and [f]Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Edited by Rachel Green, Johns Hopkins University, Baltimore, MD, and approved October 19, 2017 (received for review April 20, 2017)

**Chronic obstructive pulmonary disease (COPD) affects over 65 million individuals worldwide, where α-1-antitrypsin deficiency is a major genetic cause of the disease. The α-1-antitrypsin gene, *SERPINA1*, expresses an exceptional number of mRNA isoforms generated entirely by alternative splicing in the 5′-untranslated region (5′-UTR). Although all *SERPINA1* mRNAs encode exactly the same protein, expression levels of the individual mRNAs vary substantially in different human tissues. We hypothesize that these transcripts behave unequally due to a posttranscriptional regulatory program governed by their distinct 5′-UTRs and that this regulation ultimately determines α-1-antitrypsin expression. Using whole-transcript selective 2′-hydroxyl acylation by primer extension (SHAPE) chemical probing, we show that splicing yields distinct local 5′-UTR secondary structures in *SERPINA1* transcripts. Splicing in the 5′-UTR also changes the inclusion of long upstream ORFs (uORFs). We demonstrate that disrupting the uORFs results in markedly increased translation efficiencies in luciferase reporter assays. These uORF-dependent changes suggest that α-1-antitrypsin protein expression levels are controlled at the posttranscriptional level. A leaky-scanning model of translation based on Kozak translation initiation sequences alone does not adequately explain our quantitative expression data. However, when we incorporate the experimentally derived RNA structure data, the model accurately predicts translation efficiencies in reporter assays and improves α-1-antitrypsin expression prediction in primary human tissues. Our results reveal that RNA structure governs a complex posttranscriptional regulatory program of α-1-antitrypsin expression. Crucially, these findings describe a mechanism by which genetic alterations in noncoding gene regions may result in α-1-antitrypsin deficiency.**

translation efficiency | RNA secondary structure | uORFs | SERPINA1 | α-1-antitrypsin deficiency

Human α-1-antitrypsin is of particular clinical interest because deficiencies in this protein are associated with chronic obstructive pulmonary disease (COPD), liver disease, and asthma (1–4). Smoking is the major environmental factor that contributes to COPD risk, although the inconsistency of COPD rates among smokers points to additional genetic factors that modulate risk (5–7). Multiple genetic variants in the gene encoding α-1-antitrypsin, *SERPINA1*, cause the disease α-1-antitrypsin deficiency (8–10), which can result in COPD, liver failure, and inflammatory conditions like panniculitus, vasculitis, and glomerulonephritis (9, 11, 12). α-1-Antitrypsin is a protease inhibitor that specifically targets neutrophil elastase, which is present at chronic low levels in the lungs (1). Deficiency of α-1-antitrypsin thus results in higher levels of neutrophil elastase, which in turn degrades elastin (especially in the lungs), resulting in COPD (13). Thus, the role of *SERPINA1* in COPD etiology is well described at the protein level; however, little is known about *SERPINA1* at the transcript level and whether alteration of potential posttranscriptional controls can contribute to α-1-antitrypsin deficiency and ultimately COPD. Genome-wide association studies identified COPD-associated variants that map to the *SERPINA1* untranslated regions (UTRs), introns, and promoter region (5, 14). Furthermore, genetic variants shown to alter *SERPINA1* splicing patterns were identified in the *SERPINA1* introns of patients with COPD (15, 16). The presence of disease-associated variants in noncoding regions suggests that posttranscriptional regulation of *SERPINA1* mRNA is an important component of disease risk. Nevertheless, variants in noncoding regions of *SERPINA1* comprise only a small fraction of its disease-associated variants discovered to date, which may reflect the tendency of variant discovery studies to focus exclusively on coding exons (10, 17).

Several features of *SERPINA1* emphasize the importance of its transcripts and their regulation. The *SERPINA1* gene is exceptionally complex; 11 different splicing isoforms occur in human tissues (18). While alternative splicing occurs in 95% of human multiexon genes (19, 20), the 11 *SERPINA1* transcripts are extreme, placing *SERPINA1* in the top 0.5% of human genes in terms of transcriptional complexity (18). A particularly salient feature of *SERPINA1* alternative splicing is that all variants differ only within their 5′-UTRs (21). Therefore, all *SERPINA1* mRNA isoforms code for the same α-1-antitrypsin protein; however, their

## Significance

**Protein and mRNA expression are in most cases poorly correlated, which suggests that the posttranscriptional regulatory program of a cell is an important component of gene expression. This regulatory network is still poorly understood, including how RNA structure quantitatively contributes to translational control. We present here a series of structural and functional experiments that together allow us to derive a quantitative, structure-dependent model of translation that accurately predicts translation efficiency in reporter assays and primary human tissue for a complex and medically important protein, α-1-antitrypsin. Our model demonstrates the importance of accurate, experimentally derived RNA structural models partnered with Kozak sequence information to explain protein expression and suggests a strategy by which α-1-antitrypsin expression may be increased in diseased individuals.**

Author contributions: M.C., A.S., G.P., A.M.M., K.M.W., N.J.M., and A.L. designed research; M.C., A.S., G.P., L.L., B.Z., H.A.V., and S.B.V.R. performed research; M.C. and A.S. analyzed data; and M.C., A.S., B.Z., A.M.M., K.M.W., N.J.M., and A.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (accession no. GSE81525). SHAPE data are also available in SNRNASM format at https://docs.google.com/spreadsheets/d/1_RpB9Jto1-UEmK-ocd9pGMOYrtet1ALuaA7XTyqI8ZA/edit?usp=drive_web.

[1]To whom correspondence should be addressed. Email: alain@unc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1706539114/-/DCSupplemental.

differing 5′-UTRs likely determine transcript-specific differences in posttranscriptional processes such as mRNA translation efficiency, subcellular localization, and stability (22, 23). Importantly, the *SERPINA1* transcript isoforms are differentially expressed across tissue types (24), suggesting that posttranscriptional regulatory mechanisms adjust α-1-antitrypsin production based on the transcripts expressed in each tissue. The presence of up to three upstream ORFs (uORFs) in the *SERPINA1* 5′-UTRs (25, 26) suggests a potentially important yet unstudied mechanism for the translation efficiency regulation of these transcripts. In addition to the sequence-based differences between *SERPINA1* transcripts, RNA secondary structure differences in the 5′-UTR could also determine their regulation (27–29).

We propose here that noncoding features of *SERPINA1* transcripts make up a posttranscriptional regulatory program that ultimately determines α-1-antitrypsin expression. We describe a complex interplay between alternative splicing and translation efficiency mediated by uORFs and RNA structure, which together control tissue-specific expression of α-1-antitrypsin in humans. Our quantitative and predictive model reveals an important and overlooked aspect of α-1-antitrypsin deficiency and suggests RNA-based targets for therapeutic consideration.

## Results

**Transcript Complexity in *SERPINA1*.** As a clinically important gene harboring numerous COPD and α-1-antitrypsin deficiency-associated variants (30) (Fig. 1*A*), *SERPINA1* is of additional interest for the exceptional number of transcript isoforms it produces. Two transcription start sites (TSSs), six splicing donor (SD), and three acceptor (SA) sites yield a total of 11 transcript isoforms (21) (Fig. 1*A* and Fig. S1), which places *SERPINA1* in the top 0.5% of transcriptionally complex human genes (18). Remarkably, all of the alternative splicing occurs in the 5′-UTR of *SERPINA1* mRNA (Fig. 1*A*). Thus, in healthy adults, α-1-antitrypsin exists as a single protein isoform that is produced from 11 different mRNAs. We sought to determine whether the mRNAs are functionally different and how any differences relate to α-1-antitrypsin production or deficiency.

We therefore began this investigation by quantifying the expression of the various *SERPINA1* transcripts in human tissues. Using data from the Illumina BodyMap 2.0 transcriptome-wide RNA-seq project, we quantified the relative amount of total *SERPINA1* transcripts in 16 human tissues (Fig. 1*B*) and we show the relative amount of each *SERPINA1* transcript in the form of a heat map (Fig. 1*C*). There are clear differences in the total amount of *SERPINA1* present in each tissue. Liver noticeably yields the highest total *SERPINA1* read counts (Fig. 1*B*), reflecting that α-1-antitrypsin is primarily expressed by hepatocytes and secreted into the bloodstream (1, 31). While the lungs are thought to acquire α-1-antitrypsin from the bloodstream (1, 32), we found that lung tissue transcribes nontrivial amounts of *SERPINA1* (Fig. 1*B*), thus potentially producing its own α-1-antitrypsin. Although some *SERPINA1* transcript isoforms are more prevalent than others, we detected all of the transcripts, with some tissues like liver expressing every transcript (Fig. 1*C*). To verify these findings with greater specificity, we designed *SERPINA1* 5′-UTR–specific primers and amplified RNA extracted from liver and lung epithelial cells (HepG2 and A549 cell lines, respectively). All 11 transcripts were expressed in HepG2 cells, and all save 1 in A549 cells (Fig. 1*C*, *Bottom*). The varied expression of the *SERPINA1* transcript isoforms across the tissues suggests that each *SERPINA1* transcript has a distinct posttranscriptional function. Given that these transcripts vary only in their 5′-UTR, we hypothesize that the splicing complexity in the 5′-UTR of *SERPINA1* plays an important role in its posttranscriptional regulation, especially, as detailed next, in *SERPINA1* mRNA translation.

**Translation Efficiency Analysis.** The 5′-UTR in an mRNA regulates translation of the coding sequence and ultimately controls the expression of protein products (22). To test the effect of different *SERPINA1* 5′-UTRs on mRNA translation, we measured the translation efficiencies of six representative *SERPINA1* 5′-UTRs with luciferase assays. Strikingly, we found significant differences in translation efficiency for the six *SERPINA1* 5′-UTRs (Fig. 2*A*). Alternative splicing determines the inclusion (or exclusion) of up to three uORFs in the final *SERPINA1* transcript isoform (26) (Figs. 1*A* and 2*B*, and Fig. S1). Because uORFs can affect translation efficiency (33, 34), the uORFs in *SERPINA1* may modulate translation of the different transcripts [an idea acknowledged decades ago (25) but untested until now]. To evaluate the effect of uORFs on *SERPINA1* translation, we mutated the start codon of a single uORF in each luciferase construct from "AUG" to "AAG" (Fig. 2*B*). In this group of mutants, we mutated every possible *SERPINA1* uORF in at least one construct. Although it is possible that translation initiation at the mutated start codons could still occur (35), the initiation efficiency of an "AAG" start codon is very low—between 0 and 3% (36).

Mutating the uORF start codon(s) resulted in large increases in the translation efficiency of three of the six transcripts (Fig. 2*C*), suggesting that these uORFs typically inhibit translation. The three transcripts with inhibitory uORFs are NM_000295.4, NM_001002236.2, and NM_001127705.1, and their mutated uORFs were uORFγ, uORFδ, and uORFδ′, respectively (Fig. 2*B*). Interestingly, uORFγ is too close to the mRNA 5′-terminus to be translated based on canonical understanding of translation initiation (37). However, our luciferase assays clearly suggest that it is functional, as it both significantly represses translation of

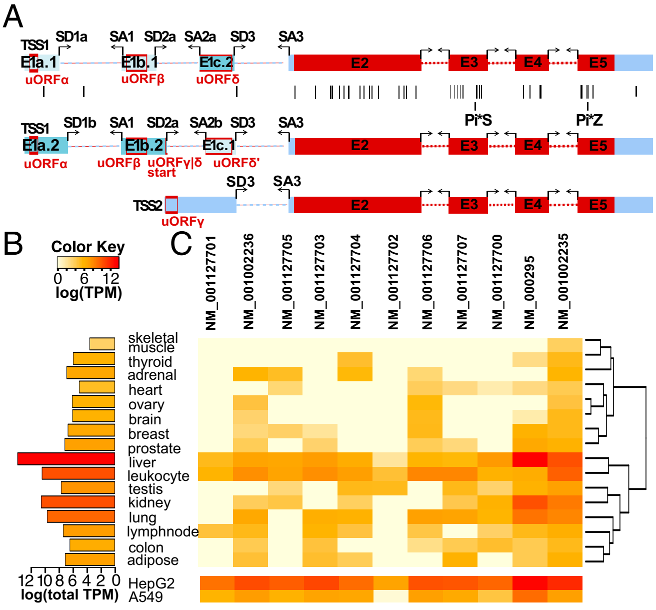
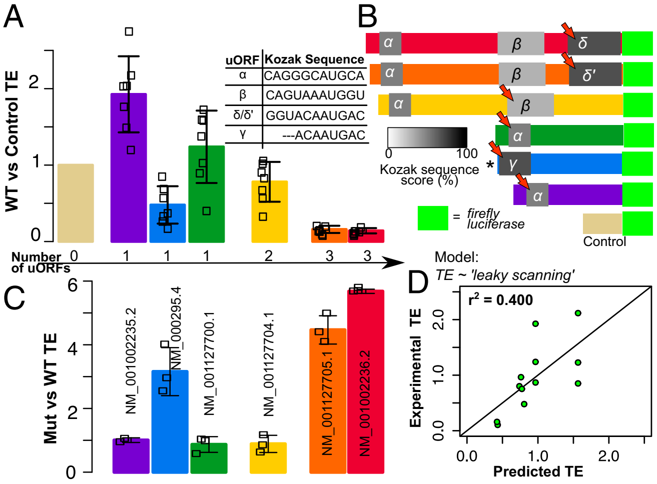


**Fig. 1.** The *SERPINA1* gene produces 11 splice isoforms, all encoding the same protein. (*A*) All exons in *SERPINA1*. Coding sequence (CDS) exons are shown in red, and untranslated regions (UTRs) in blue. Each exon, splice donor (SD), and splice acceptor (SA) is identified by a unique name. The two *SERPINA1* TSSs are labeled TSS1 and TSS2. Disease-associated variants, as cataloged by the Human Gene Mutation Database, are indicated with black lines, including the common α-1-antitrypsin deficiency-associated Pi*S and Pi*Z alleles. Upstream ORFs (uORFs) are indicated by red boxes and named. uORF δ/δ′ spans a splice junction and is present only in isoforms with exon E1b.2. (*B*) The total amount of expressed *SERPINA1* differs across 16 human tissue types. Total *SERPINA1* transcript amounts were estimated from the Illumina BodyMap 2.0 project and are shown in log relative transcripts per million (TPM). (*C*) The *SERPINA1* transcript isoforms are expressed, with different frequencies, across different tissues. Transcripts are specified with their NCBI names. The log(TPM) of each *SERPINA1* transcript is shown for each tissue and for A549 and HepG2 cells. TPMs are relative to liver, which expresses the most *SERPINA1* and is set to a total of $10^6$.



| uORF | Kozak Sequence |
|---|---|
| α | CAGGGCAUGCA |
| β | CAGUAAAUGGU |
| δ/δ' | GGUACAAUGAC |
| γ | ---ACAAUGAC |

**Fig. 2.** Translation efficiency (TE) differs between *SERPINA1* transcripts and is affected by uORFs. (*A*) The TEs of six *SERPINA1* 5′-UTRs and their SDs, as measured by luciferase reporter assays. Replicate TE values are shown as open squares. Transcripts are labeled by NCBI name. Measurements are relative to the luciferase assay control. The number of uORFs in each transcript is indicated (*Bottom*). The Kozak sequence of each uORF is listed. (*B*) Schematic of the *SERPINA1* luciferase constructs and empty vector control. Luciferase CDS not to scale. uORFs in each transcript are indicated with Greek letters and shaded by Kozak sequence score (see color scale). Red arrows indicate uORFs selected for mutation. (*C*) TEs of the six *SERPINA1* constructs with disrupted (mutated) uORFs and their SDs, relative to the wild type (above). (*D*) TEs of wild type and uORF mutant *SERPINA1* constructs predicted with a leaky-scanning model of translation (Eq. **1**) fit to experimental TEs, as measured by luciferase assays ($r^2$ = 0.400, $n$ = 12).

transcript NM_000295.4 relative to other single-uORF transcripts (Fig. 2*A*), and relieves translational inhibition when it is mutated (Fig. 2*C*). The use of an alternate upstream TSS in the luciferase reporter construct likely accounts for the translation of uORFγ in our assays (38). 5′-RACE indicates variability of start site usage in the constructs, including transcripts with additional 5′ sequence (Fig. S3).

**Modeling Translation Efficiency.** The inhibitory uORFγ, uORFδ, and uORFδ′ uORFs identified above all have different sequences. However, closer inspection revealed that uORFγ, uORFδ, and uORFδ′ share highly similar Kozak sequences (Fig. 2*A*), the well-characterized sequence element that determines translation initiation efficiency (39, 40). Indeed, uORFγ, uORFδ, and uORFδ′ have much stronger Kozak sequences compared with the *SERPINA1* uORFs determined to be nonfunctional in our luciferase assays (40) (grayscale in Fig. 2*C*).

Confident that uORFs and their Kozak sequences play an important role in regulating translation, we next used uORF Kozak sequence strengths (40) to model the differences in translation efficiency between the *SERPINA1* transcripts. We first modeled translation efficiency with a previously derived "leaky-scanning" model of translation (34), which we expanded to accommodate multiple nonoverlapping uORFs (*Methods* and Eq. **1**). The model assumes the scanning mechanism of translation, whereby ribosomes migrate along the 5′-UTR until encountering a start codon, and calculates the probability that ribosomes "leak through" any uORF to ultimately translate the primary coding sequence (41). The leaky-scanning model is based solely on the strength of the Kozak sequence of each ORF. The leaky-scanning model moderately predicted the translation efficiencies given by our luciferase assay data ($r^2 = 0.40$; Fig. 2*D*). To control for potential inaccuracies in the Kozak sequence strengths (40), we repeated the leaky-scanning model analysis using a 95% confidence interval range for each Kozak sequence strength, but this adjustment only increased the $r^2$ value to 0.46 at most. Other features in the *SERPINA1* transcripts beyond Kozak sequence thus heavily influence their translation.

As an alternative model that may explain our translation efficiency measurements, we considered the variable TSS usage indicated by 5′-RACE (Fig. S3). A significantly shorter 5′-UTR could remove uORFγ and potentially uORFα. If we adjust predictors in the leaky-scanning model (Eq. **1**) to ignore these uORFs, we see no improvement in fit ($r^2$ = 0.42, Table S1). Another factor that could affect translation efficiency is reinitiation after uORF translation (42). We therefore fit a "reinitiation leaky-scanning" model (Eq. **2**) to the experimental translation efficiencies, but observed no improvement ($r^2$ = 0.33, Table S1). The rules that govern uORF reinitiation are admittedly poorly understood (43). It is possible that uORFα, which has a strong Kozak sequence (Fig. 2*B* and Dataset S1), nevertheless fails to inhibit coding sequence (CDS) translation (Fig. 2) due to efficient reinitiation after uORFα translation. Adjusting uORFα Kozak strength in the leaky-scanning model to reflect this idea, we observe a moderate improvement in fit ($r^2$ = 0.60, Table S1). An additional factor that can modulate translation efficiency is mRNA secondary structure. Evidence for the effect of secondary structure on translation has been conflicting (27–29, 34, 40), but such studies have typically relied on theoretical structure prediction, which falls far short of the accuracy achieved with direct chemical probing experiments (44, 45). We next sought to predict translation efficiency for the *SERPINA1* transcripts using secondary-structure features derived from chemical structure probing.

**Secondary Structure of *SERPINA1* Transcripts.** Recent advances in RNA structural mapping techniques, in particular selective 2′-hydroxyl acylation by primer extension and mutational profiling (SHAPE-MaP) (46), have enabled accurate, high-throughput, whole-transcript structural interrogation of RNA (47–49). SHAPE-MaP interrogates the reactivity of each 2′-hydroxyl in an RNA toward the reagent 1-methyl-7-nitroisatoic anhydride where the relative reactivity estimates the tendency of each nucleotide to be structured (i.e., base paired) or unstructured (i.e., unpaired). To measure structure differences between the *SERPINA1* transcripts, we performed SHAPE-MaP separately on the six *SERPINA1* transcript isoforms whose 5′-UTRs were examined in luciferase assays. The resulting data are highly correlated between replicates (Fig. S4), with average correlation coefficients of 0.89 or more. Our experimental SHAPE-MaP data provide SHAPE reactivity profiles at nucleotide resolution for each of the six *SERPINA1* transcripts. Regions with lower median SHAPE values (low SHAPE reactivities) consist of largely unreactive nucleotides (Fig. 3*A*), whereas regions with higher median SHAPE values indicate the reverse (Fig. 3*B*). The median-centered SHAPE reactivities of each transcript illustrate the relative reactivity of regions in the transcripts and indicate structured regions (Fig. 3*C*). The high-reproducibility of SHAPE-MaP is immediately apparent in the median-centered SHAPE profiles: the reactivity patterns in the coding sequences (CDSs) are nearly identical across the six transcripts, corresponding to the transcripts' identical CDS sequences (Fig. 3*C* and Fig. S5). In addition, shared exons in the 5′-UTRs also exhibit comparable SHAPE reactivities despite existing in unique contexts in the different transcripts (Fig. S5).

We next used our SHAPE-MaP data to derive minimum free-energy structure models for the six *SERPINA1* transcripts (44, 50). SHAPE reactivities were incorporated as pseudo–free-energy terms to guide RNA structure modeling with RNAfold (51). Importantly, this approach has been extensively validated and generally yields structure models with accuracies above 90% (44, 46, 51, 52). Even in the case where there is not SHAPE data for the entirety of an RNA (as is common at the ends of transcripts), incorporating available SHAPE data still greatly improves the accuracy of structure predictions (53, 54). As an internal control, we initially compared the structure

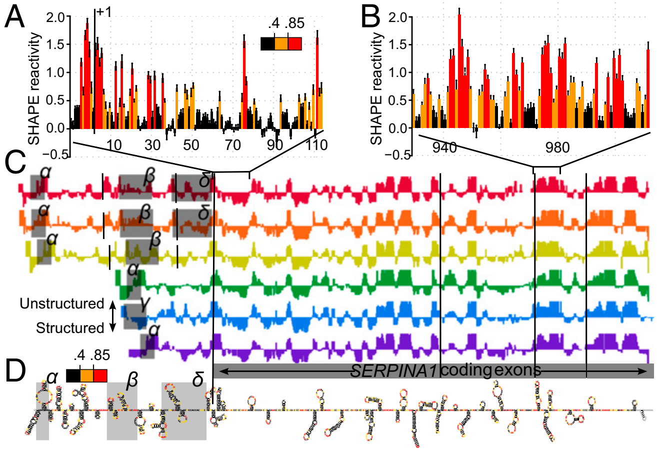


**Fig. 3.** SHAPE-MaP structure probing data for *SERPINA1* transcripts. (*A*) SHAPE reactivity of each nucleotide in a region of low median SHAPE values around the start codon of transcript NM_001002236.2. Each value is shown with its SE and colored by SHAPE reactivity according to the color scale. Nucleotides are numbered by their relative position within the transcript; the start codon is labeled +1. (*B*) SHAPE reactivity of each position in a region of high median SHAPE values in the coding sequence of transcript NM_001002236.2. (*C*) The windowed, median-centered SHAPE profiles of six *SERPINA1* transcripts ordered by length. Higher SHAPE values indicate unstructured (unpaired) regions, while lower SHAPE values indicate structured (base-paired) regions. uORFs are indicated with gray shaded regions and named with Greek letters. Vertical bars separate exons. (*D*) The minimum free-energy (MFE) secondary structure of transcript NM_001002236.2, modeled by computational folding with SHAPE reactivity information.

models derived for the CDS regions of different transcripts. Consistent with the high correlation observed between SHAPE-MaP profiles, the secondary-structure models are highly similar in the CDS regions, supporting the robustness of the models (Fig. S6). However, we were most interested in the structures around the uORFs and the beginning of the CDS, and how this information could be used to model translation efficiency.

**Modeling Translation Efficiency with Structure.** We next sought to gain a quantitative understanding of the contribution of RNA structure to translation. The interplay of transcript structural elements with the translational machinery is not well understood, although studies in bacterial and mammalian systems suggest that secondary structures near start codons are most likely to affect mRNA translation (27, 28, 55, 56). We established above that the uORFs in *SERPINA1* affect translation efficiency and found that a model that incorporates only Kozak sequence strength did not quantitatively explain a large portion of the translation efficiency differences (Fig. 2*D*). We hypothesized that, in addition to Kozak sequence strength, the model requires structural data encompassing the Kozak sequence to accurately capture the probability of the ribosome initiating at a given ORF. SHAPE-MaP data provided us with a high-confidence structure of each transcript (Fig. 3*D* and Fig. S6), including the structures surrounding each Kozak sequence (Fig. 4 *A* and *B* and Fig. S5). Studies in prokaryotes suggest that translation initiation occurs in proportion to the exponent of the free energy ($\Delta G$) of unfolding of the local structure (57), which is the energy required to "unfold" a region of RNA (and is thus a positive value). We modified the leaky-scanning model from Eq. **1** to include the $\Delta G$ of unfolding around the Kozak sequence (*Methods* and Eq. **3**). The SHAPE data-driven "structure leaky-scanning" model dramatically improves the predictive power of the model to 94% (Fig. 4*C*). The structural terms in the model weigh each Kozak sequence by its accessibility in addition to its strength. From their location in uORF secondary structures, it is immediately clear that not all of the uORF Kozak sequences are equally accessible (Fig. 4 *A* and *B*). For example, the Kozak sequence for uORFδ resides in a single-stranded loop, while the Kozak sequence for uORFα is engaged in a based-paired stem structure. It appears that uORFs δ, δ′, and γ are the only uORFs that have a Kozak sequence that is both strong and structurally accessible (Figs. 2*C* and 4 *A* and *B*, and Fig. S7), potentially explaining why only these uORFs inhibit *SERPINA1* translation in our assays. The specific $\Delta G$ of unfolding associated with each uORF Kozak sequence is important: permuting the $\Delta G$ values' assignments and refitting the structure leaky-scanning model never produces $r^2$ values reaching 0.94 (value of $P < 0.001$). Furthermore, refitting the structure leaky-scanning model using $\Delta G$ values predicted without SHAPE data yields a lower correlation ($r^2 = 0.79$; Dataset S1), supporting the importance of using accurate SHAPE-based structure models. Finally, we also varied the size of the unfolding region around the Kozak sequence used for calculating $\Delta G$ of unfolding values. Supporting the physical relevance of our structure leaky-scanning model, the optimal predictive power was obtained for an unfolding window size of 30 nt, consistent with the known size of the eukaryotic ribosomal footprint (58, 59). Either smaller or larger unfolding regions exhibited significantly worse agreement with the translation efficiency data (Table S2).

It is important to note that our structure models for the *SERPINA1* transcripts focus on local structures (*Methods*). Although long-range interactions in large RNAs can occur, local structure is thought to dominate the folding of mRNAs (60, 61). To explore the possibility of longer-range secondary structures, we recalculated the $\Delta G$ of unfolding values, allowing for greater pairing distances in RNA structure predictions and refit our structure leaky-scanning model in each case (Table S3). Predictive

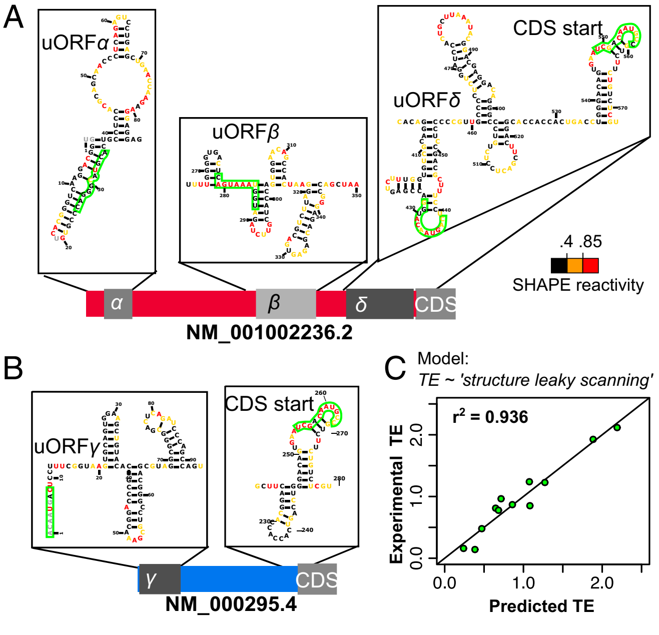


**Fig. 4.** Structural data greatly improve the leaky-scanning model of translation efficiency (TE). (*A*) SHAPE-based predicted structures around the uORFs and coding sequence start in transcript NM_001002236.2. uORFs are labeled by name. Bases are colored according to their SHAPE reactivity, as measured by SHAPE-MaP. Bases with unknown SHAPE data are colored gray. Kozak sequences are outlined in green. (*B*) SHAPE-based predicted structures around the uORF and coding sequence start in transcript NM_000295.4. (*C*) TEs of wild type and uORF mutant *SERPINA1* constructs predicted with the structure leaky-scanning model of translation (Eq. **3**) fit to experimental TEs, as measured by luciferase assays ($r^2 = 0.936$, $n = 12$).

performance of the model generally decreases as the max pairing distance increases (Table S3), suggesting that local structure is most important in determining translation of these transcripts (Table S3). However, we cannot exclude the possibility of long-range interactions. While a few recent structure probing methods can directly detect long-range interactions (62, 63), SHAPE-directed modeling accuracy decreases for long-range interactions, which could also contribute to the decreased performance as max pairing distance increases.

Based on this analysis, we propose that the Kozak sequence determines the likelihood of initiating translation, but the secondary structure determines whether the Kozak sequence can in fact be accessed. Thus, the translation efficiency of each *SERPINA1* transcript is a combination of the initiation strength and structure of its CDS Kozak sequence, attenuated by the translation efficiency of any uORFs as governed by the same parameters.

**Mutating Secondary Structure to Change Translation Efficiency.** While our luciferase assays suggest that little to no translation occurs at uORFα (Fig. 2), available ribosomal profiling data (64) show minimal yet detectable signal at uORFα (Fig. S8). This indicates that uORFα is capable of translation and is thus capable of being translationally regulated, including by structural manipulation. To further assess the role of secondary structure in controlling uORF function, we designed structure mutants for uORFα (in transcript NM_001002235.2). We designed three mutants with low free energies of unfolding to be predominantly single stranded within 30 nt of the uORF Kozak sequence. Mutants contained altered sequences upstream and downstream of the Kozak sequence, without altering the Kozak sequence itself. The wild-type structure of uORFα has an unfolding energy of 22.4 kcal/mol, while the three mutants have unfolding energies below 4 kcal/mol and are expected to enhance the activity of uORFα, thus diminishing translation of the CDS. The translation efficiency of each structure mutant was measured by luciferase assays relative to wild type. As expected, the mutants show reduced translation efficiencies (of the CDS) relative to wild type (Fig. 5*A*) that are consistent with predictions from the structure leaky-scanning model (Fig. 5*B*).

In transcript NM_001002236.2, which contains uORFδ, we designed four mutants with increased energies of unfolding within 30 bases of the uORF Kozak sequence and an additional mutant with greatly decreased Δ*G* of unfolding. As with uORFα, a reduction in the Δ*G* of unfolding causes a reduction in overall translation efficiency relative to wild type (Fig. 5*C* and Fig. S9). Conversely, increasing the Δ*G* of unfolding around uORFδ increases translation efficiency in one structure mutant, but as the structure-mutant energies of unfolding increase above ∼25 kcal/mol, overall translation efficiency begins to decrease (Fig. S9). These results could be interpreted to indicate that the Δ*G* of unfolding is not a significant factor controlling uORFδ translation initiation. Alternatively, the overall decrease in translation efficiency as hairpin sizes increase exactly replicates multiple experiments in which hairpins of increasing size added to the 5′-UTR progressively reduce translation efficiency (39, 65–67). Thus, increasing the Δ*G* of unfolding around a uORF may increase overall translation efficiency up to a point (56), beyond which strong secondary structures begin to impede ribosomal scanning altogether (65).

Overall, our data from *SERPINA1* wild type, uORF mutant, and structure mutant luciferase assays strongly support that the Δ*G* of unfolding around the Kozak sequence is an important determinant of translation efficiency. Including the additional structure mutants, the leaky-scanning model moderately predicts translation efficiencies ($r^2$ = 0.55), but most of the variation in translation efficiency is explained by the structure leaky-scanning model ($r^2$ = 0.83). Changing the Δ*G* of unfolding around a single Kozak sequence in a given 5′-UTR leads to changes in translation efficiency that are well predicted by the structure leaky-scanning model, but unanticipated by the leaky-scanning model (Fig. 5 *C* and *D*).

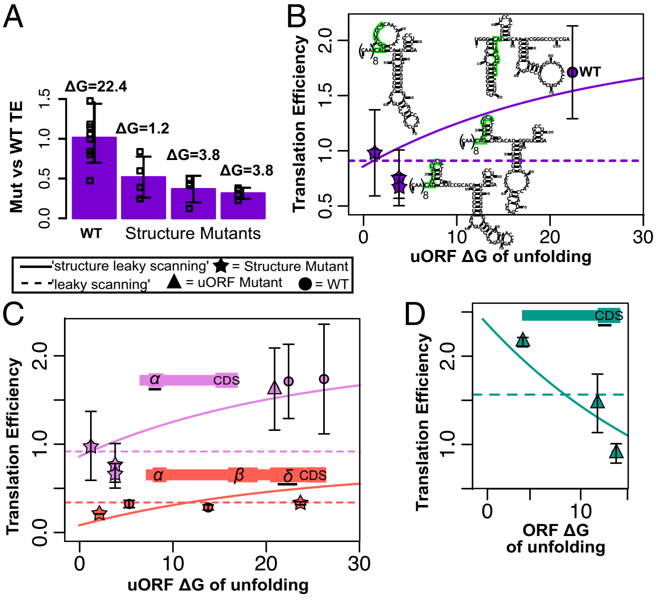


**Fig. 5.** Structure mutants show translation efficiency (TE) is a function of Δ*G* of unfolding around the uORF Kozak sequence. (*A*) TE relative to wild type (WT) for three uORFα structure mutants in transcript NM_001002235.2. Replicate TE values are shown as open squares. The predicted Δ*G* of unfolding is shown for each structure mutant. (*B*) Structure mutant and WT TEs plotted with the structure leaky-scanning (solid line) and leaky-scanning (dotted line) models as functions of uORFα Δ*G* of unfolding. The predicted structure for each mutant and the WT uORFα is shown. Kozak sequences are outlined in green. CAA repeats are abbreviated in the mutants. (*C*) The structure leaky-scanning and leaky-scanning models as functions of uORFα Δ*G* of unfolding (lilac), or uORF δ/δ′ Δ*G* of unfolding (peach). Experimental TEs are plotted for *SERPINA1* structure mutants (stars), uORF mutants (triangles), and WT constructs (circles) that contained only uORFα or uORFα, β, and δ/δ′. (*D*) The structure leaky-scanning and leaky-scanning models as functions of ORF (CDS) Δ*G* of unfolding. Experimental TEs are plotted for *SERPINA1* constructs that contained no uORFs.

**Modeling α-1-Antitrypsin Expression in Tissue.** A goal of transcriptomics is to develop models that accurately describe transcript dynamics and expression in living tissue. As we have seen from tissue-specific transcriptome data, *SERPINA1* transcription is not limited to the liver, and different tissues express different combinations of the *SERPINA1* transcript isoforms (24) (Fig. 1 *B* and *C*). Thus, optimized combinations of *SERPINA1* transcripts could regulate the amount of α-1-antitrypsin protein produced in each tissue. Based on available protein quantification data (68), we calculated the overall *SERPINA1* translation efficiency in each tissue as the ratio of α-1-antitrypsin protein to *SERPINA1* transcript totals. If the translation efficiency of *SERPINA1* mRNA were equal in every tissue, then we expect to observe that α-1-antitrypsin amounts and total *SERPINA1* transcript amounts are correlated. However, we observed no such correlation (Fig. 6*A*), indicating that different tissues have different net α-1-antitrypsin translation rates, potentially due to their unique combinations of *SERPINA1* transcript isoforms. Assuming that the overall translation efficiency in a tissue is the average of the translational efficiencies of all its *SERPINA1* transcripts weighted by abundance, we can use the two scanning models described above to predict *SERPINA1* translation efficiency in tissues (Eq. **4**). While our luciferase assays show uORFγ to repress translation in transcript NM_000295.4 (Fig. 2*C*), it is likely that this uORF is not functional in vivo given its close proximity to the transcript 5′ termini and lack of a canonical translation initiator of short 5′-UTR (TISU) sequence (37). Indeed, when we assessed our ability to model translation efficiency treating the uORFγ as functional (Fig. S10) or nonfunctional (Fig. 6 *B* and *C*), the nonfunctional assumption yielded better prediction of

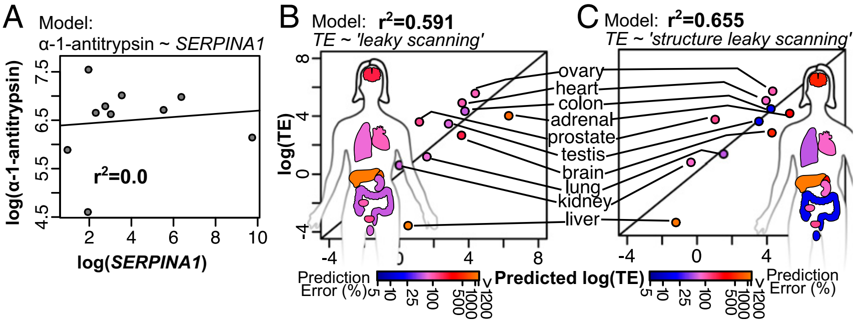


**Fig. 6.** Predictions of *SERPINA1* translation efficiency (TE) in 10 human tissues are improved with the structure leaky-scanning model. (*A*) Total *SERPINA1* transcript versus α-1-antitrypsin protein measurements show no correlation ($r^2$ = 0.0, $n$ = 10). Protein measurements are in normalized spectral counts (68); transcript measurements are in transcripts per million (TPM). (*B*) Leaky-scanning model predictions of TE versus measured TE in each tissue ($r^2$ = 0.591, $n$ = 10). Each tissue is labeled and colored in the plot and in the human figure according to its prediction percent error (Eq. **5**). (*C*) Structure leaky-scanning model predictions of TE versus measured TE in each tissue ($r^2$ = 0.655, $n$ = 10).

tissue-specific translation efficiencies. The leaky-scanning model of translation (Fig. 2*D* and Eq. **1**) explains 59% of the variation in translation efficiency between tissues (Fig. 6*B*), whereas the structure leaky-scanning model (Fig. 4*C* and Eq. **3**) explains 66% (Fig. 6*C*). The addition of RNA structural data to the model of translation thus improves predictions of translation efficiency in human tissues.

## Discussion

The amount of protein produced from a gene is not a simple function of the abundance of the transcript (69, 70). The complex path between transcript expression and protein expression is often a missing link in our understanding of cellular phenotype, indicating a need for integrative models that bridge this divide. *SERPINA1* is exemplary of the effects of posttranscriptional regulation on protein output. While each of the *SERPINA1* transcripts produces the same protein isoform, they do so with different translation efficiencies. Differences in uORF content and 5′ -UTR secondary structure combine to differentiate the translational efficiencies of *SERPINA1* transcripts. Secondary structure plays a surprisingly important role in accounting for these differences, and in determining the repressive effect of individual *SERPINA1* uORFs. When considering the role of secondary structure in a system, correctly defining an RNA secondary structure demands more than a cursory computational prediction. Structural data accurate enough for successful biological models require comprehensive chemical or enzymatic probing of the RNA molecules of interest (46–48). Previously, no correlation was found between secondary structure and translation rate in experiments that measured the protein expression of constructs with varied uORF or CDS Kozak sequences (34, 40). In these studies, it is likely that the purely computational RNA structure models were inadequate for predicting structures around Kozak sequences. In this study, we used SHAPE-MaP chemical probing to successfully improve mRNA translation efficiency predictions (44) (Figs. 2*D*, 4*C*, and 5). While this model aptly describes the translation of *SERPINA1* transcripts, additional experiments measuring the translation efficiencies of simultaneous uORF and structure mutants are necessary to determine the contribution of secondary structure in more detail. Additionally, a more generalizable model of translation efficiency will require modifications to capture additional factors that regulate translation, including overlapping uORFs, reinitiation after uORF translation (43), non-AUG uORF translation (71), and 5′ cap secondary structure (65).

Transcript-specific translation efficiencies may play an important role in tissue-specific protein expression, especially in the case of α-1-antitrypsin, which shows a complex and varied expression pattern across human tissues. However, overall α-1-antitrypsin output in a tissue is not solely a consequence of translation efficiency. Transcripts travel through a coordinated posttranscriptional program, or “regulon” (72), and may diverge from their fellow isoforms at each step. Tissues could also have different overall rates of translation (for example, in a fast- versus slow-growing tissue) or have different rates of protein export.

These additional layers of regulation likely explain why our model of translation efficiency performs better in tissue culture cells than in tissues. However, our model still provides insights into the regulation of α-1-antitrypsin expression in tissues. First, liver tissue is a considerable outlier in both models of *SERPINA1* translation efficiency (Fig. 6 *B* and *C*). Interestingly, predicting much higher translational efficiencies in liver tissue than observed based on measured levels of α-1-antitrypsin is consistent with the understanding that liver exports most of its α-1-antitrypsin into the bloodstream (1, 31, 73). This artifact indicates a need for tissue-specific cellular import/export dynamics to inform models of protein expression. Conversely, the models predict translation efficiency in lung tissue fairly accurately, suggesting that translation of *SERPINA1* mRNA is a major source of its α-1-antitrypsin, which contrasts the paradigm that lung tissue derives its α-1-antitrypsin from the bloodstream (1, 32). Our detection of *SERPINA1* transcripts in cultured lung cells (A549 cells; Fig. 1*C*) and recent quantification of *SERPINA1* transcripts in lung tissue (24) further support the conclusion that cells in the lung itself express α-1-antitrypsin. This surprising conclusion contradicts current models of the role of α-1-antitrypsin in disease. The most common genetic variant in *SERPINA1* associated with COPD and α-1-antitrypsin deficiency, the Pi*Z allele, is thought to cause α-1-antitrypsin to be poorly exported from the liver, leading to deficient α-1-antitrypsin levels in the lungs and eventual neutrophilic overload (1, 8, 32). If lung tissue produces its own α-1-antitrypsin, however, then this disease model is likely incomplete. Instead, disease-associated variants must also impact α-1-antitrypsin levels in lung tissue, either by producing unviable α-1-antitrypsin or reducing its translation. For example, genetic variants could reduce α-1-antitrypsin production if they shift *SERPINA1* transcription to isoforms with the lowest translation efficiencies. A recent study quantified α-1-antitrypsin and different *SERPINA1* transcripts in the serum of α-1-antitrypsin deficiency patients and healthy controls to determine whether patients have different combinations of the transcripts (24). Unfortunately, the primer design in that study did not differentiate between the transcripts with the lowest and highest translational efficiencies, but the data did show a change in transcript proportions for at least one patient population (24).

Ultimately, COPD in α-1-antitrypsin deficiency is caused by the diminished levels of α-1-antitrypsin. Current therapies attempt to deliver donor serum-derived α-1-antitrypsin i.v. to affected individuals, but this treatment is costly and of unknown efficacy (74).

Our work suggests a therapeutic strategy: α-1-antitrypsin levels could be increased in situ, perhaps with antisense oligonucleotides (ASOs) that target the Kozak sequences around the uORFs in *SERPINA1* transcripts, as shown recently for other uORF-containing mRNAs (56). ASOs would likely act as double-stranded regions that increase the $\Delta G$ of unfolding around uORF Kozak sequences, blocking the uORFs in *SERPINA1* transcripts and increasing in situ α-1-antitrypsin expression. Our findings illustrate the importance of the numerous *SERPINA1* transcript isoforms and their translation in disease and the impact of post-transcriptional regulation and secondary structure on phenotype in general.

## Materials and Methods

***SERPINA1* Annotation.** The known *SERPINA1* transcript annotations were taken from RefSeq, version hg38. In each transcript, uORFs are defined by a start and stop codon in the same frame within the 5′-UTR. Distinct uORFs are named here with the Greek letters α, β, γ, δ, and δ′.

**Heat Map of Tissue-Specific Isoform Expression.** Paired-end RNA-seq reads from 16 different tissues were downloaded from the Illumina BodyMap 2.0 project [Gene Expression Omnibus (GEO) accession number GSE30611]. Abundance estimates of the 11 known *SERPINA1* transcripts were quantified with Sailfish, version Beta 0.7.6 (75), using the full human transcriptome (RefSeq, version hg38) as the reference. Estimates of total *SERPINA1* expression in each tissue were calculated as the sum of transcripts per million (TPM) estimates of each transcript. For better visualization in Fig. 1 *B* and *C*, total expression in liver was adjusted by constant to $10^6$, and all other tissues' TPM measurements were adjusted by the same constant. All TPM measurements are provided in Dataset S1.

**Cell Line-Specific Transcript Expression.** A549 and HepG2 cells were provided by the Tissue Culture Facility at University of North Carolina at Chapel Hill. RNA was isolated using TRIzol. Using Phase-Lock Heavy (Eppendorf) to remove the organic phase, the aqueous phase was then purified using a PureLink RNA mini kit (Life Technologies) and subjected to TurboDNase to digest DNA. The total RNA from each cell line was then reverse transcribed with SuperScript III (Life Technologies and New England Biolabs Hot Start Q5; NEB) and amplified with 35 cycles in a reverse transcription–PCR (RT-PCR). Because *SERPINA1* transcript NM_000295.4 has a unique TSS, reverse transcription reactions with reverse primer GCCCCACGAGACAGAAGACGG were split into two different PCRs using forward primers TGGGCAGGAACTGGGCACTG and ACAATGACTCCTTTCGGTAAGTGCAGTGG to amplify NM_000295.4 and all other transcripts, respectively. Following purification with a PureLink PCR cleanup kit (Life Technologies), samples were assessed on an agarose gel. Double-stranded DNA was prepared using a Nextera DNA Library Prep Kit (Illumina). Following concentration determination via Qubit and library analysis with a Bioanalyzer, libraries were sequenced on a miSeq (Illumina). Transcript isoform abundances in A549 and HepG2 cells were estimated with Sailfish, version Beta 0.7.6 (69), mapping the sequenced reads to a reference that includes all known transcripts in RefGene hg38 excepting NM_000295.4, due to its separate primer set. Relative abundance of NM_000295.4 in A549 and HepG2 cells was estimated using a dilution series amplified separately in 35 cycles of RT-PCR with two primer sets: ACTTAGCCCCTGTTTGCTCC (forward) and TGTCGATTCACTGTCCCAGG (reverse) for NM_000295.4 and ACCCTCAGAGTCCTGAGCTG (forward) and CTCTGTCTCTTCTGGCAGGC (reverse) for all other *SERPINA1* transcripts. Both primer sets were designed to amplify ∼150 bp of sequence. Products from the dilution series of NM_000295.4 and other *SERPINA1* transcripts from A549 and HepG2 cells were run on a 2% SEAkem GTG (Lonza) agarose gel and stained with 1× GelStar (Lonza) and were quantified with a gel imager. The quantifications of each dilution series were fit to logistic curves, and inflection points were determined for the NM_000295.4 and other transcript curves. The ratio between the two inflection points was used as the ratio of NM_000295.4 transcript to all other *SERPINA1* transcripts. TPM measurements for each transcript in A549 and HepG2 were adjusted based on their respective NM_000295.4:other ratios. *SERPINA1* transcript abundance estimates in the cell lines are provided in Dataset S1.

**Luciferase Assays.** To assess the translation efficiency of *SERPINA1* transcripts, we built six luciferase constructs containing 5′-UTRs from six selected *SERPINA1* transcripts: NM_001002235.2, NM_000295.4, NM_001127700.1, NM_001127704.1, NM_001127705.1, and NM_00100236.2. The 5′-UTRs were cloned via double digestion with NcoI and SacII into a modified pGL3 that minimizes the amount of plasmid 5′-UTR in the product. For each of the *SERPINA1* and control constructs, 0.5 μg of plasmid was transfected into HeLa cells. Cells were harvested with Cell Culture Lysis Reagent (Promega; E153A) 24 h posttransfection. Luciferase activity of the samples was measured by Luciferase Assay Substrate (Promega; E151C) and Luciferase Assay Buffer (Promega; E152B) with a luminometer (Molecular Devices). Luciferase activity measurements were taken in duplicate and averaged for each sample. The luciferase activity measurement for each sample was normalized to total sample protein concentration, as determined by Bradford assay ($n = 4$), and reported in Dataset S1. Luciferase measurements were further normalized to the abundance of luciferase RNA in each sample to obtain (luciferase activity)/(luciferase RNA), as described previously (76). To quantify luciferase RNA abundance, after measuring luciferase activity, total RNA was extracted with TRIzol. Samples were depleted of DNA with Ambion Turbo DNA-free (AM1907) and reverse transcribed with High Capacity cDNA Reverse Transcription Kit (Applied Biosystems; 4368814). Luciferase and GAPDH cDNAs were quantified by real-time PCR (qRT-PCR) on a Bio-Rad CFX96 Real-Time System. Luciferase and GAPDH primers used were 5′-ACAAAGGCTATCAGGTGGCT-3′ (forward), 5′-CGTGCTCCAAAACAACAACG-3′ (reverse), and 5′-CTGTTGCTGTAGCCAAATTCGT-3′ (forward), 5′-ACCCACTCCTCCACCTTTGAC-3′ (reverse), respectively. Luciferase RNA abundance was determined by the ΔΔCT method ($n = 4$). All (luciferase activity)/(luciferase RNA) measurements are reported relative to an empty vector control to correct for systematic variations between experiments.

**uORF Mutants.** To disrupt uORFs in the original six *SERPINA1* plasmid constructs, we designed primers to substitute the start codon of selected uORFs from AUG to AAG using the NEB Q5 site-directed mutagenesis kit. uORFδ and uORFδ′ were mutated in NM_001002236.2 and NM_001127705.1 luciferase plasmids using primers uORFT435A1F: CCAGGTACAAAGACTCCTTTC/uORFT435AR: CTCAGAAACCACAGCGTC. uORFβ was mutated in NM_001127704.1 luciferase plasmid using primers uORFT285AF: ACTCAGTAAAAGGTAGATCTTGCTAC/uORFT285AR: CACCCCAAAATGCCTGATG. uORFα was mutated in NM_001002235.2 and NM_001127700.1 luciferase plasmids using primers uORFT32AF: GCCCAGGGCAAGCACTGCCTC/uORFT32AR: ACAGTGCCCAGTTCCTGCC. uORFγ was mutated in NM_000295.4 luciferase plasmid using primers uORFT4AF: CCGCGGACAAAGACTCCTTTC/uORFT4AR: CCTCGGCCTCTGCATAAA. Mutant constructs were verified by sequencing (Dataset S2). Luciferase assays were performed on the mutant constructs as above, and results are reported in Fig. 2, Fig. S2, and Dataset S1.

**uORF Structure Mutants.** The structure mutants of uORFα in transcript NM_001002235.2 and of uORFδ in transcript NM_001002236.2 were computationally designed by altering sequences adjacent to the uORF Kozak sequences and predicting the resulting change in secondary structure. To design structure mutants, codons in the given uORF were either permuted or mutated with CodonShuffle (77) to preserve dinucleotide frequency and codon usage. To design mutants with increased uORF $\Delta G$ of unfolding, sequence upstream of the uORF was changed to complement the new uORF sequence. To design mutants with decreased uORF $\Delta G$ of unfolding, sequence upstream of the Kozak sequence was substituted with CAA repeats, which adopt a single-stranded structure (65). Kozak sequences (6 nt upstream and 2 nt downstream of the AUG) were left unchanged. The $\Delta G$ of unfolding was then predicted for all ORFs in the transcript, selecting mutants that exhibited the desired change in uORF $\Delta G$ of unfolding without affecting the predicted $\Delta G$ of other open read frames. Three mutants were selected for NM_001002235.2 and five for NM_001002236.2. The structure mutant 5′-UTRs were cloned via double digestion into modified pGL3 plasmids and verified by sequencing (Dataset S2) as described above. Luciferase assays were performed as before, except without normalization to luciferase RNA levels. Previous luciferase assay data (wild type and uORF mutant constructs) show a very strong linear correlation ($r = 0.95$) between luciferase activities alone and luciferase activities normalized by luciferase RNA, indicating that luciferase activity alone is sufficient to estimate translation efficiency for these constructs. Luciferase assays on wild-type NM_001002236.2 were performed in parallel for comparison. Structure mutant luciferase activities adjusted to the scale of RNA-normalized luciferase values according to the following: adjusted luciferase activity = (luciferase activity)*0.20650 + 0.20141. Structure mutant luciferase assay results are reported in Fig. 5, Fig. S9, and Dataset S1.

**5′-RACE.** To characterize the 5′ ends of luciferase construct transcripts, HeLa cells transfected with NM_001002235.2-luciferase were treated with the RLM-RACE kit (Ambion). Briefly, total RNA was isolated and incubated with calf intestinal phosphatase, and intact 5′-methylguanosine caps were removed by treatment with tobacco acid phosphatase. The 5′ ends of transcripts were ligated to a linker sequence and primed with random hexamers in reverse transcription. cDNA was then amplified via nested PCR with forward primer CTGCATACGACGATTCTGTGATTTG and reverse primer CCCATATCGTTTCATAGCTTCTGC, complementary to the linker sequence and luciferase coding sequence, respectively. PCR products were shotgun cloned into pCR-Blunt vectors (Invitrogen) and Sanger sequenced using forward primer M13F to determine 5′-end sequence.

**Ribosome Profiling Data.** Ribosomal profiling datasets with sufficient coverage over the entire *SERPINA1* locus were identified by RPFdb [Eichhorn et al. (64), U2OS cells] (78). Single-end ribosomal profiling sequencing reads were downloaded from SRA (identifiers SRX680698 and SRX680702), trimmed on the 3′ end to 26 nt as described in ref. 78, and mapped to the human genome build hg38 by Bowtie2, allowing for multimapped reads. Read coverage mapping to the *SERPINA1* transcripts is visualized in Fig. S8.

**SHAPE-MaP Sequencing and Analysis.** 5′-UTRs and coding sequences of six selected *SERPINA1* transcripts were cloned into pBLUNTII using overlap extension PCR and verified by sequencing (Dataset S2). The selected transcripts are the same set analyzed by luciferase assays. Plasmids were named as follows: NM_000295: pAL0108; NM_001002235: pAL0096; NM_001127700: pAL0110 and pAL0111; NM_001002236: pAL0098; NM_001127704: pAL0100; NM_001127705: pAL0103 and pAL0105. Templates for transcription were amplified from ∼100 ng of plasmid using Phusion high-fidelity polymerase (NEB) and primers TAATACGACTCACTATAGGGTGGGCAGGAACTGGGCACT (forward) and TTATTTTTGGGTGGGATTCACCAC (reverse) except for NM_000295:pAL0108, which required TAATACGACTCACTATAGGGACAATGACTCCTTTCGGTAAGTGC as a forward primer. The T7 promoter was added by the forward primers. PCR products were transcribed using a HiScribe T7 High Yield RNA Synthesis Kit (NEB), and the RNA was purified using an Ambion MEGAClear Transcription Clean-up kit (Thermo Fisher) or an RNEasy mini kit (Qiagen). Transcripts were verified using denaturing agarose gel electrophoresis with 2% SEAkem gold agarose and the Amresco Formaldehyde-Free RNA Gel Kit. The 0.5–2 pmol of RNA was used for each reaction in SHAPE-MaP library preparation, as described previously (46) with some modifications. Briefly, RNA was diluted in water, denatured at 95 °C for 1 min, and snap cooled on ice. After the addition of folding buffer (100 mM KCl, ∼10 mM $MgCl_2$, 100 mM Hepes, pH 8.0, final concentration), the RNA was folded at 37 °C for 10–15 min. Then 45 μL of folded RNA was either mixed with 5 μL of DMSO (untreated control) or 5 μL of 100 mM 1-methyl-7-nitroisatoic anhydride (1M7) in DMSO (treated sample). After 5 min, reactions were desalted using G25 or G50 columns. A denatured control was performed in parallel in which the RNA was diluted into 50 mM Hepes, pH 8.0, 4 mM EDTA, and 50% formamide, then heated to 95 °C and treated with 5 μL of 100 mM 1M7 in DMSO. After 1 min, reactions were desalted using G25 or G50 columns. The RNA was reverse transcribed using SuperScript II (Life Technologies) and random nonamers followed by cleanup with a G25 or G50 column. The second strand was synthesized using the NEBNext mRNA Second Strand Synthesis Module (NEB). The double-stranded DNA was then prepared using a Nextera or Nextera XT DNA Library Prep Kit (Illumina). Following DNA library concentration determination via Qubit and analysis by Bioanalyzer, libraries were run on a miSeq (Illumina) and resulting data were analyzed using the ShapeMapper pipeline (46), version 1.2, which calculates the SHAPE reactivity of each nucleotide *i* as follows:

$$R = \frac{mutr_S - mutr_U}{mutr_D},$$

where $mutr_S$ is the mutation rate in the sample treated with the SHAPE reagent, $mutr_U$ is the mutation rate in the untreated control, and $mutr_D$ is the mutation rate in the denatured control.

SHAPE-MaP sequencing data and processed SHAPE reactivity profiles are available in the National Center for Biotechnology Information (NCBI) GEO accession number GSE81525. SHAPE data are also available in SNRNASM format at https://docs.google.com/spreadsheets/d/1_RpB9Jto1-UEmK-ocd9pGMOYrte-t1ALuaA7XTyqI8ZA/edit?usp=drive_web. SHAPE-MaP experiments were performed twice for each transcript, and the average of the two replicate profiles was used for subsequent analyses requiring SHAPE data. For visualization, the median SHAPE profiles in Fig. 3*C* were generated for each transcript by calculating the median SHAPE value in windows of 20 bases (step size = 1) and subtracting the global median.

**Secondary-Structure Analysis.** Each transcript with SHAPE-MaP data was folded with RNAfold, version 2.2.4, incorporating their respective SHAPE data with the –shape option and a max distance of 50 (–maxBPspan = 50) to focus on local structures. SHAPE reactivities were incorporated into structure modeling as pseudo free energies according to ref. 44 using a slope of 1.8 and an intercept of −0.6. The 3′-UTRs were excluded from structure modeling since these regions were not covered by our SHAPE-MaP experiments. Structure models were also generated for the six luciferase constructs, which consist of a specific *SERPINA1* 5′-UTR followed by 700 bases of luciferase coding sequence. For these luciferase construct models, SHAPE data from the endogenous transcripts were used to restrain the 5′-UTRs; SHAPE data for the luciferase coding sequence are unavailable. Δ*G* of unfolding measurements were calculated around Kozak sequences in the structure models by removing base pairs that occur within ±15 bases around the "A" in the start codon. The free energy of the "relaxed" structure was subtracted by the free energy of the original structure to arrive at the Δ*G* of unfolding around the Kozak sequence. (The Δ*G* of unfolding = −Δ*G* of folding.) The Δ*G* of unfolding was calculated around the coding sequence and uORF Kozak sequences in the luciferase constructs when fitting to experimental translation efficiencies, and in the wild-type transcripts when fitting to tissue translation efficiencies (Dataset S1). SHAPE-MaP was not performed on uORF mutant *SERPINA1* transcripts, but, because point mutations rarely cause perceptible changes in secondary structure (79, 80), it is assumed that the structures of the wild-type transcripts closely approximate the structures of the uORF mutants. Δ*G* of unfolding values around *SERPINA1* structure mutant uORFs were calculated in the same manner as above, with the exception that the underlying structure models were generated using naive prediction due to the absence of SHAPE-MaP data for structure mutants.

**Models.** The performance of a number of different translation efficiency models is described in Dataset S1. Predictor(s) were fit with simple linear regression to the (luciferase activity)/(luciferase RNA) measurements of the six *SERPINA1* constructs and six uORF mutant constructs. Adjusted $r^2$ values and predictor *P* values were determined by the lm function in R, version 3.2.3. The models we feature in the results are the leaky-scanning and the structure leaky-scanning models (Eqs. **1** and **3**). Eq. **1**, the leaky-scanning model, is our expansion of a previously published model (34) to allow multiple nonoverlapping uORFs. Eq. **2** is a rederivation of Eq. **1** that allows for translation reinitiation after uORF translation, dependent on a logistic model that assumes the ribosome has a 50% probability of reinitiating 35 bases downstream from the end of a uORF (42). Eq. **3**, the structure leaky-scanning model, is our variation of Eq. **1** that incorporates the Δ*G* of unfolding around Kozak sequences, as assumed to be exponentially related to ribosomal initiation (57). TE is "translation efficiency"; *k*, *k*′, and *i* are constants; and $P_n$ is the strength of the given Kozak sequence as determined previously (40). Kozak strengths are converted to probabilities by dividing by the maximum Kozak strength, 150. Because uORF order matters in Eq. **2**, $P_n{}'$ refers to the Kozak sequence strength of the uORF that is *n*th closest to the CDS. $d_n$ refers to the distance between the end of the *n*th uORF and the beginning of the next ORF. $\Delta G_n$ in Eq. **3** corresponds to the Δ*G* of unfolding ±15 bases around the given Kozak sequence (calculation described above). The subscripts of $P_n$ and $\Delta G_n$ indicate either the coding sequence or the *n*th uORF, numbered 5′–3′ in each transcript. $P_n$ and $\Delta G_n$ values for a transcript without an *n*th uORF are simply zero. $P_n$ and $\Delta G_n$ values are provided for every transcript's CDS and uORF(s) in Dataset S1. Constants *k*′ and *i* were optimized in the structure leaky-scanning model fit to *SERPINA1* wild type and uORF mutant (luciferase activity)/(luciferase RNA) values, and the constant *k* is the original published value (34):

$k = 0.86$
$k' = 0.39$
$i = 0.037$

$$TE \sim kP_{cds}(1 - kP_1)(1 - kP_2)(1 - kP_3), \qquad [1]$$

$$TE \sim kP_{cds}\left[1 - kP_1'\left(1 - \frac{1}{1 + e^{-d_1+35}}\right)\times\left(1 - kP_2'\left(1 - \frac{1}{1 + e^{-d_2+35}}\right)\left(1 - kP_3'\left(1 - \frac{1}{1 + e^{-d_3+35}}\right)\right)\right)\right], \qquad [2]$$

$$TE \sim k'P_{cds}e^{-i\Delta G_{cds}}\left(1 - k'P_1e^{-i\Delta G_1}\right)\left(1 - k'P_2e^{-i\Delta G_2}\right)\left(1 - k'P_3e^{-i\Delta G_3}\right). \qquad [3]$$

When including data from structure mutants, constants were optimized in the structure leaky-scanning model fit to the adjusted luciferase activities measured for wild type, uORF mutant, and structure mutant constructs, where the adjusted luciferase activity = (luciferase activity)*0.20650 + 0.20141. Constants for these models are as follows: *k* = 0.86, *k*′ = 1.0, *i* = 0.044. Structure mutants affected by strong hairpin inhibition were excluded from model fitting.

**Models in Tissues.** Total tissue *SERPINA1* concentrations (in transcripts per million) were fit to their α-1-antitrypsin protein concentrations (in parts per million) (Fig. 5*A*) with simple linear regression with the lm function in R, version 2.3.2. The *SERPINA1* transcript concentrations are described above (Fig. 1*B* and Dataset S1), and the α-1-antitrypsin protein measurements are derived from mass spectrometry data on the human proteome (68) (Dataset S1). *SERPINA1* translation efficiency in each tissue was measured by dividing α-1-antitrypsin protein concentration by the total *SERPINA1* concentration. To predict tissue translation efficiencies with a given model, the model's translation efficiency estimates of all 11 *SERPINA1* transcripts were used to predict the average translation efficiency of each tissue as in Eq. **4**. $TE_j$ is the model-predicted translation efficiency of tissue *j*, $TPM_{i,j}$ is the transcript abundance in transcripts per million of *SERPINA1* transcript *i* in tissue *j*, and $m(i)$ is the function for the translation efficiency of transcript *i* with parameters from fitting the model to the luciferase data:

$$TE_j = \frac{\sum_{i=1}^{11} m(i) \times TPM_{i,j}}{\sum_{i=1}^{11} TPM_{i,j}}. \qquad [4]$$

The model-predicted values for tissue translation efficiency were then fit to the measured tissue translational efficiencies with simple linear regression. Models fit best to the log of the measured tissue translational efficiencies. *R*-squared values and model *P* values are reported in Dataset S1. The structure leaky-scanning model requires SHAPE-based secondary structure information (Δ*G* of unfolding values), which is not available for transcripts NM_001127701.1, NM_001127702.1, NM_001127703.1, NM_001127705.1, NM_001127706.1, and NM_001127707.1. In their case, free energies of unfolding were assigned based on the most similar transcript with available secondary-structure data. The Δ*G* of unfolding measurements used in tissue predictions were derived from the *SERPINA1* transcript secondary-structure models (described above). Percent error of the leaky-scanning and structure leaky-scanning models (Fig. 5 *B* and *C*) in each tissue was calculated according to Eq. **5**, where TE is translation efficiency:

$$\mathrm{Error}(\%) = \frac{|\mathrm{modelPredictedTE} - \mathrm{measuredTE}|}{\mathrm{measuredTE}}. \qquad [5]$$

**ACKNOWLEDGMENTS.** A.M.M. is an Arnold O. Beckman Postdoctoral Fellow. L.L. is an American Cancer Society Postdoctoral Fellow. This work was supported by National Institutes of Health–National Heart, Lung, and Blood Institute Grant R01 HL111527 (to A.L.), National Institutes of Health–National Institute of Allergy and Infectious Diseases Grant R01 AI03311 (to N.J.M.), and National Institutes of Health Grant R35 GM122532 (to K.M.W.), and by the North Carolina University Cancer Research Fund (to N.J.M.).

1. Crystal RG (1989) The alpha 1-antitrypsin gene and its deficiency states. *Trends Genet* 5:411–417.
2. Castaldi PJ, et al. (2010) The COPD genetic association compendium: A comprehensive online database of COPD genetic associations. *Hum Mol Genet* 19:526–534.
3. Eden E, et al. (1997) Atopy, asthma, and emphysema in patients with severe alpha-1-antitrypysin deficiency. *Am J Respir Crit Care Med* 156:68–74.
4. Mahadeva R, Gaillard M, Pillay V, Halkas A, Lomas D (2001) Characterization of a new variant of alpha(1)-antitrypsin E(Johannesburg) (H15N) in association with asthma. *Hum Mutat* 17:156.
5. Chappell S, et al. (2006) Cryptic haplotypes of SERPINA1 confer susceptibility to chronic obstructive pulmonary disease. *Hum Mutat* 27:103–109.
6. Pillai SG, et al.; ICGN Investigators (2009) A genome-wide association study in chronic obstructive pulmonary disease (COPD): Identification of two major susceptibility loci. *PLoS Genet* 5:e1000421.
7. Løkke A, Lange P, Scharling H, Fabricius P, Vestbo J (2006) Developing COPD: A 25 year follow up study of the general population. *Thorax* 61:935–939.
8. Primhak RA, Tanner MS (2001) Alpha-1 antitrypsin deficiency. *Arch Dis Child* 85:2–5.
9. Fregonese L, Stolk J (2008) Hereditary alpha-1-antitrypsin deficiency and its clinical consequences. *Orphanet J Rare Dis* 3:16.
10. Ferrarotti I, et al. (2014) Identification and characterisation of eight novel SERPINA1 null mutations. *Orphanet J Rare Dis* 9:172.
11. Edmonds BK, Hodge JA, Rietschel RL (1991) Alpha 1-antitrypsin deficiency-associated panniculitis: Case report and review of the literature. *Pediatr Dermatol* 8:296–299.
12. Lewis M, et al. (1985) Severe deficiency of alpha 1-antitrypsin associated with cutaneous vasculitis, rapidly progressive glomerulonephritis, and colitis. *Am J Med* 79: 489–494.
13. Sandhaus RA, Stoller JK (2013) Introduction to the 50th anniversary of the description of alpha-1 antitrypsin deficiency. *COPD* 10:1–2.
14. Morgan K, Scobie G, Kalsheker NA (1993) Point mutation in a 3′ flanking sequence of the alpha-1-antitrypsin gene associated with chronic respiratory disease occurs in a regulatory sequence. *Hum Mol Genet* 2:253–257.
15. Laubach VE, Ryan WJ, Brantly M (1993) Characterization of a human alpha 1-antitrypsin null allele involving aberrant mRNA splicing. *Hum Mol Genet* 2:1001–1005.
16. Seixas S, Mendonça C, Costa F, Rocha J (2002) Alpha1-antitrypsin null alleles: Evidence for the recurrence of the L353fsX376 mutation and a novel G→A transition in position +1 of intron IC affecting normal mRNA splicing. *Clin Genet* 62:175–180.
17. Zorzetto M, et al.; SAPALDIA Team (2008) SERPINA1 gene variants in individuals from the general population with reduced alpha1-antitrypsin concentrations. *Clin Chem* 54:1331–1338.
18. Lackey L, McArthur E, Laederach A (2015) Increased transcript complexity in genes associated with chronic obstructive pulmonary disease. *PLoS One* 10:e0140885.
19. Braunschweig U, et al. (2014) Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res* 24:1774–1786.
20. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40:1413–1415.
21. Wheeler DL, et al. (2008) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 36:D13–D21.
22. Araujo PR, et al. (2012) Before it gets started: Regulating translation at the 5′ UTR. *Comp Funct Genomics* 2012:475731.
23. Sgourou A, et al. (2004) Thalassaemia mutations within the 5′UTR of the human beta-globin gene disrupt transcription. *Br J Haematol* 124:828–835.
24. Matamala N, et al. (2015) Alternative transcripts of the SERPINA1 gene in alpha-1 antitrypsin deficiency. *J Transl Med* 13:211.
25. Perlino E, Cortese R, Ciliberto G (1987) The human alpha 1-antitrypsin gene is transcribed from two different promoters in macrophages and hepatocytes. *EMBO J* 6: 2767–2771.
26. Grillo G, et al. (2010) UTRdb and UTRsite (RELEASE 2010): A collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res* 38:D75–D80.
27. Gu W, Zhou T, Wilke CO (2010) A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput Biol* 6: e1000664.
28. Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324:255–258.
29. Pickering BM, Willis AE (2005) The implications of structured 5′ untranslated regions on translation and disease. *Semin Cell Dev Biol* 16:39–47.
30. Stenson PD, et al. (2003) Human gene mutation database (HGMD): 2003 update. *Hum Mutat* 21:577–581.
31. Lomas DA, Evans DL, Finch JT, Carrell RW (1992) The mechanism of Z alpha 1-antitrypsin accumulation in the liver. *Nature* 357:605–607.
32. Brantly M, Nukiwa T, Crystal RG (1988) Molecular basis of alpha-1-antitrypsin deficiency. *Am J Med* 84:13–31.
33. Calvo SE, Pagliarini DJ, Mootha VK (2009) Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci USA* 106:7507–7512.
34. Ferreira JP, Overton KW, Wang CL (2013) Tuning gene expression with synthetic upstream open reading frames. *Proc Natl Acad Sci USA* 110:11284–11289.
35. Touriol C, et al. (2003) Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biol Cell* 95:169–178.
36. Peabody DS (1989) Translation initiation at non-AUG triplets in mammalian cells. *J Biol Chem* 264:5031–5035.
37. Elfakess R, Dikstein R (2008) A translation initiation element specific to mRNAs with very short 5′UTR that also regulates transcription. *PLoS One* 3:e3094.
38. Wasylyk B, Wasylyk C, Matthes H, Wintzerith M, Chambon P (1983) Transcription from the SV40 early-early and late-early overlapping promoters in the absence of DNA replication. *EMBO J* 2:1605–1611.
39. Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44:283–292.
40. Noderer WL, et al. (2014) Quantitative analysis of mammalian translation initiation sites by FACS-seq. *Mol Syst Biol* 10:748.
41. Kozak M (2002) Pushing the limits of the scanning mechanism for initiation of translation. *Gene* 299:1–34.
42. Kozak M (1987) Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. *Mol Cell Biol* 7:3438–3445.
43. Gunišová S, Beznosková P, Mohammad MP, Vlčková V, Valášek LS (2016) In-depth analysis of *cis*-determinants that either promote or inhibit reinitiation on GCN4 mRNA after translation of its four short uORFs. *RNA* 22:542–558.
44. Deigan KE, Li TW, Mathews DH, Weeks KM (2009) Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci USA* 106:97–102.
45. Seetin MG, Mathews DH (2012) RNA structure prediction: An overview of methods. *Methods Mol Biol* 905:99–122.
46. Siegfried NA, Busan S, Rice GM, Nelson JA, Weeks KM (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* 11:959–965.
47. Wan Y, et al. (2012) Genome-wide measurement of RNA folding energies. *Mol Cell* 48:169–181.
48. Ding Y, et al. (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505:696–700.
49. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505: 701–705.
50. Lorenz R, et al. (2011) ViennaRNA package 2.0. *Algorithms Mol Biol* 6:26.
51. Hajdin CE, et al. (2013) Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc Natl Acad Sci USA* 110:5498–5503.
52. Low JT, Weeks KM (2010) SHAPE-directed RNA secondary structure prediction. *Methods* 52:150–158.
53. Ramachandran S, Ding F, Weeks KM, Dokholyan NV (2013) Statistical analysis of SHAPE-directed RNA secondary structure modeling. *Biochemistry* 52:596–599.
54. Lotfi M, Zare-Mirakabad F, Montaseri S (2015) RNA secondary structure prediction based on SHAPE data in helix regions. *J Theor Biol* 380:178–182.

55. Goodman DB, Church GM, Kosuri S (2013) Causes and effects of N-terminal codon bias in bacterial genes. *Science* 342:475–479.
56. Liang XH, et al. (2016) Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames. *Nat Biotechnol* 34:875–880.
57. Salis HM (2011) The ribosome binding site calculator. *Methods Enzymol* 498:19–42.
58. Wolin SL, Walter P (1988) Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J* 7:3559–3569.
59. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324:218–223.
60. Lange SJ, et al. (2012) Global or local? Predicting secondary structure and accessibility in mRNAs. *Nucleic Acids Res* 40:5215–5226.
61. Flamm C, Fontana W, Hofacker IL, Schuster P (2000) RNA folding at elementary step resolution. *RNA* 6:325–338.
62. Lu Z, et al. (2016) RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* 165:1267–1279.
63. Krokhotin A, Mustoe AM, Weeks KM, Dokholyan NV (2017) Direct identification of base-paired RNA nucleotides by correlated chemical probing. *RNA* 23:6–13.
64. Eichhorn SW, et al. (2014) mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Mol Cell* 56:104–115.
65. Babendure JR, Babendure JL, Ding JH, Tsien RY (2006) Control of mammalian translation by mRNA structure near caps. *RNA* 12:851–861.
66. Kozak M (1989) Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol Cell Biol* 9:5134–5142.
67. Koromilas AE, Lazaris-Karatzas A, Sonenberg N (1992) mRNAs containing extensive secondary structure in their 5′ non-coding region translate efficiently in cells overexpressing initiation factor eIF-4E. *EMBO J* 11:4153–4158.
68. Kim MS, et al. (2014) A draft map of the human proteome. *Nature* 509:575–581.
69. Maier T, Güell M, Serrano L (2009) Correlation of mRNA and protein in complex biological samples. *FEBS Lett* 583:3966–3973.
70. Lundberg E, et al. (2010) Defining the transcriptome and proteome in three functionally different human cell lines. *Mol Syst Biol* 6:450.
71. Young SK, Baird TD, Wek RC (2016) Translation regulation of the glutamyl-prolyl-tRNA synthetase gene EPRS through bypass of upstream open reading frames with noncanonical initiation codons. *J Biol Chem* 291:10824–10835.
72. Keene JD (2007) RNA regulons: Coordination of post-transcriptional events. *Nat Rev Genet* 8:533–543.
73. Carrell RW, Aulak KS, Owen MC (1989) The molecular pathology of the serpins. *Mol Biol Med* 6:35–42.
74. Gøtzsche PC, Johansen HK (2016) Intravenous alpha-1 antitrypsin augmentation therapy for treating patients with alpha-1 antitrypsin deficiency and lung disease. *Cochrane Database Syst Rev* 9:CD007851.
75. Patro R, Mount SM, Kingsford C (2014) Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol* 32:462–464.
76. Kutchko KM, et al. (2015) Multiple conformations are a conserved and regulatory feature of the RB1 5′ UTR. *RNA* 21:1274–1285.
77. Jorge DM, Mills RE, Lauring AS (2015) CodonShuffle: A tool for generating and analyzing synonymously mutated sequences. *Virus Evol* 1:vev012.
78. Xie SQ, et al. (2016) RPFdb: A database for genome wide information of translated mRNA generated from ribosome profiling. *Nucleic Acids Res* 44:D254–D258.
79. Corley M, Solem A, Qu K, Chang HY, Laederach A (2015) Detecting riboSNitches with RNA folding algorithms: A genome-wide benchmark. *Nucleic Acids Res* 43:1859–1868.
80. Solem AC, Halvorsen M, Ramos SB, Laederach A (2015) The potential of the riboSNitch in personalized medicine. *Wiley Interdiscip Rev RNA* 6:517–532.