Absolute quantification of translational regulation and burden using combined sequencing approaches

Thomas E. Gorochowski[1,2,*], Irina Chelysheva[3], Mette Eriksen[4], Priyanka Nair[3], Steen Pedersen[4], Zoya Ignatova[3,*]

[1] BrisSynBio, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol, UK

[2] School of Biological Sciences, University of Bristol, Tyndall Avenue, Bristol, UK

[3] Biochemistry and Molecular Biology, Department of Chemistry, University of Hamburg, Hamburg, Germany.

[4] Biomolecular Sciences, Department of Biology, University of Copenhagen, Copenhagen, Denmark

* Correspondence should be addressed to T.E.G (thomas.gorochowski@bristol.ac.uk) and Z.I. (zoya.ignatova@uni-hamburg.de)

Keywords: translation; transcription; Ribo-seq; RNA-seq; genetic circuits; biometrology; systems biology; synthetic biology

Abstract

Translation of mRNAs into protein is a key cellular process. Ribosome binding sites and stop codons provide signals to initiate and terminate translation, while stable secondary mRNA structures can induce translational recoding events. Fluorescent proteins, commonly used to characterize such elements, require the modification of a part's natural context and allow only a few parameters to be monitored concurrently. Here, we develop a methodology that combines ribosome profiling (Ribo-seq) with quantitative RNA sequencing (RNA-seq) to enable the high-throughput characterization of genetic parts controlling translation in absolute units. We simultaneously measure 743 translation initiation rates and 754 termination efficiencies across the *Escherichia coli* transcriptome, in addition to translational frameshifting induced at a stable RNA pseudoknot structure. By analyzing the transcriptional and translational response, we discover that sequestered ribosomes at the pseudoknot causes a σ^{32}-mediated stress response, codon-specific pausing, and drop in translation initiation rates across the cell. Our work demonstrates the power of integrating global approaches to give a comprehensive and quantitative understanding of gene regulation and burden in living cells.

Introduction

Gene expression is a multi-step process involving the transcription of DNA into messenger RNA (mRNA) and the translation of mRNAs into protein. To fully understand how a cell functions and adapts to changing environments and adverse conditions (e.g., disease or chronic stress), quantitative methods to precisely observe these processes are required [1]. Gene regulatory networks (also known as "genetic circuits") control where and when these processes take place and underpin many important cellular phenotypes. Recently, there has been growing interest in building synthetic genetic circuits to understand the function of natural gene regulatory networks through precise perturbations and/or creating systems *de novo* [2,3].

In synthetic biology, genetic circuits are designed to control gene expression in a desired way [4]. Circuits have been built to implement a range of digital [5,6] and analog functions [7], and have been integrated with endogenous pathways to control cellular behaviors [8,9]. The construction of a genetic circuit requires the assembly of many DNA-encoded parts that control the initiation and termination of transcription and translation. A major challenge is predicting how a part will behave when assembled with many others [10]. The sequences of surrounding parts [11], interactions with other circuit components or the host cell [10,12–14], and the general physiological state of the cell [15,16] can all alter a part's behavior. Although biophysical models have been refined to capture some contextual effects [17–19], and new types of part created to insulate against these factors [6,7,20–23], we have yet to reach a point where large and robust genetic circuits can be reliably built on our first attempt.

Fluorescent proteins and probes are commonly used to characterize the function of genetic parts [24,25] and debug the failure of genetic circuits [26]. When used for characterization, the part of interest is generally placed into a new genetic backbone (often a plasmid) where its behavior is directly linked to the expression of one or more fluorescent proteins [27]. When debugging a circuit failure, it is not possible to extract the part of interest as its behavior in the context of the circuit is required. For circuits where transcription rate (i.e. RNAP flux) is used as a common signal between components [28], debugging plasmids containing a promoter responsive to the signal of interest have been used to track the propagation of signals and reveal the root causes of failures [26]. Alternatively, any genes whose expression is controlled by the part of interest can be tagged by fusion to a fluorescent protein [29]. Such modifications allow for a direct readout of protein level but come at the cost of alterations to the circuit. This is problematic as there is no guarantee the fluorescent tag itself will not affect a part's function [30,31].

The past decade has seen tremendous advances in sequencing technologies. This has resulted in continuously falling costs and a growing range of information that can be captured [32]. Sequencing methods exist to measure chromosomal architecture [33], RNA secondary structure [34],

DNA and RNA abundance [35], and translation efficiency [36]. New developments have expanded the capabilities even further towards more quantitve measurements of transcription and protein synthesis rates with native elongating transcript sequencing (NET-seq) [37] and ribosome profiling (Ribo-seq) [38,39]. Ribo-seq provides position-specific information on the translating ribosomes through sequencing of ribosome-protected fragments (RPFs; approximately 25–28 nt) which allows genome-wide protein synthesis rates to be inferred with accuracy similar to quantitative proteomics [38].

Sequencing technologies offer several advantages over fluorescent probes for characterization and debugging genetic parts and circuits. First, they do not require any modification of the circuit DNA. Second, they provide a more direct measurement of the processes being controlled (e.g. the RNAs synthesized during transcription), and third, they capture information regarding the host response and the indirect effects that this may have on a part's function. Furthermore, for large multi-component circuits or synthetic genomes, sequencing is the only way of gaining a comprehensive view of the system's behavior, offering a scalable approach which goes beyond the limited numbers of fluorescent probes that can be measured simultaneously. Recently, RNA-seq has been used to characterize every transcriptional component in a large logic circuit composed of 46 genetic parts [40]. While successful in demonstrating the ability to characterize genetic part function, observe internal transcriptional states, and find the root cause of circuit failures, the use of RNA-seq alone restricts the method to purely transcriptional elements and does not allow for quantification of this process in physically meaningful units.

Here, we address these limitations by combining Ribo-seq with a modified version of RNA-seq to quantitatively characterize genetic parts controlling transcription and translation at a nucleotide resolution. By supplementing this sequencing data with other experimentally measured cell parameters, we are able to generate transcription and translation profiles that capture the local flux of RNA polymerases and ribosomes governing these processes in absolute units. We apply our method to *Escherichia coli* and demonstrate how local changes in these profiles can be interpreted using biophysical models to measure the absolute performance of many types of genetic part. Finally, we demonstrate how genome-wide shifts in transcription and translation can be used to dissect the burden synthetic genetic constructs place on their host cell and the role that allocation of shared cellular resources such as ribosomes play.

Results

Generating transcription and translation profiles in absolute units

To enable quantification of both transcription and translation in absolute units, we modified the RNA-seq protocol and extended the Ribo-seq protocol with quantitative cell measures (red elements in

Figure 1A). For RNA-seq, we introduced RNA spike-ins to our samples at known molar concentrations before fragmentation (left panel, **Figure 1A**). These RNAs span a wide range of lengths (250–2000 nt) and share no homology with the transcriptome of the cells being studied. Using the RNA spike-ins as a standard, numbers of mapped reads are converted to absolute molecule counts and then further normalized by cell counts to give absolute transcript copy numbers per cell [41,42] (**Materials and Methods**). These conversion factors are also used with a previously described method [40] to generate transcription profiles that capture RNA polymerase flux (in RNAP/s units) across the genome.

For Ribo-seq, we use direct ligation of the adaptors to the RPFs [43] to allow for the capture of low-abundance transcripts [44] and complemented this protocol with additional measurements of cellular properties such as growth rate, cell count, and protein mass (right panel, **Figure 1A**). To generate the translation profile, we start by first calculating an RPF profile $C(x)$ that captures the number of RPFs mapping to nucleotide x. Each RPF read in the Ribo-seq data set is mapped to reference sequences of all transcripts. For each mapped read, we consider the physical constraints of how a transcript is shielded by a ribosome and estimate the nucleotide corresponding to the central position of the codon in the P site (**Materials and Methods**). The value of $C(x)$ at this point is incremented by one and the remaining reads processed in the same way. Because each ribosome translates at a relatively constant speed [38,45], the number of RPFs mapping to a nucleotide is proportional to the overall translation rate (i.e. ribosome flux). Given that each RPF corresponds to a single ribosome, and assuming that each translating ribosome produces a single protein, then the number of RPFs mapping within a gene f_i as a fraction of the total f_t equals the proportion of the proteome corresponding to that gene. By measuring the total mass of a cell's proteome m_t and calculating the mass of each encoded protein m_i, we can then calculate the protein copy number per cell of gene i as $n_i = \frac{f_i m_t}{f_t m_i}$. This can be converted to a total number of amino acids $a_t = \sum_i n_i a_i$, where a_i is the number of amino acids in gene i. Because each amino acid is attached to the peptide chain by a ribosome translocating from the A to P site, a single RPF read captures a_t/f_t translocating ribosomes. Furthermore, as the cells are in exponential phase growth, the proteome will be duplicated after each doubling time t_d (in seconds), and therefore the ribosome flux (in ribosomes/s units) passing through a single footprint is given by $r_f = \frac{a_t}{f_t t_d}$. We finally multiply $C(x)$ by r_f and normalize by the transcript copy number $m(x)$ from the RNA-seq data to generate the translation profile,

$$R(x) = \frac{3\, r_f C(x)}{m(x)}. \quad (1)$$

An additional factor of three is required because translation profiles are defined at a nucleotide resolution and ribosomes translocate three nucleotides (a codon) at a time when attaching an amino acid.

Importantly, because both the transcription and translation profiles are given in absolute units (RNAP/s and ribosomes/s, respectively), they can be directly compared across samples without any further normalization.

Characterizing genetic parts controlling translation

Genetic parts controlling translation alter the ribosome flux along a transcript. These changes are captured by the translation profiles. We developed biophysical models to interpret these signals and quantify the performance of RBSs, stop codons and translational recoding (or ribosome frameshifting) at stable secondary structures.

In prokaryotes, RBSs support translation initiation and cause a jump in the translation profile after the start codon of the associated gene due to an increased ribosome flux originating at that location (**Figure 1B**). If initiation is rate-limiting [38], then the translation initiation rate of an RBS (in ribosomes/s units) is given by the absolute increase in ribosome flux after the RBS,

$$\delta R = \sum_{i=x_s}^{x_e} \frac{R(i)}{(x_e - x_s)} - \sum_{i=x_0-n}^{x_0} \frac{R(i)}{n}. \quad (2)$$

In this equation, x_0 is the start point of the RBS, and x_s and x_e are the start and end point of the coding region associated with this RBS, respectively (**Figure 1B**). A window of n = 30 nt (10 codons) is used to average fluctuations in the translation profile upstream of the RBS, which is equal to the approximate length of a ribosome's footprint on a transcript. If the transcription start site (TSS) of the promoter expressing this transcript falls in the upstream window then the start point ($x_0 - n$) is adjusted to the TSS to ensure incoming ribosome flux is not underestimated. A similar change is made if the coding sequence is in an operon structure and the end of an upstream gene falls in this window. In this case, the start point is adjusted to 9 nt (3 codons) downstream of the overlapping gene's stop codon.

In eukaryotes, genes are generally monocistronic with translation initiation occurring through scanning of the 5' untranslated region (5'-UTR) by the 43S preinitiation complex until a suitable start codon is found. This allows a translation-competent 80S ribosome to assemble and translation elongation to begin [46]. In this case, no ribosome flux is generated by upstream genes. Therefore, when calculating the initiation rate of a 5'-UTR, the second term in Equation 2 is set to zero (i.e. $\sum_{i=x_0-n}^{x_0} \frac{R(i)}{n} = 0$).

Ribosomes terminate translation and disassociate from a transcript when a stop codon (TAA, TAG or TGA) is encountered. This leads to a drop in the translation profile at these points (**Figure 1C**). Although this process is typically efficient, ribosomes can sometimes read through a stop codon and continue translating downstream [47]. The termination efficiency of a stop codon (i.e. the fraction of ribosomes terminating) is given by

$$T_e = 1 - \frac{\sum_{i=x_1}^{x_1+n} R(i)/n}{\sum_{i=x_s}^{x_0} R(i)/(x_o - x_s)}, \quad (3)$$

where x_0 and x_1 are the start and end nucleotide of the stop codon, respectively, x_s is the start of the coding region associated to this stop codon, and n = 30 nt (10 codons) is the window used to average fluctuations in the translation profile downstream of the stop codon (**Figure 1C**). If additional stop codons are present in the downstream window, the end point of this window is adjusted to ensure that the termination efficiency of only the (first) stop codon of interest is measured. A similar adjustment is made if the end of a transcript generated by an upstream promoter ends within this window.

Translation requires the conversion of triplets of nucleotides (a codon) into one of the proteinogenic amino acids. Because every nucleotide can be either in the first, second or third position of a codon, three reading frames can be defined for every transcript. A single sequence can therefore encode three different proteins. Although synthetic biology approaches rarely use multiple reading frames, natural systems exploit this feature in many different ways [48–51]. In our characterization workflow, the RPFs used to generate the translation profiles are aligned to the middle nucleotide of the codon in the ribosomal P site, providing the frame of translation. Although the flexible nature of RNA hampers accuracy, if there are sufficient numbers of RPFs, the major frame along a transcript can be determined [52,53].

To characterize genetic parts that cause translational recoding through ribosomal frameshifting, two approaches can be used (**Figure 1D**). First, the major frame of translation can be estimated by separately summing all RPFs for each reading frame in the region before and after the expected point of frameshifting and summarizing these as a fraction of all RPFs in these regions (**Figure 1D**). Strong frameshifting will cause the fraction of RPFs to shift from the original frame to a new one when comparing the regions. Second, the frameshifting efficiency can be calculated from the translation profile using,

$$F_e = 1 - \frac{\sum_{i=x_1}^{x_e} R(i)/(x_e - x_1)}{\sum_{i=x_s}^{x_0} R(i)/(x_o - x_s)}, \quad (4)$$

where x_0 is the nucleotide at the start of the region where frameshifting occurs, and x_1 is the end nucleotide of the stop codon for the first coding sequence (**Figure 1D**).

Measuring genome-wide translation initiation and termination in Escherichia coli

We applied our method to *Escherichia coli* cells harboring a *lacZ* gene whose expression is induced using isopropyl β-D-1-thiogalactopyranoside (IPTG) (**Figure 2A**). After induction for 10 min, *lacZ* expression reached 14% of the total cellular protein mass (**Supplementary Table S1**). Samples from non-induced and induced cells were subjected to our combined sequencing workflow. Sequencing yielded between 41–199 million reads per sample (**Supplementary Table S2**) with a high correlation between biological replicates (R^2 > 0.96; **Supplementary Figure S1**).

Transcription and translation profiles were generated from this data and used to measure translation initiation rates of RBSs and termination efficiencies of stop codons across the genome. To remove the bias due to the RPF enrichment in the 5'-end of coding regions [39] (**Figure 2B**), x_s was adjusted to 51 bp (17 codons) downstream of the start codon when estimating average ribosome flux across a coding region in Equations 2 and 3. To verify that this correction gave translation rates that were constant across each gene (a necessary condition for our models), we compared the number of RPFs mapping to the first and second half of each coding region. If ribosomes traverse the whole coding sequence at a constant speed, then the two halves of a transcript should have a near identical numbers. As required, we found a high correlation between both halves for non-induced and induced cells (**Supplementary Figure S2**).

We characterized chromosomal RBSs in *E. coli* by assuming that each covered a region spanning from the start codon to 15 bp upstream. The measured translation initiation rates varied over two orders of magnitude with a median initiation rate of 0.1 ribosomes/s (**Figure 2C**; **Supplementary Data S1**). This closely matches previously measured rates for single genes [54]. A few RBSs mostly related to stress response functions (*tabA*, *hdeA*, *uspA*, *uspG*), the ribosomal subunit protein L31 (*rpmE*), and some unknown genes (*ydiH*, *yjdM*, *yjfN*, *ybeD*), reached much higher rates of up to 2.45 ribosomes/s.

For estimating termination efficiency at the stop codons, we considered 9 nt (3 codons) up and downstream of the stop codon to ensure local fluctuations in the translation profile due to termination did not affect our measurements (**Figure 2B**). Our analysis showed a median termination efficiency across the genome of 0.987, with 339 stop codons (45% of all those measured) having termination efficiencies >0.99 (**Figure 2D**; **Supplementary Data S2**).

Similar performance for both RBSs (R^2 = 0.84) and terminators (R^2 = 0.52) was found between non-induced and induced conditions (**Figures 2E** and **2F**; **Supplementary Data S1** and **S2**).

Quantifying transcription and translation of a synthetic genetic construct

The quantitative measurements produced by our methodology allow both transcription and translation to be monitored simultaneously. To demonstrate this capability, we focused on the LacZ expression construct and quantified the inducible promoter and terminator controlling transcription, and the RBS and stop codon controlling translation.

The transcription and translation profiles clearly show the beginning and end of both the transcript produced and protein coding region, with sharp increases and decreases at the relevant points (**Figure 3A**). Induction causes a large increase in the number of *lacZ* transcripts from 0.18 to 110 copies per cell. This is directly observed in the transcription profile, while the translation profile remains stable across conditions. Analysis of the profiles shows that the P_{tac} promoter has a transcription initiation rate of 0.0009 RNAP/s in the absence and 0.73 RNAP/s in the presence of IPTG (1 mM) (**Figure 3B**). The RBS for the *lacZ* gene has consistent translation initiation rates of between 0.21 and 0.35 ribosomes/s. Similarly, both the transcriptional terminator and stop codon showed similar efficiencies of 0.93–0.95 and 0.9–0.93, respectively. This is as expected given that both these processes act independently per transcript.

Characterizing a synthetic pseudoknot that induces translational recoding

Pseudoknots (PKs) are stable tertiary structures that regulate gene expression and combined with slippery sequences are frequently employed in compact viral genomes to stimulate translation recoding and produce multiple protein products from a single gene [48,50,55,56]. PK are the most common type of structures used to facilitate mostly –1 frameshifting [57] and in much rarer cases stimulate +1 frameshifting, e.g., in the eukaryotic antizyme genes [58]. PKs consist of a hairpin with an additional loop that folds back to stabilize the hairpin via extra base pairing (**Figure 4A**). In addition to stimulating recoding events, PKs regulate translational initiation, where they interfere with the RBS through antisense sequences that base pair with the binding site [51,59]. PKs located in the coding sequence induce –1 programmed frameshifting which is an evolutionary tool to reduce the length of sequence needed to encode a biological system and acts as a form of compression. The percentage of frameshifting reflects the stoichiometry of the translated proteins (e.g. capsule proteins for virus assembly), and helps overcome problems where the stochastic nature of transcription and translation make maintenance of specific ratios difficult [49].

Two elements signal and stimulate frameshifting. The first is a slippery site consisting of a heptanucleotide sequence of the form XXXYYYZ which enables out-of-zero-frame paring in the A or P site of the ribosome and facilitates recoding events. The second is a PK situated 6–8 nt downstream of the slippery site. The distance between the slippery site and the 5'-end of the PK

positions the mRNA in the entry channel of the 30S ribosomal subunit, allowing contact with the PK and pauses translation to provide an extended time window for frameshifting to occur [48].

To demonstrate our ability to characterize this process, we created an inducible genetic construct (referred to as PK-LacZ) that incorporated a virus-inspired PK structure within its natural content (*gene10*) fused to *lacZ* in a –1 frame (**Figure 4A**) [60]. A slippery site UUUAAAG preceded the PK. *Gene10* of bacteriophage T7 produces two proteins, one through translation in the zero frame and one through a –1 frameshift; both protein products constitute the bacteriophage capsid [61].

We generated translation profiles to assess ribosome flux along the entire construct (**Figure 4B**). The profiles showed high-levels of translation up to the PK with a major drop of 80–90% at the PK to the end of the *gene10* coding region, and a further drop of ~97% after this region (**Figure 4B**).

To analyze the frameshifting within *gene10*, we divided the construct into three regions: (1) the *gene10* gene up to the slippery site, (2) the middle region, which covers the slippery site along with the PK up to the *gene10* stop codon, and (3) the downstream *lacZ* gene in a –1 frame. For each region, we calculated the fraction of RPFs in each frame as a total of all three possible frames. We found that the zero and –1 frames dominate the *gene10* and *lacZ* regions, respectively, with >46% of all RPFs being found in these frames (top row, **Figure 4C**). The middle region saw a greater mix of all three, while the lacZ region saw a further drop in the zero-frame. This is likely due to a combination of ribosomes that have passed the PK successfully and terminated in zero-frame at the end of *gene10* and those that have frameshifted. Similar results were found with and without induction by IPTG (**Figure 4C**). An identical analysis of the reading frames from the RNA-seq data revealed no specific frame was preferred, with equal fractions of each (bottom row, **Figure 4C**). This suggests the reading frames recovered for the RPFs were not influenced by any sequencing bias. We further tested if the major translation frame could be recovered by analyzing the entire genome and measuring the fraction of each frame across every gene. The correct zero-frame dominated in most cases (**Figure 4D**).

Finally, to calculate the efficiency of frameshifting by the PK, we compared the density of RPFs per nucleotide for the middle and *lacZ* regions. Because the PK is known to cause ribosomes to stall, the assumption of constant ribosome speed is broken for the *gene10* region before the PK. Therefore, when calculating the frameshifting efficiency using Equation 4, x_s and x_0 were set to the start and end nucleotide of the middle region, directly downstream of the PK where pausing was not expected to occur. We found that the PK caused 2–3% of ribosomes to frameshift, ~3-fold less than the 10% reported for the PK in its natural context [49].

Cellular response to a strong synthetic pseudoknot

It is known that expression of strong PKs can severely impact cell growth, but the reason for this remains unclear [60]. We observed a large number of RPF reads within the *gene10* region (**Figure 4B**). However, because PKs are known to cause ribosomal stalling [48] many of these reads may not capture actively translating ribosomes. Such stalling leads to an abundance of partially formed protein products and limits availability of translational resources, raising the question as to whether expression of the PK-LacZ construct elicits cellular stress by sequestering ribosomes.

To better understand the burden that expression of both *lacZ* and *PK-lacZ* had on the cell, we compared shifts in the transcription (mRNA counts) and translational efficiency [39] of endogenous genes after induction by IPTG (**Figure 5A**; **Supplementary Data S3**). No major changes were observed for the LacZ construct (**Figure 5A**). In contrast, the PK-LacZ construct caused significant shifts in the regulation of 491 genes (**Supplementary Data S4**). Of these, 341 were transcriptionally and 204 translationally regulated, with little overlap (54 genes) between the different types of regulation (**Figure 5B**). Transcriptional regulation mostly acted to reduce gene expression, while similar numbers of genes were translationally up and down regulated. Gene ontology (GO) analysis revealed a clustering of transcriptionally downregulated genes in categories mostly linked to translation, e.g. ribosomal proteins, amino acid biosynthesis, amino acid activation (amino acyl synthetases), and genes involved in respiration and catabolism (**Supplementary Data S5**). Transcriptionally upregulated genes were associated with ATP binding, chaperones (*ftsH*, *lon*, *clpB*, *dnaJK*, *groLS*, *htpG*), ion binding, proteolytic activities (*ftsH*, *prlC*, *htpX*), and an endoribonuclease (*ybeY*). Interestingly, all of these are under σ^{32} regulation. Since we do not subject cells to heat shock, upregulation of σ^{32}-dependent proteins most likely occurs via DnaK/DnaJ and GroEL/S release of σ^{32} that is already present in the cell upon binding of misfolded proteins from ribosomes stalled at the PK. The DnaK/DnaJ and GroEL/S chaperone systems negatively regulates σ^{32} by binding to it. When the ratio of chaperones to misfolded proteins is shifted, the chaperones release σ^{32}, which can then bind to the RNA polymerases under its control and induce heat shock genes [62,63]. This is supported by the fact that *dnaJ*, *groL/S*, and *grpE* were transcriptionally upregulated during PK induction as well as *ftsh*, which encodes the protease that degrades σ^{32}. The precise mechanism of σ^{32} upregulation by expressing synthetic constructs is not known, although it has been reported to be a general response [64].

To test whether this stress caused changes in translational dynamics (e.g. ribosome pausing at particular codons), we next compared the change in the dwell time of ribosomes at each codon (also known as codon occupancy) across the genome after induction [52]. Notable increases in occupancy were found for the codons AGA, CTA, CCC, TCC, which encode for arginine, leucine, proline and serine, respectively (**Figure 5C**). All of these codons are rarely used in the genome for

their cognate amino acid but were found in higher proportions across *gene10*. For example, the CTA codon that codes for leucine is only used by 4% of codons in the genome, while accounted for 8% in the *gene10* region. When coupled with the strong expression of *gene10*, the stress induced by this abnormal demand on cellular resources would be amplified.

The broad shifts in cell-wide regulation and changes in codon occupancy suggest that expression of *PK-lacZ* may significantly limit the availability of shared cellular resources. From a translational perspective, this would manifest as a cell-wide drop in translation initiation rate as the pool of free ribosomes would be reduced [14]. To test this hypothesis, we compared the RBS initiation rates of endogenous genes before and after induction and found a consistent reduction across all genes for both synthetic constructs (**Figure 5D**; **Supplementary Data S1**). While relatively small for the LacZ construct (18%) where no notable stress was seen, the PK-LacZ construct triggered a large (43%) drop in translation initiation rates across the cell (**Figure 5D**). Analysis of transcriptome composition and distribution of engaged ribosomes across these further revealed that the PK-LacZ construct made up 40% of all mRNAs and captured 47% of the shared ribosome pool engaged in translation (**Figure 5E**). This would account for the global drop in translation initiation rates and agglomeration of the partially translated capsule protein Gene10 by these ribosomes would help explain the strong σ^{32}-mediated response.

Discussion

In this work, we present a new methodology to quantify transcription and translation in living cells at a nucleotide resolution. This is based on a sequencing workflow that combines a modified version of RNA-seq and Ribo-seq with measures of key cellular parameters, and then use biophysical models to interpret this data. We show that our approach is able to simultaneously characterize the translation initiation rate of RBSs and termination efficiency of stop codons across the *E. coli* transcriptome: in addition, we measure the precise behavior of the genetic parts controlling transcription and translation in a synthetic genetic construct that expresses *lacZ*. Because our approach is based on sequencing it can scale beyond what is possible with commonly used fluorescence protein-based approaches, and through the use of spike-in standards we are able to give measurements in absolute units. Specifically, transcription and translation rates in RNAP/s and ribosomes/s units.

To demonstrate the ability to quantitatively assess various translational processes that have been difficult to measure, we studied the behavior of a genetic construct that contains a strong virus-inspired PK structure that induces a translational frameshift. Following expression of *PK-lacZ*, the main reading frame shifts, but at an efficiency ~3-fold less in in PKs native viral context. In contrast

to *lacZ* expression, *PK-lacZ* also causes a major burden to the cell. We observe transcriptome-wide increases in ribosome dwell times at rare codons encoding arginine, leucine, proline and serine. These codons are used far more frequently in the synthetic construct than in the *E. coli* genome, suggesting that strong expression of gene places significant demands on the translational resources of the cell. This burden also results in significant changes in gene regulation (both transcriptional and translational), which was mediated by the alternative polymerase subunit, σ^{32} that remodels the bacterial proteome following thermal stress [65]. The likely cause of σ^{32} activation is the partial unfolding of translation products from the ribosomes stalled at the PK [48,65]. Supportive of this mechanism is the upregulation of the DnaJ/K and GroEL/ES/GrpE chaperone complexes that modulate the release of σ^{32} and activation of the stress response [62,63]. To our knowledge, the stress response induced by a strong pseudoknot has not been reported before. Our findings highlight the importance of considering the burden a synthetic construct places on a cell (e.g. the sequestering shared resources like RNAPs or ribosomes) and their indirect effects on integrated stress responses [13,14,40,64].

Mature engineering fields rely on predictive models to efficiently design complex systems by reducing the need to physically construct and test each iteration. To date, the accuracy of models in synthetic biology have been hampered by a lack of reliable and quantitative measurements of genetic parts and devices, and their effects on the host cell. Attempts have been made to improve this situation by using standard calibrants to increase reproducibility across labs and equipment [66–68] and by including synthetic RNA spike-ins to enable absolute quantification of transcription [69]. Our methodology complements these efforts by combining multiple sequencing methods and spike-in standards to quantify the two major cellular processes controlled by genetic circuits (i.e. transcription and translation). As we attempt to implement ever more complex functionalities in living cells [26] and push towards a deeper understanding of the processes sustaining life, scalable and comprehensive methodologies for quantitative measurement of these processes are paramount. Such capabilities will move us beyond a surface-level view of living cells to one that allows the exploration of their inner most regulation and homeostasis.

Materials and Methods

Strain, media, and inducers. The *E. coli* K12 strain, [K-12, *recA1* Δ*(pro-lac) thi ara F':lacIq1 lacZ::Tn5 proAB*+], harbours a pBR322-derived plasmid containing either *lacZ* with a fragment insert that contains a truncated *lac* operon with the P_{tac} promoter and the wildtype *lacZ* under *lacI* control, or a pseudoknot-lacZ (*PK-lacZ*) consisting of *gene10*, a virus-derived RNA pseudoknot [60], 22/6a, fused upstream of the *lacZ*. Bacteria were grown in MOPS minimal medium supplemented with

0.4% glycerol, 2.5 µg/ml vitamin B1, 100 µg/ml ampicillin, 20 µg/ml kanamycin and additionally 50 µg/ml arginine for the *lacZ* expressing strain. The cells were grown for at least 10 generations at 37°C to ensure stable exponential growth before induction.

Gene expression and preparation of the sequencing libraries. LacZ and *PK-lacZ* expression were induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM at OD_{600} ≈ 0.4 for 10 min and 15 min, respectively. One aliquot of each culture was used to isolate RPFs and prepare the cDNA library for Ribo-seq as described in Bartholomäus *et al*. [41] In parallel, from another aliquot, total RNA was isolated with TRIzol (Invitrogen) and subjected to random alkaline fragmentation for RNA-seq as described in Bartholomäus *et al*. [41] The total RNA was spiked in with RNA standards (ERCC RNA Spike-In Mix; Ambion) which were used to calculate the copy numbers per cell. Total protein concentration (g wet mass per ml culture) were determined by the Bradford assay using serial dilutions of the cells from the exponential phase until the induction time at OD 0.4 and following induction with 1 mM IPTG. Using the cell number and the volume of *E. coli* as 1 femtoliter, the protein mass was recalculated as g wet protein mass per cell.

Processing of sequencing data. Sequenced reads were quality trimmed using fastx-toolkit version 0.0.13.2 (quality threshold: 20), sequencing adapters were cut using cutadapt version 1.8.3 (minimal overlap: 1 nt) and the reads were uniquely mapped to the genome of *E. coli* K-12 MG1655 strain using Bowtie version 1.1.2 allowing a maximum of two mismatches. *LacZ* and other similar parts of the plasmids were masked in the genome. Reads aligning to more than one sequence including tRNA and rRNA were excluded from the data. The raw reads were used to generate gene read counts by counting the number of reads whose middle nucleotide (for reads with an even length the nucleotide 5' of the mid-position) fell in the coding sequence (CDS). Gene read counts were normalized by the length of the unique CDS per kilobase (RPKM units) and the total mapped reads per million (RPM units) [42]. Biological replicates were performed for all sequencing reactions. Based on the high correlation between replicates (**Supplementary Figure S1**), reads from both biological replicates were merged into metagene sets [39]. Differential gene expression was performed using DESeq2 version 1.20 ($P < 0.01$ for the translational efficiency data, and $P < 0.001$ and a $\log_2$ fold-change > 1.37 for the mRNA counts due to very high correlation in this data) and the GO terms with significant enrichment ($P < 0.01$) were calculated using GO.db version 2.10. *P*-values were adjusted for multiple testing using false-discovery rate (FDR) according to Benjamini and Hochberg.

Calculating absolute transcript numbers. To calculate the transcript copy number, we used a method previously described by Bartholomäus *et al*. and Mortazavi *et al*. [41,42]

Calibration of ribosome profiling reads. RPFs were binned in groups of equal read length, and each group was aligned to the stop codons as described previously by Mohammad *et al*. [70] For each read length we calculated the distance between the point a transcript leaves the ribosome and the middle nucleotide in the P site, and used this distance to determine the center of each P site codon along each mRNA. As expected, the majority of our sequence reads were 23–28 nt and these read lengths were used for the further analysis. The ribosome dwelling occupancy per codon over the whole transcriptome was calculated as described by Lareau *et al*. [52], where the reads over each position within a gene were normalized to the average number of footprints across this gene. Metagene analysis of the ribosome occupancies within the start and stop codon regions was performed as described by Baggett *et al*. [71] Thereby, only genes with at least 5 RPFs in the chosen window were considered. Overlapping genes were excluded from the analysis.

Data analysis and visualization. Data analysis was performed using custom scripts run with R version 3.4.4 and Python version 3.6.3. Plots was generated using matplotlib version 2.1.2 and genetic constructs were visualized using DNAplotlib version 1.0 [72] with Synthetic Biology Open Language Visual (SBOLv) notation [73].

Data availability. Sequencing data from RNA-Seq and Ribo-Seq were deposited in the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra/) under accession number SRP144594.

Acknowledgements

We thank Alexander Bartholomäus for the initial mapping and earlier data analysis. This work was supported by BrisSynBio, a BBSRC/EPSRC Synthetic Biology Research Centre (grant BB/L01386X/1), a Royal Society University Research Fellowship (grant UF160357) to T.E.G., the MOLPHYSX program of the University of Copenhagen to S.P., and by the European Union (grants NICHE ITN and SynCrop ETN) to Z.I.

Author Contributions

Z.I. and T.E.G. conceived of the study. M.E. performed the sequencing experiments, P.N. performed the quantitative determination of cellular parameter. S.P. provided the LacZ and PK-LacZ constructs

and advised the experimental acquisition of sequencing data. T.E.G. developed the biophysical models. I.C. processed the sequencing data. Z.I., T.E.G. and I.C. analyzed the data. Z.I., T.E.G. and I.C. wrote the manuscript.

Conflict of Interest

The authors declare no competing financial interest.

References

1. Belliveau, N. M. *et al.* Systematic approach for dissecting the molecular mechanisms of transcriptional regulation in bacteria. *Proc. Natl. Acad. Sci.* **115,** E4796–E4805 (2018).
2. Wang, L.-Z., Wu, F., Flores, K., Lai, Y.-C. & Wang, X. Build to understand: synthetic approaches to biology. *Integr. Biol.* **8,** 394–408 (2016).
3. Smanski, M. J. *et al.* Synthetic biology to access and expand nature's chemical diversity. *Nat. Rev. Microbiol.* **14,** 135–149 (2016).
4. Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. *Nat. Methods* **11,** 508–20 (2014).
5. Fernandez-Rodriguez, J., Yang, L., Gorochowski, T. E., Gordon, D. B. & Voigt, C. A. Memory and Combinatorial Logic Based on DNA Inversions: Dynamics and Evolutionary Stability. *ACS Synth. Biol.* **4,** (2015).
6. Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. & Voigt, C. A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491,** 249–253 (2012).
7. Daniel, R., Rubens, J. R., Sarpeshkar, R. & Lu, T. K. Synthetic analog computation in living cells. *Nature* **497,** 619–623 (2013).
8. Tan, S. Z., Manchester, S. & Prather, K. L. J. Controlling Central Carbon Metabolism for Improved Pathway Yields in Saccharomyces cerevisiae. *ACS Synth. Biol.* **5,** 116–124 (2016).
9. Nielsen, A. A. & Voigt, C. A. Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. *Mol. Syst. Biol.* **10,** 763–763 (2014).
10. Cardinale, S., Joachimiak, M. P. & Arkin, A. P. Effects of genetic variation on the E. coli host-circuit interface. *Cell Rep.* **4,** 231–237 (2013).
11. Poole, E. S., Brown, C. M. & Tate, W. P. The identity of the base following the stop codon determines the efficiency of in vivo translational termination in Escherichia cobl. *EMBO J.* **14,** 151–158 (2000).
12. Ceroni, F., Algar, R., Stan, G. B. & Ellis, T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Methods* **12,** 415–418 (2015).
13. Gyorgy, A. *et al.* Isocost Lines Describe the Cellular Economy of Genetic Circuits. *Biophys. J.* **109,** 639–646 (2015).
14. Gorochowski, T. E., Avcilar-Kucukgoze, I., Bovenberg, R. A. L., Roubos, J. A. & Ignatova, Z. A Minimal Model of Ribosome Allocation Dynamics Captures Trade-offs in Expression between Endogenous and Synthetic Genes. *ACS Synth. Biol.* **5,** (2016).
15. Gorochowski, T. E., Van Den Berg, E., Kerkman, R., Roubos, J. A. & Bovenberg, R. A. L. Using synthetic biological parts and microbioreactors to explore the protein expression

characteristics of escherichia coli. *ACS Synth. Biol.* **3,** 129–139 (2014).

16. Wohlgemuth, S. E., Gorochowski, T. E. & Roubos, J. A. Translational sensitivity of the Escherichia coli genome to fluctuating tRNA availability. *Nucleic Acids Res.* **41,** 8021–8033 (2013).
17. Salis, H. M., Mirsky, E. a & Voigt, C. a. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27,** 946–50 (2009).
18. Espah Borujeni, A., Channarasappa, A. S. & Salis, H. M. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.* gkt1139- (2013). doi:10.1093/nar/gkt1139
19. Seo, S. W. *et al.* Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. *Metab. Eng.* **15,** 67–74 (2013).
20. Shendure, J. *et al.* DNA sequencing at 40: Past, present and future. *Nature* **550,** 345–353 (2017).
21. Yang, L. *et al.* Permanent genetic memory with >1-byte capacity. *Nat. Methods* **11,** 1261–1266 (2014).
22. Siuti, P., Yazbek, J. & Lu, T. K. Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **31,** 448–452 (2013).
23. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10,** 354–360 (2013).
24. Hecht, A. *et al.* Measurements of translation initiation from all 64 codons in E. coli. *Nucleic Acids Res.* **45,** 3615–3626 (2017).
25. Jones, D. L., Brewster, R. C. & Phillips, R. Promoter architecture dictates cell-to-cell variability in gene expression. *Science* **346,** 1533–1536 (2014).
26. Nielsen, A. A. K. *et al.* Genetic circuit design automation. *Science* **352,** (2016).
27. Cambray, G. *et al.* Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res.* **41,** 5139–5148 (2013).
28. Canton, B., Labno, A. & Endy, D. Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.* **26,** 787–93 (2008).
29. Snapp, E. Design and Use of Fluorescent Fusion Proteins in Cell Biology. *Curr. Protoc. Cell Biol.* (2005). doi:10.1002/0471143030.cb2104s27.Design
30. Baens, M. *et al.* The dark side of EGFP: Defective polyubiquitination. *PLoS One* **1,** 1–6 (2006).
31. Margolin, W. The price of tags in protein localization studies. *J. Bacteriol.* **194,** 6369–6371 (2012).

32. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17,** 333–351 (2016).
33. Lieberman-aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **33292,** 289–294 (2009).
34. Lucks, J. B. *et al.* Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci.* **108,** 11063–11068 (2011).
35. Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17,** 1–19 (2016).
36. Ingolia, N. T. Ribosome profiling: New views of translation, from single codons to genome scale. *Nat. Rev. Genet.* **15,** 205–213 (2014).
37. Mayer, A. *et al.* Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161,** 541–544 (2015).
38. Li, G. W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157,** 624–635 (2014).
39. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324,** 218–23 (2009).
40. Gorochowski, T. E. *et al.* Genetic circuit characterization and debugging using RNA-seq. *Mol. Syst. Biol.* **13,** (2017).
41. Bartholomäus, A. *et al.* Bacteria differently regulate mRNA abundance to specifically respond to various stresses. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **374,** 20150069 (2016).
42. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5,** 621–628 (2008).
43. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466,** 835–840 (2010).
44. Del Campo, C., Bartholomäus, A., Fedyunin, I. & Ignatova, Z. Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function. *PLoS Genet.* **11,** 1–23 (2015).
45. Gorochowski, T. E., Ignatova, Z., Bovenberg, R. A. L. & Roubos, J. A. Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate. *Nucleic Acids Res.* **43,** 3022–3032 (2015).
46. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **11,** 113–127 (2010).

47. Arribere, J. A. *et al.* Translation readthrough mitigation. *Nature* **534,** 719–723 (2016).
48. Giedroc, D. P. & Cornish, P. V. Frameshifting RNA pseudoknots: Structure and mechanism. *Virus Res.* **139,** 193–208 (2009).
49. Condron, B. G., Atkins, J. F. & Gesteland, R. F. Frameshifting in gene 10 of bacteriophage T7. *J. Bacteriol.* **173,** 6998–7003 (1991).
50. Tsuchihashi, Z. & Kornberg, A. Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proc. Natl. Acad. Sci.* **87,** 2516–2520 (1990).
51. Bordeau, V. & Felden, B. Curli synthesis and biofilm formation in enteric bacteria are controlled by a dynamic small RNA module made up of a pseudoknot assisted by an RNA chaperone. *Nucleic Acids Res.* **42,** 4682–4696 (2014).
52. Lareau, L. F., Hite, D. H., Hogan, G. J. & Brown, P. O. Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *Elife* **2014,** 1–16 (2014).
53. Woolstenhulme, C. J., Guydosh, N. R., Green, R. & Buskirk, A. R. High-Precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. *Cell Rep.* **11,** 13–21 (2015).
54. Kennell, D. & Riezman, H. Transcription and translation initiation frequencies of the Escherichia coli lac operon. *J. Mol. Biol.* **114,** 1–21 (1977).
55. Brierley, I., Pennell, S. & Gilbert, R. J. C. Viral RNA pseudoknots: Versatile motifs in gene expression and replication. *Nat. Rev. Microbiol.* **5,** 598–610 (2007).
56. Sharma, V. *et al.* Analysis of tetra-and hepta-nucleotides motifs promoting-1 ribosomal frameshifting in Escherichia coli. *Nucleic Acids Res.* **42,** 7210–7225 (2014).
57. Atkins, J. F., Loughran, G., Bhatt, P. R., Firth, A. E. & Baranov, P. V. Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* **44,** 7007–7078 (2016).
58. Ivanov, I. P., Anderson, C. B., Gesteland, R. F. & Atkins, J. F. Identification of a new antizyme mRNA +1 frameshifting stimulatory pseudoknot in a subset of diverse invertebrates and its apparent absence in intermediate species. *J. Mol. Biol.* **339,** 495–504 (2004).
59. Unoson, C. & Wagner, E. G. H. Dealing with stable structures at ribosome binding sites: Bacterial translation and ribosome standby. *RNA Biol.* **4,** 113–117 (2007).
60. Tholstrup, J., Oddershede, L. B. & Sørensen, M. A. MRNA pseudoknot structures can act as ribosomal roadblocks. *Nucleic Acids Res.* **40,** 303–313 (2012).
61. Condron, B. G., Gesteland, R. F. & Atkins, J. F. An analysis of sequences stimulating frameshifting in the decoding of gene 10 of bacteriophage T7. *Nucleic Acids Res.* **19,** 5607–

5612 (1991).

62. Mogk, A., Huber, D. & Bukau, B. Integrating protein homeostasis strategies in prokaryotes. *Cold Spring Harb. Perspect. Biol.* **3,** 1–19 (2011).
63. Guisbert, E., Herman, C., Lu, C. Z. & Gross, C. A. A chaperone network controls the heat shock response in E. coli. *Genes Dev.* **18,** 2812–2821 (2004).
64. Ceroni, F. *et al.* Burden-driven feedback control of gene expression. *Nat. Methods* 177030 (2018). doi:10.1101/177030
65. Guo, M. S. & Gross, C. A. Stress-induced remodeling of the bacterial proteome. *Curr. Biol.* **24,** R424–R434 (2014).
66. Castillo-Hair, S. M. *et al.* FlowCal: A User-Friendly, Open Source Software Tool for Automatically Converting Flow Cytometry Data from Arbitrary to Calibrated Units. *ACS Synth. Biol.* **5,** 774–780 (2016).
67. Davidsohn, N. *et al.* Accurate Predictions of Genetic Circuit Behavior from Part Characterization and Modular Composition. *ACS Synth. Biol.* **4,** 673–681 (2015).
68. Beal, J. *et al.* Reproducibility of fluorescent expression from engineered biological constructs in E. coli. *PLoS One* **11,** 1–22 (2016).
69. Owens, N. D. L. *et al.* Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. *Cell Rep.* **14,** 632–647 (2016).
70. Mohammad, F., Woolstenhulme, C. J., Green, R. & Buskirk, A. R. Clarifying the Translational Pausing Landscape in Bacteria by Ribosome Profiling. *Cell Rep.* **14,** 686–694 (2016).
71. Baggett, N. E., Zhang, Y. & Gross, C. Global analysis of translation termination in E. coli. *PLoS Genet.* **13,** e1006676 (2017).
72. Der, B. S. *et al.* DNAplotlib: Programmable Visualization of Genetic Designs and Associated Data. *ACS Synth. Biol.* **6,** (2017).
73. Myers, C. J. *et al.* A standard-enabled workflow for synthetic biology. *Biochem. Soc. Trans.* **45,** (2017).

Figures and Captions

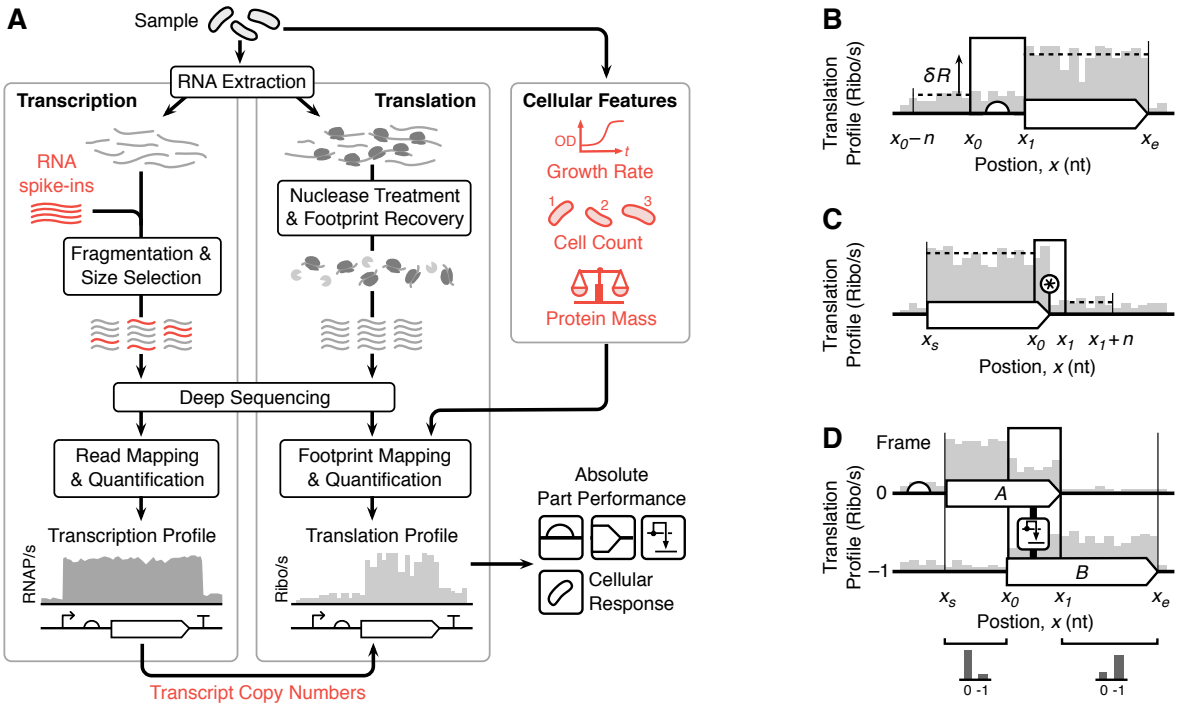

Figure 1: Overview of the workflow. (**A**) Major steps involved when quantifying transcription (RNA-seq) and translation (Ribo-seq) and the additional cellular features measured. Elements required for quantification in absolute units are highlighted in red. (**B**) Model for calculating the translation initiation rate of a ribosome binding site (Eq. 2). (**C**) Model for calculating the termination efficiency of a stop codon (Eq. 3). Star denotes the location of the stop codon. (**D**) Model for calculating translational frameshifting efficiency between two coding regions 'A' and 'B' in zero and –1 reading frames, respectively (Eq. 4).

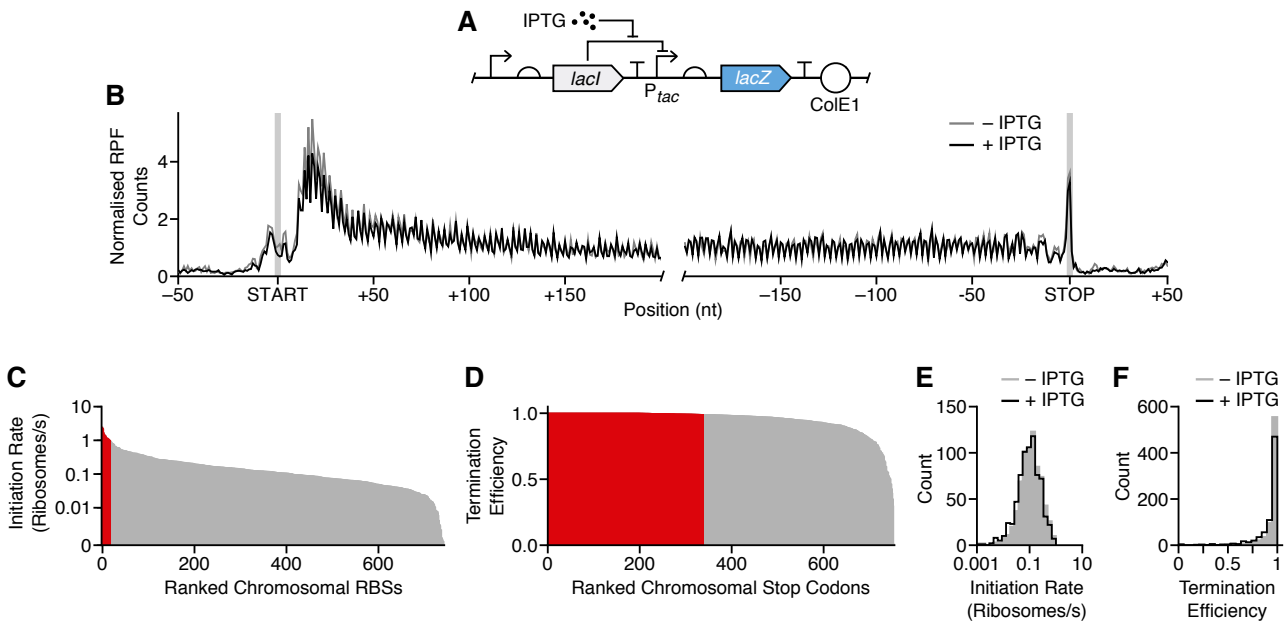

Figure 2: Measuring translation initiation and termination signals across the *E. coli* transcriptome. (**A**) Genetic design of the LacZ reporter construct whose expression is activated by the inducer IPTG. (**B**) Normalized RPF count profile averaged for all *E. coli* transcripts. Profiles generated for cells grown in the absence and presence of IPTG (1 mM). Start and stop codons are shaded. (**C**) Bar chart of all measured RBS initiation rates ranked by their strength. Strong RBSs with initiation rates >1 ribosome/s are highlighted in red. (**D**) Distribution of initiation rates for cells grown in the absence and presence of IPTG (1 mM). (**E**) Bar chart of all measured stop codon termination efficiencies ranked by their strength. Stop codons with termination efficiency >0.99 are highlighted in red. (**F**) Distribution of stop codon termination efficiencies for cells grown in the absence and presence of IPTG (1 mM).

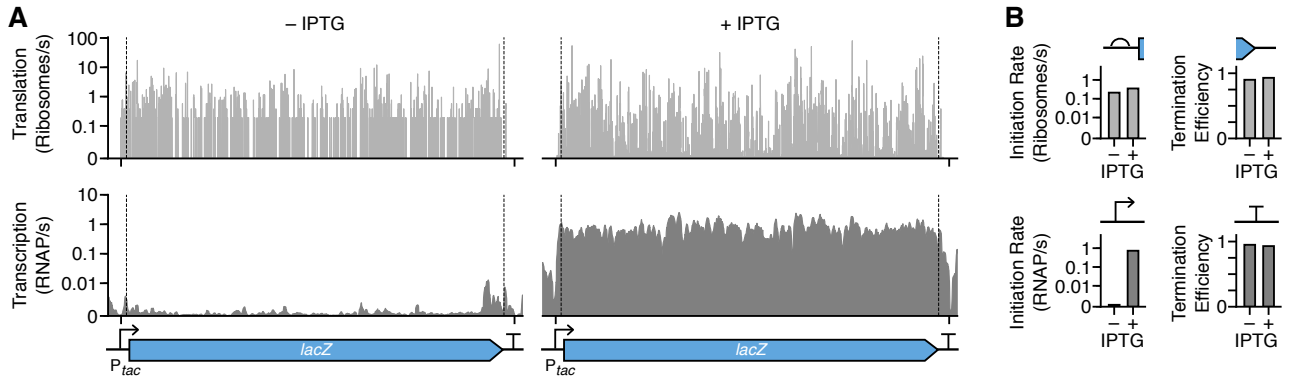

Figure 3: Simultaneous quantification of transcription and translation in a synthetic genetic construct. (**A**) Transcription (bottom) and translation (top) profiles for *lacZ*, computed from the RNA-seq and Ribo-seq data, respectively. Profiles are shown for cells in the absence and presence of IPTG (1 mM). Position of genetic parts and gene is shown below the profiles. RBS is omitted from the genetic design due to its size. (**B**) Measured performance of promoter strength in RNAP/s units, RBS strength and initiation rate in ribosomes/s units, and the transcriptional terminator and stop codon termination efficiency for *lacZ*. Data shown for cells in the absence and presence of IPTG (1 mM).

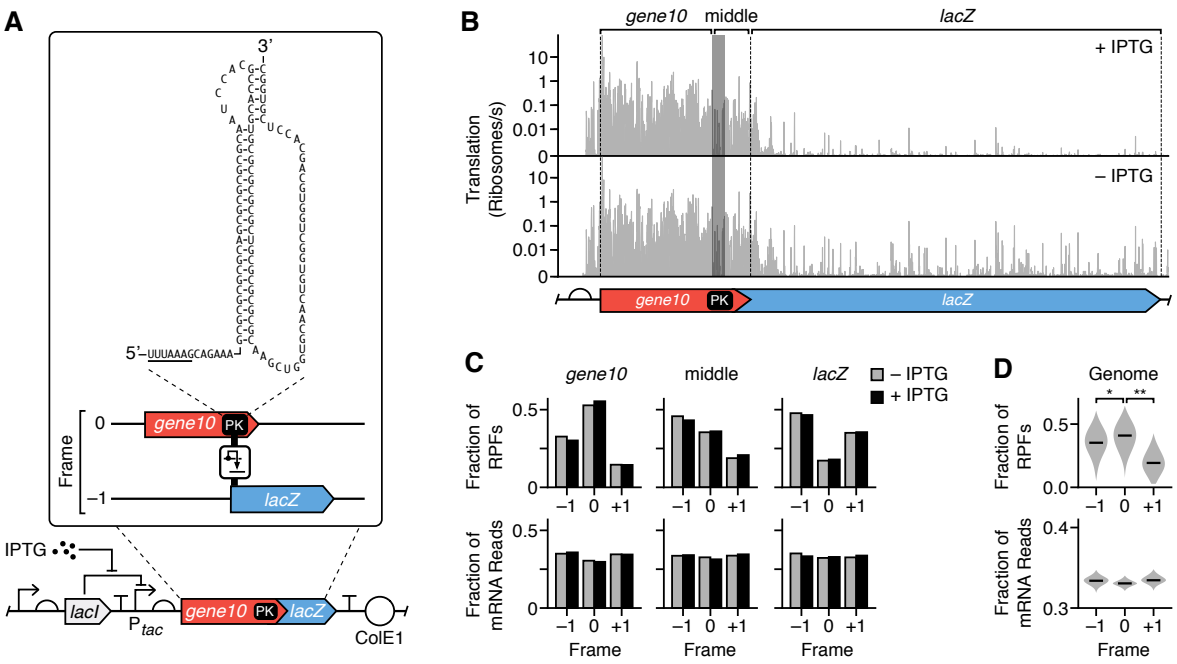

Figure 4: Characterization of a synthetic pseudoknot construct that induces translational frameshifting. (**A**) Genetic design of the PK-LacZ construct. Expanded sequence shows the PK secondary structure with the slippery site underlined, as well as the two genes (*gene10* and *lacZ*) in differing reading frames. (**B**) Translation profiles for the PK-LacZ construct in cells cultured in the absence (bottom) and presence (top) of IPTG (1 mM). The *gene10*, middle, and *lacZ* regions are labelled above the profiles. Shaded region denotes the PK, and dashed lines denote the start codon and stop codons of *gene10* and *LacZ*. (**C**) Fraction of the total RPFs and mRNA reads in each reading frame for the *gene10*, PK or middle, and *lacZ* regions schematically shown below and are of the PK-LacZ construct. Data shown separately for cell cultured in the absence and presence of IPTG (1 mM). (**D**) Violin plots of the distributions of fractions of total RPFs and mRNA reads in each reading frame for all *E. coli* transcripts. Median values shown by horizontal bars. *, $P = 0.049$; **, $P = 1.6 \times 10^{-9}$.

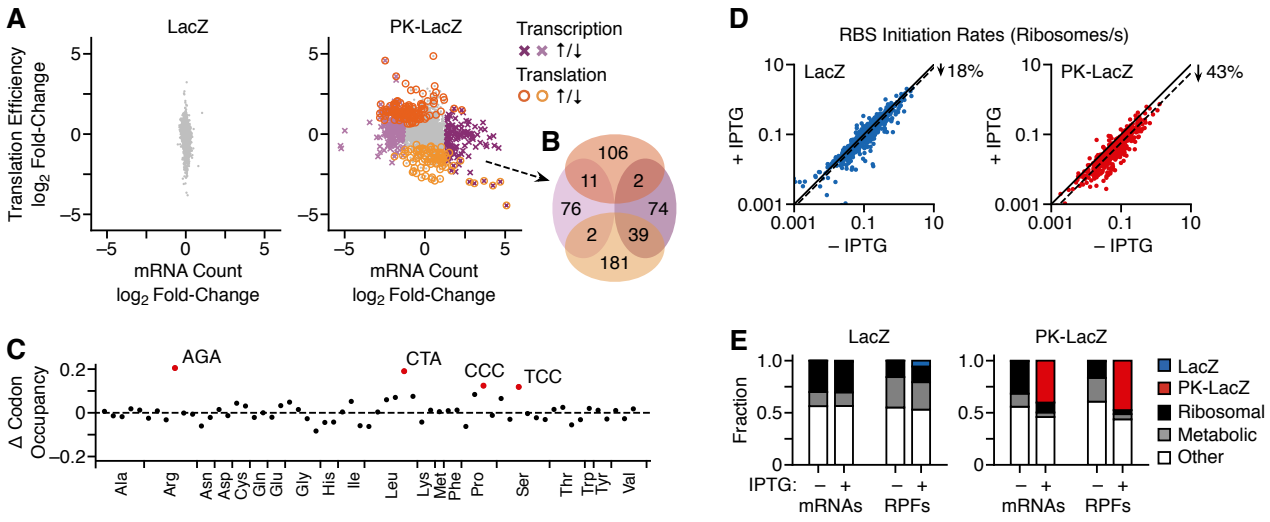

Figure 5: Cellular response to the expression of a synthetic pseudoknot construct. (**A**) Change in expression of chromosomal genes in *E. coli* cells following induction of *PK-lacZ* expression (1 mM IPTG). Each point represents a transcript. Differentially expressed genes (mRNA count: $P < 0.001$ and absolute $\log_2$ fold-change > 1.37; translation efficiency: $P < 0.01$) are highlighted in color and by an alternative point shape (transcriptional regulation: purple cross; translational regulation: orange open circle). (**B**) Venn diagram of genes significantly regulated transcriptionally and translationally after induction of the PK-LacZ construct. Colors match those in panel A. (**C**) Change in codon occupancy for cells harboring the PK-LacZ construct after induction by IPTG (1 mM) calculated from the Ribo-seq data. Each point corresponds to a codon, which are ordered by amino acid identity and then by abundance in the genome (left most abundant, right least abundant). Dashed horizontal line denotes no change. Outliers are labelled and highlighted in red (Tukey test: 1.5 times the interquartile range below the first quartile or above the third quartile). (**D**) Translation initiation rates for all *E. coli* RBSs in cells harboring the LacZ and PK-LacZ constructs in the absence and presence of IPTG (1 mM). Solid line shows the same initiation rate for both conditions. Dotted lines denote linear regressions for the data with no offset. (**E**) Fractions of mRNA reads and RPFs mapping to each synthetic expression construct (LacZ and PK-LacZ) and *E. coli* transcripts, which are divided into three major categories: ribosomal, metabolic, and other functions. Data shown for cells cultured in the absence and presence of IPTG (1 mM).