nature chemical biology

ARTICLE

PUBLISHED ONLINE: 16 SEPTEMBER 2012 | DOI: 10.1038/NCHEMBIO.1070

Discovery and biological characterization of geranylated RNA in bacteria

Christoph E Dumelin[1,2], Yiyun Chen[1,2], Aaron M Leconte[1,2], Y Grace Chen[1,2] & David R Liu[1,2]*

A general MS-based screen for unusually hydrophobic cellular small molecule–RNA conjugates revealed geranylated RNA in *Escherichia coli*, *Enterobacter aerogenes*, *Pseudomonas aeruginosa* and *Salmonella enterica* var. Typhimurium. The geranyl group is conjugated to the sulfur atom in two 5-methylaminomethyl-2-thiouridine nucleotides. These geranylated nucleotides occur in the first anticodon position of $tRNA^{Glu}_{UUC}$, $tRNA^{Lys}_{UUU}$ and $tRNA^{Gln}_{UUG}$ at a frequency of up to 6.7% (~400 geranylated nucleotides per cell). RNA geranylation can be increased or abolished by mutation or deletion of the *selU* (*ybbB*) gene in *E. coli*, and purified SelU protein in the presence of geranyl pyrophosphate and tRNA can produce geranylated tRNA. The presence or absence of the geranyl group in $tRNA^{Glu}_{UUC}$, $tRNA^{Lys}_{UUU}$ and $tRNA^{Gln}_{UUG}$ affects codon bias and frameshifting during translation. These RNAs represent the first reported examples of oligoisoprenylated cellular nucleic acids.

The number of known roles of RNA in living systems has rapidly expanded over the past three decades. The development and application of methods such as bioinformatic analysis, oligonucleotide microarray technology and high-throughput sequencing have revealed that RNA is an active contributor to many biological processes, including gene regulation[1,2], catalysis[3] and viral defense[4]. Although the known biological repertoire of cellular RNA has grown dramatically, the observed chemical diversity of RNA has remained relatively unchanged over the past several decades. Indeed, approximately two-thirds of all known RNA modifications were reported before 1980 (ref. 5). The application of modern analytical methods to either a small subset of RNAs or to the study of specific modifications of RNA have enabled the recent discovery of agmatine-modified tRNA in archaea[6,7] and led to the elucidation of the biosynthesis of several nucleosides[8–10]. We recently developed general and highly sensitive chemical screens to reveal cellular small molecule–RNA conjugates in a broad and unbiased manner that is not limited to any particular class of modifications or any specific subset of RNAs. These screens were designed to capture chemically reactive[11] and hydrophilic[12] new RNA nucleotides, resulting in the discovery of coenzyme A–linked RNA and NAD-linked RNA isolated from several bacteria[11,12].

Large, hydrophobic small molecules conjugated to cellular RNAs could potentially influence the structure, function or subcellular localization of RNA. Such conjugates could also be evolutionary fossils from ancient RNA-mediated lipid synthesis, as has been previously speculated[13]. We therefore implemented a chemical screen to isolate and identify unusually hydrophobic small molecules conjugated to biological RNA. Here we report the discovery, structural elucidation and initial biological characterization of geranylated RNAs from bacterial cells. These findings represent the first reported examples of oligoisoprenylated nucleotides from cells.

RESULTS

Screen for hydrophobic small molecule–RNA conjugates

We modified our previously reported small molecule–RNA conjugate screen[12] to increase the likelihood of detecting unusually hydrophobic nucleotides. The complete experimental details are provided in the Methods and **Supplementary Methods**. Briefly, we isolated total cellular RNA isolated in a manner that minimizes contamination with small molecules, DNA or protein. We digested the resulting RNA pool with nuclease P1 and analyzed the products by high-resolution LC/MS. To ensure that the identified species were covalently attached to RNA, we treated an identical sample of RNA with heat-inactivated nuclease P1 and analyzed it in parallel using LC/MS (**Fig. 1a** and **Supplementary Methods**). Small molecules conjugated to RNA generate noncanonical nucleotides (or fragments thereof) in the nuclease P1–digested sample that are absent in the control sample treated with inactivated nuclease P1. To focus the screen on hydrophobic RNA modifications, we (i) performed size-exclusion chromatography after digestion in the presence of organic solvents, (ii) implemented a C8 reverse-phase column for LC/MS analysis and (iii) considered only species that eluted later than AMP-Trp, the most hydrophobic of the aminoacyl adenylates, for further evaluation. The application of this screen to RNA isolated from *E. coli* revealed eight unknown species that are more hydrophobic than AMP-Trp and that are enriched at least threefold in the active nuclease P1 samples relative to the control samples (**Supplementary Results**, **Supplementary Table 1**).

Two new hydrophobic small molecule–RNA conjugates

We chose two of these species, with $[M\text{-}H]^-$ m/z = 824.200 and $[M\text{-}H]^-$ m/z = 868.189, for further investigation (**Fig. 1b**) on the basis of their hydrophobicity, their high degree of enrichment and their relatedness; MS/MS fragmentation of either species resulted in ions consistent with uridine monophosphate, an unmodified ribose and an unknown group (307.171 Da or 351.161 Da) (**Fig. 2a**). We hypothesized that the unknown groups were two previously unidentified nucleobases and that breakage of the glycosidic bond between one of these nucleobases and a ribose gives rise to the observed UMP-ribose adduct ($[M\text{-}H]^-$ m/z = 517.029). Further MS/MS/MS fragmentation of the UMP-ribose adduct showed that the dinucleotides are of the sequence 5′-X-U-3′ (**Supplementary Fig. 1**).

To determine the molecular formula of the unknown nucleotides, we isolated RNA from *E. coli* grown with [^{13}C]glucose as the sole carbon source or [^{15}N]ammonium chloride as the sole nitrogen source. The resulting increase in mass of the parent and the daughter ions revealed the molecular formulas of the unknown dinucleotides

[1]Department of Chemistry & Chemical Biology, Harvard University, Cambridge, Massachusetts, USA. [2]Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts, USA. *e-mail: drliu@fas.harvard.edu

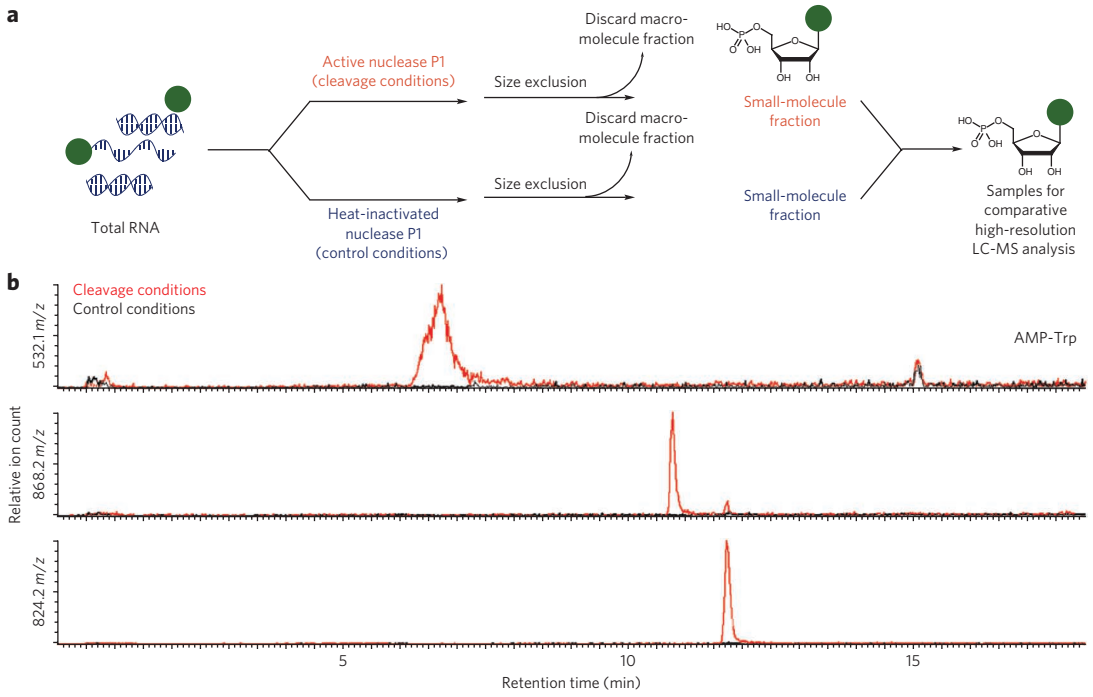

Figure 1 | Discovery of two hydrophobic small molecule–RNA conjugates with $[M-H]^-$ m/z = 824.200 and 868.189. (**a**) Scheme of the general method for discovering biological small molecule–RNA conjugates applied in this work. Total cellular RNA samples were treated with active nuclease P1 or with heat-inactivated nuclease P1 under otherwise identical conditions and then subjected to comparative LC-MS analysis. Nucleotide ions that are more abundant in the active nuclease sample than in the heat-inactivated nuclease (control) sample represent candidate cellular small molecule–RNA conjugates. (**b**) Extracted ion chromatogram of AMP-tryptophan and the two unknown nucleotides elucidated in this work.

and the corresponding unknown nucleobases: 825.2-Da dinucleotide, $C_{30}H_{45}N_5O_{16}P_2S$; 307.2-Da nucleobase, $C_{16}H_{25}N_3OS$; 869.2-Da dinucleotide, $C_{31}H_{45}N_5O_{18}P_2S$; 351.2-Da nucleobase, $C_{17}H_{25}N_3O_3S$ (**Supplementary Fig. 2**). These empirical formulas are consistent with 5′-X-U-3′ structures. Given the high degree of hydrophobicity of the parental species, we reasoned that the unknown nucleobases probably contain a lipid-like group. The empirical formulas also suggested that the two nucleobases are likely to differ by the presence or absence of a carboxylic acid.

To further characterize the structures of the unknown nucleobases, we performed MS/MS experiments in positive and negative ion mode on both unlabeled and isotopically labeled samples (**Fig. 2b**,**c** and **Supplementary Fig. 3**). On the basis of the isotope labeling data, a key fragment, $[M+H]^+$ m/z = 137.14, had a predicted neutral molecular formula of $C_{10}H_{16}$, corresponding to three degrees of unsaturation (**Fig. 2d**). We hypothesized that this ion might arise from the elimination of a geranyl group attached to the newly identified nucleobases (**Supplementary Fig. 4a**). The presence of a characteristic ion fragment with $[M+H]^+$ m/z = 81.07 in the MS/MS of both nucleobases that we also observed in the MS/MS spectra of geraniol, geranyl acetate, farnesol and geranylgeraniol, four potentially related lipid-like molecules analyzed for comparison, supports this hypothesis (**Supplementary Fig. 4b**).

The removal of a geranyl group ($C_{10}H_{16}$) from the empirical formulas of the unknown nucleobases resulted in molecular formulas matching two known modified nucleobases, 5-methylaminomethyl-2-thiouridine (mnm5s2U) and 5-carboxymethylaminomethyl-2-thiouridine (cmnm5s2U). We therefore speculated that the unknown nucleobases could be geranylated mnm5s2U and geranylated cmnm5s2U. These structures are consistent with all ion species arising from the fragmentation of the nucleobases (**Supplementary Table 2**). Notably, the $[M+H]^+$ m/z = 277.139 fragment observed in the positive mode MS/MS of both nucleobases, a fragment consistent with the presence of the geranyl group and the absence of the (carboxy)methylamino group, strongly suggests that the proposed geranyl group is not linked to the exocyclic amine of either new nucleobase (**Fig. 2d**).

To confirm that the unknown nucleotides were derivatives of mnm5s2U and cmnm5s2U and to verify the regiochemistry of the geranyl group, we prepared RNA from *E. coli* strains lacking genes involved in the biosynthesis of mnm5s2U and cmnm5s2U. *E. coli* Δ*mnmA* or Δ*mnmE* lack the ability to install sulfur at the C2 of uridine or to incorporate the methylaminomethyl group at the C5 position, respectively[14,15]. Neither sample contained detectable quantities of the previously observed unknown nucleotides (**Supplementary Fig. 5**), suggesting that the geranyl group is bound to either the sulfur atom or to the methylaminomethyl group. Whereas RNA isolated from *E. coli* Δ*mnmA* did not contain any corresponding geranylated product (for example, geranyl-mnm5U-U), RNA from *E. coli* Δ*mnmE* strains did contain a geranyl-2-thiouridine–containing dinucleotide (ges2U-U). These results indicate that either the 2-thiouridine position is the position of geranylation (**Fig. 3a**), or it is necessary for recognition by enzyme (or enzymes) responsible for installing the geranyl group.

To test the hypothesis that geranylation occurs at the 2-thiouridine position, we developed a synthetic route to geranyl-2-thiouridine (1-((2*R*,3*R*,4*S*,5*R*)-3,4-dihydroxy-5-(hydroxymethyl) tetrahydrofuran-2-yl)-2-(((E)-3,7-dimethylocta-2,6-dien-1-yl) thio)pyrimidin-4(1H)-one; ges2U; **1**) from 2-thiouridine (**Supplementary Fig. 6**). Because it is challenging to discriminate the desired S-alkylation product from the undesired N-substituted and O-substituted adducts, we used two-dimensional NMR of synthetic ges2U as well as X-ray crystallography of the nucleobase (geranyl-2-thiouracil; **Supplementary Data Set 1**) derived from synthetic ges2U to unambiguously confirm the structure of the synthetic ges2U (**Supplementary Figs. 15–22**). To generate a biological sample for comparison, we subjected RNA from *E. coli* Δ*mnmE* to P1 digestion, base hydrolysis and dephosphorylation. The resulting biological sample and the synthetic ges2U showed identical

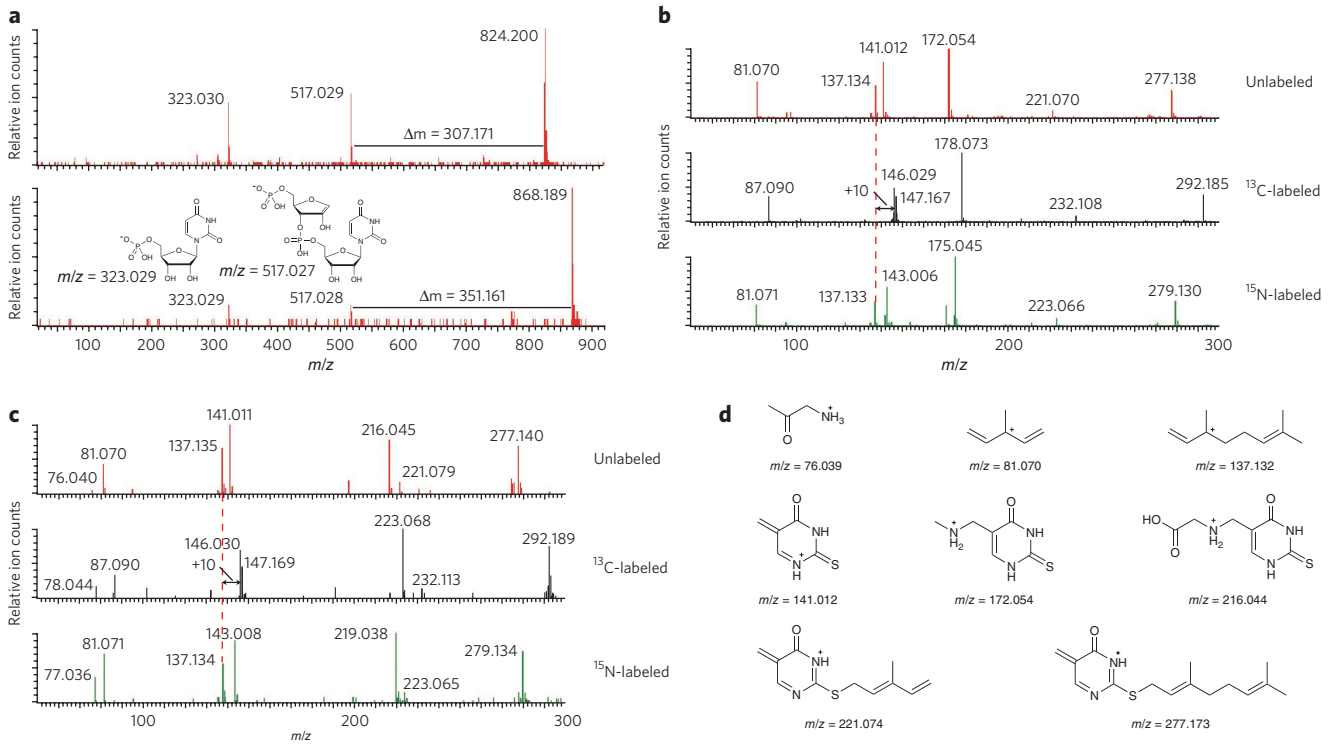

Figure 2 | MS characterization of two new hydrophobic small molecule–RNA conjugates. (**a**) Negative ion–mode MS/MS of the two unknown nucleotides reveals a dinucleotide structure with uracil and two unknown nucleobases of 307.171 Da and 351.161 Da. (**b**,**c**) Positive ion–mode MS/MS of the unknown nucleobases of 307.171 Da (**b**) and 351.161 Da (**c**) from unlabeled as well as ^{13}C-labeled and ^{15}N-labeled RNA. The spectra of the unlabeled RNA samples are shown in red, the ^{13}C-labeled samples are in black, and the ^{15}N-labeled samples are in green. The +10-Da shifts, indicating 10 carbon atoms, of the geranyl fragment in the ^{13}C-labeled RNA spectra are shown. (**d**) Proposed structures for the individual ion fragments observed in the MS/MS experiments.

retention times and eluted as a single peak upon co-injection and HPLC (**Fig. 3b**). In addition, side-by-side MS/MS comparison at varying collision energies[16] revealed an identical ion fragmentation pattern for both molecules (**Fig. 3c**). Collectively, these results unambiguously indicate that these new *E. coli* nucleotides contain a geranyl group that is linked to the 2-thiouridine group, and they confirm that our proposed structures for cellular mnm5ges2U (**2**) and cmnm5ges2U (**3**) are correct (**Fig. 3a**). We further used the synthetic standard of ges2U to estimate the amount of cellular ges2U in *E. coli* Δ*mnmE* cells by generating a standard curve with known quantities of authentic ges2U. Ion counts of ges2U in digested *E. coli* Δ*mnmE* RNA correspond to 398 ± 81 molecules of ges2U per *E. coli* cell under the growth conditions used (**Supplementary Fig. 7**).

Geranylated nucleotides are in anticodon loops of tRNAs

Next we sought to identify the RNA sequence (or sequences) that contained the geranylated nucleotides. We fractionated total RNA by size and analyzed individual fractions by P1 digestion followed by LC/MS. Both modifications were present on RNAs of length 50–80 nucleotides (**Supplementary Fig. 8**), suggesting their possible attachment to tRNA. Considering that mnm5s2U is known to be present at the first uridine (U34) in the anticodon loop of the $tRNA^{Glu}_{UUC}$ and $tRNA^{Lys}_{UUU}$ and that $tRNA^{Gln}_{UUG}$ contains a potentially related s2U derivative at the same position[17–19], we speculated that the geranylated nucleotides might also be present at U34 of these tRNAs. Because these three tRNAs all contain a uridine at position 35 as well, this hypothesis is consistent with our observation of geranylated bases in 5′-X-U-3′ dinucleotides.

To test this hypothesis, we used complementary oligonucleotides specific for $tRNA^{Glu}_{UUC}$, $tRNA^{Lys}_{UUU}$ and $tRNA^{Gln}_{UUG}$ to isolate these tRNAs from whole *E. coli* tRNA. In addition, we isolated $tRNA^{Asn}_{GUU}$, which does not contain any 2-thiouridine derivatives and therefore should not contain any geranylated nucleotides, as a negative control[20]. P1 nuclease digestion and LC/MS analysis revealed that geranylated mnm5s2U is primarily present in $tRNA^{Glu}_{UUC}$ and $tRNA^{Lys}_{UUU}$ and that geranylated cmnm5s2U is present on $tRNA^{Gln}_{UUG}$ (**Fig. 4a**). We observed that $tRNA^{Gln}_{UUG}$ also contains mnm5ges2U to a lesser extent, suggesting that this tRNA contains either mnm5s2U or cmnm5s2U at U34. As expected, we observed no geranylated nucleotides in the isolated $tRNA^{Asn}_{GUU}$.

To probe the position of the geranylated nucleotide in these tRNAs, we digested these isolated tRNAs with RNase T1, which specifically cleaves after G. We observed fragments corresponding to both (c)mnm5s2U and (c)mnm5ges2U at position U33 or U34 (**Supplementary Fig. 9**). Samples digested with RNase A, which cleaves specifically after pyrimidines[21], contained the dinucleotides mnm5ges2U-U and cmnm5ges2U-U. These results are consistent with geranylation of the anticodon loop because uridine is immediately 5′ of the anticodon in the tRNAs in question. MS/MS analysis of RNase T1 digestion products provided a partial sequence[22] of the anticodon loop and definitively confirmed U34 as the site of geranylation in $tRNA^{Glu}_{UUC}$, $tRNA^{Lys}_{UUU}$ and $tRNA^{Gln}_{UUG}$ (**Fig. 4b** and **Supplementary Fig. 10**). Collectively, these findings suggest that geranylated (c)mnm5s2U coexists with (c)mnm5s2U at position U34 in $tRNA^{Glu}_{UUC}$, $tRNA^{Lys}_{UUU}$ and $tRNA^{Gln}_{UUG}$. On the basis of previously reported abundances of tRNAs in *E. coli* cells[23,24], we estimate that approximately 2.8–6.7% of these three tRNAs are geranylated.

SelU geranylates mnm5s2U and cmnm5s2U

To characterize the evolutionary conservation of these geranylated nucleotides, we analyzed whole-cellular RNA or tRNA from 11

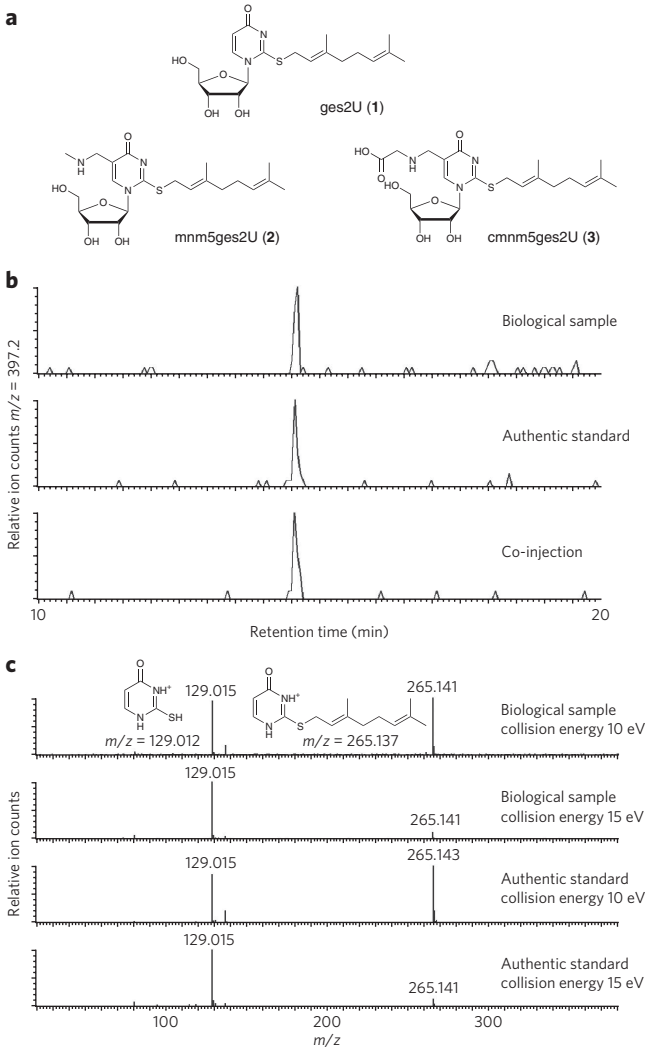

Figure 3 | Structural elucidation of two geranylated nucleosides. (**a**) Structures of geranylated 2-thiouridine (ges2U), geranylated 5-methylaminomethyl-2-thiouridine (mnm5ges2U) and geranylated 5-carboxymethylaminomethyl-2-thiouridine (cmnm5ges2U). (**b**) LC comparison of biologically generated and authentic synthetic ges2U. (**c**) Comparison of biologically generated and authentic ges2U by MS/MS fragmentation. Proposed structures of ion fragments are shown.

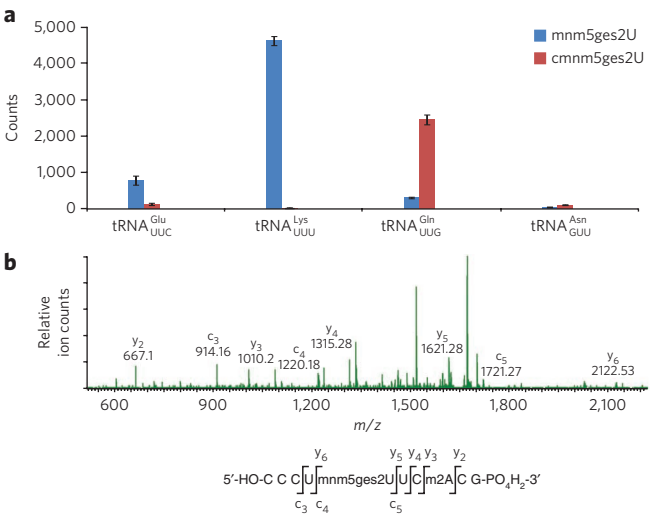

Figure 4 | Characterization of geranylated cellular RNAs. (**a**) MS analysis of individual tRNAs isolated from cells shows that mnm5ges2U is present on $tRNA^{Glu}_{UUC}$, $tRNA^{Lys}_{UUU}$ and (to a lesser extent) $tRNA^{Gln}_{UUG}$ and that cmnm5ges2U is present on $tRNA^{Gln}_{UUG}$. The same analysis is shown from $tRNA^{Asn}_{GUU}$ as a negative control. (**b**) MS/MS fragmentation of the T1 digestion product of $tRNA^{Glu}_{UUC}$ shows that the geranylated nucleotide is U34.

different organisms. Both geranylated nucleotides were present in *E. aerogenes*, *P. aeruginosa* and *S.* Typhimurium but were undetectable in *Vibrio fischeri*, *Bacillus subtilis* and several samples of eukaryotic RNA (**Supplementary Table 3**). These results are consistent with a previous report describing the presence of an uncharacterized $C_{10}H_{17}$ group, noted by the authors to have the same molecular formula as a geranyl group, on mnm5s2U in *S.* Typhimurium whose mutated *sufY* gene resulted in a G67E mutant[17]. SelU, the *E. coli* homolog of SufY, is known to catalyze the substitution of the sulfur at the C2 position of (c)mnm5s2U with selenium when selenium is present at high concentrations[25]. Notably, the presence or absence of homologs of *sufY* in each of the 11 organisms tested perfectly predicts the observed presence or absence of geranylated RNA (**Supplementary Table 3**), suggesting a potential role of this gene in RNA geranylation.

To test the role of *selU* in producing geranylated RNA, we isolated RNA from a *selU*-deficient *E. coli* strain (*E. coli* Δ*selU*) as well as from *E. coli* Δ*selU* bearing a plasmid with a complementary copy of either wild-type *selU* or mutant selU encoding the G67E mutation (*selU*(G67E)) . In *E. coli* Δ*selU* we observed neither mnm5ges2U nor cmnm5ges2U. *E. coli* Δ*selU* complemented with wild-type *selU* contained levels of geranylated RNA similar to those found in *E. coli*, whereas complementation with *selU*(G67E) resulted in increased amounts of mnm5ges2U and cmnm5ges2U (**Fig. 5a**). Collectively, these results confirm that *selU* is involved in the production of geranylated RNA. We obtained similar results when complementing *E. coli* Δ*selU* with *S.* Typhimurium *sufY*(G67E), *S.* Typhimurium *sufY*(G67R) (a second mutant with a similar phenotype to that of *sufY*(G67E)[17]) and *E. coli selU*(G67R) (**Supplementary Fig. 11**). As *selU* is part of the biosynthetic pathway of 5-methylaminomethyl-2-selenouridine (mnm5se2U) and 5-carboxymethylaminomethyl-2-selenouridine (cmnm5se2U), we analyzed RNA from *E. coli* grown under varying selenium concentrations[25]. Notably, selenation occurs at the expense of geranylation at selenium concentrations above 10 nM (**Supplementary Fig. 12**).

To determine whether SelU directly catalyzes geranylation or has a regulatory role in the reaction, we expressed and purified *E. coli* SelU. Upon incubation of *E. coli* tRNA in the presence or absence of purified SelU and geranyl pyrophosphate, we observed a six- to seven-fold increase in the amounts of mnmges2U and cmnm5ges2U in samples containing SelU and geranyl pyrophosphate (**Fig. 5b**), thus confirming that SelU geranylates tRNA containing (c)mnm5s2U using geranyl pyrophosphate as a cosubstrate.

Geranylation affects codon bias and frameshifting

We hypothesized that a large hydrophobic modification at the first position of an anticodon could affect recognition of the third position in the corresponding codon during translation. $tRNA^{Glu}_{UUC}$, $tRNA^{Lys}_{UUU}$ and $tRNA^{Gln}_{UUG}$ recognize the codons GAA and GAG, AAA and AAG, and CAA and CAG, respectively. Although CAG is further decoded by a separate $tRNA^{Gln}_{CUG}$, only $tRNA^{Glu}_{UUC}$ and $tRNA^{Lys}_{UUU}$ decode the codons GAA and GAG and the codons CAA and CAG, respectively. Therefore, we compared the translation efficiency of the glutamate codons GAA and GAG at varying degrees of geranylation using a luciferase reporter containing a six-codon leader sequence of either $(GAA)_6$ or $(GAG)_6$ after the ATG start codon. Consistent with previous reports, we observed that *E. coli*

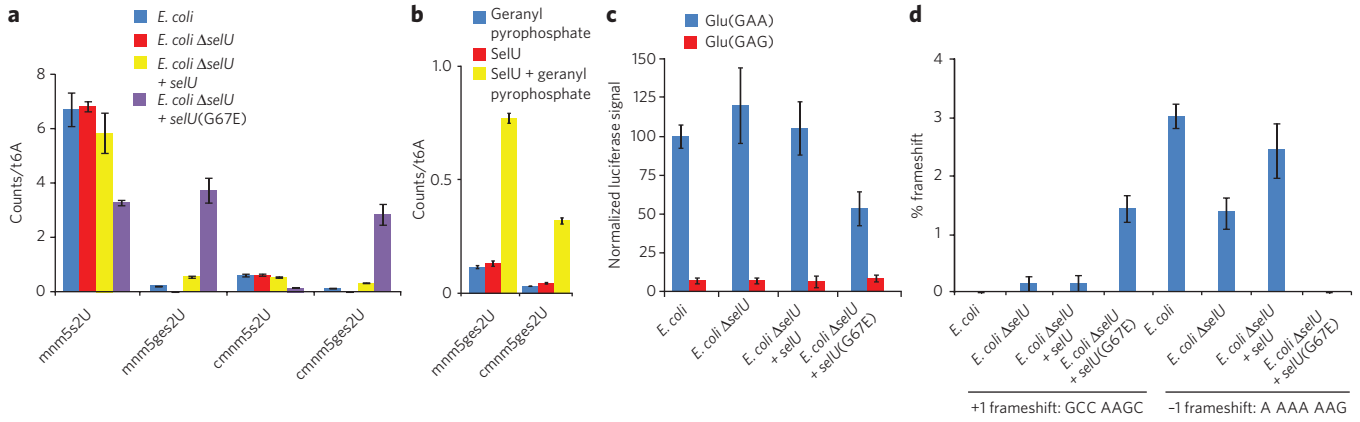

Figure 5 | Biological abundance and properties of geranylated RNA. (**a**) *E. coli selU* mutations affect the amount of geranylated nucleotide. No geranylated nucleotides were detected in *E. coli* cells lacking *selU* (Δ*selU*). Complementation with wild-type *selU* restores geranylation. Complementation with mutant *selU*(G67E) results in high geranylation. Amount of nucleoside is shown as relative ratios of MS counts of the individual nucleosides versus t6A. Error bars represent the s.d. of three independent biological replicates. (**b**) SelU geranylates mnm5s2U and cmnm5s2U on tRNA *in vitro* in a geranyl pyrophosphate–dependent manner. Amount of nucleoside is shown as relative ratios of MS counts of the individual nucleosides versus t6A. Error bars represent the s.d. of three analytical replicates. (**c**) Luciferase reporter assay of glutamate codon translation efficiency for GAA and GAG reveals a strong bias favoring GAA translation under low geranylation conditions and a substantially lower preference under high geranylation conditions owing to reduced efficiency of GAA decoding. Error bars represent the s.d. of three independent biological replicates. (**d**) Frameshift efficiency (+1 and −1) at the sequences GCC AAGC and A AAA AAG inserted between GST and MBP reading frames. The *y* axis indicates the amount of GST-MBP fusion protein detected by western blotting as a percentage of total GST-containing proteins produced. Error bars reflect the range of two technical replicates.

$tRNA^{Glu}_{UUC}$ decodes GAA ten times more efficiently than GAG[26–28]. Although this ratio did not change substantially for *E. coli* Δ*selU* and *E. coli* Δ*selU* complemented with wild-type *selU*, complementation with *selU*(G67E) strongly reduced the translation efficiency of GAA codons, altering the relative GAA/GAG translation efficiency ratio from 13:1 to 6:1 (**Fig. 5c**). We verified that these differences do not arise from changes in transcript levels, which did not differ substantially among all of the strains assayed (**Supplementary Fig. 13**). Collectively, these findings indicate that tRNA geranylation affects translational efficiency in a codon-dependent manner.

Next, we investigated whether the amount of tRNA geranylation can affect the efficiency of frameshifting during translation. In *S.* Typhimurium, the *sufY*(G67E) mutant resulting in the uncharacterized $C_{10}H_{17}$ fragment has been previously reported to enable the read-through of a single C insertion in the *his* operon[17]. We studied the effect of varying the amount of tRNA geranylation on frameshift of the nucleotide sequence GCC AAGC, which promotes +1 frameshifting, and A AAA AAG, which promotes −1 frameshifting, at the lysine codon AAG[29,30]. The sequences were inserted between the genes encoding glutathione *S*-transferase (GST) and maltose-binding protein (MBP). The inserted sequence was followed by a stop codon in frame with the GST-encoding sequence, and the MBP-encoding sequence was in the +1 or −1 frame, such that production of MBP requires a +1 or −1 translational frameshift. High amounts of geranylation in *E. coli* Δ*selU* complemented with *selU*(G67E) induced the +1 frameshift (GCC AAGC) by more than ten-fold and reduced the −1 frameshift (A AAA AAG) to less than a tenth relative to that of wild-type *E. coli*. For the +1 frameshift, we observed no detectable change in frameshift efficiency relative to wild-type *E. coli* for *E. coli* Δ*selU* or *E. coli* Δ*selU* complemented with *selU*, whereas for the −1 frameshift we observed a decrease in frameshift efficiency to one-half for *E. coli* Δ*selU* and no substantial change for *E. coli* Δ*selU* complemented with *selU* (**Fig. 5d**). It has been previously reported that +1 frameshifting at GCC AAGC can be induced by uncharged lysyl-tRNAs as a result of ribosome stalling at lysine codons in lysine-limiting conditions[30]. However, the amount of lysyl-tRNA or $tRNA^{Lys}_{UUU}$ does not change as a function of geranylation (**Supplementary Fig. 14**), suggesting that geranylation either results in ribosome stalling by affecting a different step of protein synthesis or induces frameshifting by a completely different mechanism.

DISCUSSION

Post-transcriptional modifications are ubiquitous on all tRNA molecules and are known to affect their function[31]. Among the most frequently modified nucleotides in tRNAs are the anticodon wobble position (nucleotide 34) and the purine adjacent to the anticodon (nucleotide 37). These modifications ensure binding to the correct codon and are required for the efficient recognition of several codons by the same tRNA[32]. Modifications at U34 are of particular importance for wobble base pairing. In bacteria, the 5-methylaminomethyl group on $tRNA^{Lys}_{UUU}$ and $tRNA^{Glu}_{UUC}$ is important for efficient recognition of AAG and GAG, respectively[33,34]. The 2-thiouridine group in modified U34 of $tRNA^{Glu}_{UUC}$ has been reported to increase recognition of GAA[33] and to be required for efficient binding to the tRNA synthetase[35]. Moreover, in $tRNA^{Lys}_{UUU}$ the 2-thiouridine modification enables binding to the ribosome[36,37].

Previously reported RNA modifications are predominantly small or hydrophilic[5]. Our identification of two lipid-linked nucleotides, mnm5ges2U and cmnm5ges2U, represent what is to our knowledge the first characterized oligoisoprenylated nucleotides. These modifications are present in *E. coli*, *E. aerogenes*, *P. aeruginosa* and *S.* Typhimurium at U34 in the anticodon of $tRNA^{Glu}_{UUC}$, $tRNA^{Lys}_{UUU}$ and $tRNA^{Gln}_{UUG}$. It is tempting to speculate that the unusual physical properties of very hydrophobic RNA modifications may have obscured their discovery using previous methods. The conditions traditionally used for TLC and HPLC optimized for hydrophilic mononucleotides often do not allow the resolution of very hydrophobic molecules[38]. Furthermore, although MS has previously been used to characterize new nucleotides, in this work, the use of MS in the primary screening method enabled the detection and analysis of low-abundance species.

Our findings suggest that RNA geranylation is an alternative to selenation at low selenium concentrations. Notably, our results indicate that *selU* can directly catalyze geranylation of (c)mnm5s2U, in addition to its previously known role in selenation. That two such disparate chemical alternatives are mediated by a single enzyme (SelU in *E. coli*) may seem surprising, but we speculate that the active

site can accommodate either selenophosphate, the substrate for 2-selenouridine synthesis[25], or geranyl pyrophosphate, the substrate for geranylation. Indeed, in *S.* Typhimurium, mutation of Cys97, the homolog of the active site residue for selenation in selU, to alanine in the rhodanese domain of sufY^{G67E} strongly reduces the amount of the previously uncharacterized $C_{10}H_{17}$ modification[17]. Even if the same residues on the protein mediate both selenation and geranylation, the two reactions are likely to proceed through distinct pathways because, for selenation, the sulfur at the 2-thiouridine position must be converted into a leaving group as selenium acts as nucleophile, whereas for geranylation, the 2-thiouridine group acts as nucleophile. Further studies on this enzyme are necessary to reveal the detailed mechanism of RNA geranylation.

High geranylation can affect codon bias and frameshift efficiency during translation. The exact mechanisms by which geranylation modulates codon bias and frameshifting remains to be elucidated. Frameshifting at A AAA AAG is thought to be based on the preferential recognition of AAA over AAG by the anticodon mnm5s2UUU[39]. Hypomodification of tRNA$^{Lys}_{UUU}$ at U34 reduces −1 frameshifting at NNAAAG sites, possibly owing to improved wobble pairing[40]. Likewise, high geranylation may also reduce the preference of tRNA$^{Lys}_{UUU}$ for AAA over AAG, thereby inhibiting frameshifting. This hypothesis is supported by our observation of decreased codon bias of tRNA$^{Glu}_{UUG}$ upon geranylation. Considering the importance of the position of U34 in codon recognition as well as tRNA synthetase and ribosome binding[35–37], it is possible that a large hydrophobic modification at this tRNA position could have a broad impact on translation.

Recent studies in *Saccharomyces cerevisiae* suggest that the large-scale coordination of tRNA modification is a translational response to cellular stress[41,42]. Modification of RNA with unusually hydrophobic groups could also affect subcellular localization by inducing association with proteins or the lipid bilayer. Small molecule–RNA conjugates may therefore represent an additional basis for translation-independent RNA localization[43,44]. Geranylation and other hydrophobic modifications of RNA may also have even broader roles in cellular function. It was recently reported that lipid-linked DNA oligonucleotides enable vesicle fusion *in vitro* by a process resembling SNARE-mediated exocytosis[45]. Although lipid–oligonucleotide conjugates that induce membrane association in living systems have yet to be discovered, analogous processes enabled by lipid-linked cellular RNA may exist in nature.

METHODS

RNA digestion with nuclease P1. For negative ion–mode MS analysis, 500 µg of isolated total RNA or commercial RNA (wheat tRNA or bovine total RNA; Sigma) were digested with 10 U nuclease P1 (Wako Chemicals) in 500 µl of 50 mM NH_4OAc, pH 4.5, at 37 °C for 40 min. The control samples were prepared by incubating 500 µg of RNA with 10 U heat-inactivated nuclease P1 (preincubated at 95 °C for 2 h) in 500 µl of 50 mM NH_4OAc, pH 4.5, at 37 °C for 40 min. The digestion products were purified by size-exclusion chromatography (NAP5, GE Healthcare), and the small-molecule fraction was collected and lyophilized. The size-exclusion chromatography step during the screen was performed in the presence of 30–50% (v/v) methanol.

RNA digestion with nuclease P1 and alkaline phosphatase. For positive ion–mode MS analysis, 500 µg of isolated total RNA or commercial RNA were digested with 10 U nuclease P1 and 100 U alkaline phosphatase (Sigma) in 500 µl of 50 mM NH_4OAc, pH 6, at 37 °C for 4 h. The digestion products were purified and lyophilized as described above.

LC/MS analysis. LC/MS was performed using a Waters Aquity ultra-performance LC (UPLC) quadrupole TOF Premier instrument. For negative ion–mode analysis of the small molecule–RNA conjugate screen and characterization of nuclease P1–digested RNA, samples were dissolved in 50 µl of 0.05% NH_3HCOOH in 1:1 deionized water/DMSO, and LC was performed using a gradient from 0.1% (w/v) aqueous ammonium formate (A1) to methanol (B1) on an Aquity UPLC BEH C8 column (1.7 µm, 2.1 mm × 50 mm, Waters) at a constant flow rate of 0.3 ml min^{-1}. The mobile phase composition was as follows: 100% A1 for 1 min; linear increase over 21 min to 100% B1; maintain at 100% B1 for 8 min; return to 100% A1 over 5 min. The column was washed several times using short gradients from A1 to B1 between subsequent LC runs. Electrospray ionization was used with a capillary voltage of 3 kV, a sampling cone voltage of 25 V and 40 V and a low-mass resolution of 4.7. The desolvation gas temperature was 300 °C, the flow rate was 800 l h^{-1}, the source temperature was 150 °C, and the detector was operated in negative-ion mode. LC/MS/MS experiments of the identified dinucleotides were performed at collision energies of 20 eV and 30 eV. LC/MS/MS/MS for fragmentation of the nucleobase of mnm5ges2U was performed with the cone voltage at 180 V and collision energies of 10 eV and 20 eV.

Positive ion–mode analysis for the characterization of RNA digested with nuclease P1 and alkaline phosphatase (Sigma) was performed by dissolving the samples in 50 µl of 0.1% (v/v) aqueous NH_3HCOOH and performing LC using a linear gradient from 0.1% (v/v) aqueous formic acid (A2) to acetonitrile (B2) on an Aquity UPLC BEH C8 column (1.7 µm, 2.1 mm × 50 mm) at a constant flow rate of 0.3 ml min^{-1}. The mobile-phase composition was as follows: 100% A2 for 1 min; linear increase over 21 min to 100% B2; maintain at 100% B2 for 8 min; return to 100% A2 over 5 min. Electrospray-ionization conditions were as described above with the detector operating in positive ion mode. LC/MS/MS experiments of the identified dinucleotides were performed at collision energies of 10 eV, 20 eV and 30 eV. LC/MS/MS/MS analysis of the nucleobase of mnm5ges2U was performed with a cone voltage of 70 V and collision energies of 10 eV, 20 eV, 30 eV and 40 eV. LC/MS/MS/MS analysis of the nucleobase of cmnm5ges2U was performed with a cone voltage of 40 V and collision energies of 10 eV, 20 eV, 30 eV and 40 eV.

Synthesis of ges2U. A solution of 2-thiouridine (26.0 mg, 0.1 mmol, Santa Cruz Biotechnology), geranyl bromide (57 µl, 0.3 mmol) and *N*,*N*-diisopropylethylamine (52.2 µl, 0.3 mmol) in methanol (1 ml) was stirred at 25 °C for 12 h. The resulting reaction mixture was concentrated *in vacuo* and directly subjected to silica gel chromatography. A white solid (38.9 mg, 98% yield) was obtained after flash chromatography: TLC R_f = 0.10 (10% (v/v) MeOH/EtOAc); ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.25 (d, *J* = 7.2 Hz, 1 H), 5.98 (d, *J* = 7.2 Hz, 1 H), 5.91 (d, *J* = 5.4 Hz, 1 H), 5.28 (t, *J* = 7.2 Hz, 1 H), 5.05 (t, *J* = 7.2 Hz, 1 H), 4.45 (m, 1 H), 4.37 (m, 1 H), 4.18 (s, 1 H), 3.88–3.96 (m, 2 H), 3.78–3.86 (m, 2 H), 2.04–2.12 (m, 2 H), 1.98–2.04 (m, 2 H), 1.72 (s, 3 H), 1.68 (s, 3 H), 1.59 (s, 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 170.4, 164.9, 143.8, 141.2, 132.1, 123.9, 115.9, 109.0, 92.2, 86.1, 75.7, 70.8, 61.5, 39.9, 31.7, 26.6, 25.9, 18.0, 17.0; HRMS-EI (*m*/*z*): $[M+H]^+$ calculated for $C_{19}H_{28}N_2O_5S$ = 397.1792, observed = 397.1800.

Accession codes. Cambridge Structural Database: crystallography data for geranyl-2-thiouracil have been deposited under accession code CCDC892714.

Received 23 April 2012; accepted 24 July 2012; published online 16 September 2012

References

1. He, L. & Hannon, G.J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **5**, 522–531 (2004).
2. Serganov, A. & Patel, D.J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.* **8**, 776–790 (2007).
3. Fedor, M.J. & Williamson, J.R. The catalytic diversity of RNAs. *Nat. Rev. Mol. Cell Biol.* **6**, 399–412 (2005).
4. Ding, S.W. RNA-based antiviral immunity. *Nat. Rev. Immunol.* **10**, 632–644 (2010).
5. Cantara, W.A. *et al.* The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res.* **39**, D195–D201 (2011).
6. Ikeuchi, Y. *et al.* Agmatine-conjugated cytidine in a tRNA anticodon is essential for AUA decoding in archaea. *Nat. Chem. Biol.* **6**, 277–282 (2010).
7. Mandal, D. *et al.* Agmatidine, a modified cytidine in the anticodon of archaeal tRNA(Ile), base pairs with adenosine but not with guanosine. *Proc. Natl. Acad. Sci. USA* **107**, 2872–2877 (2010).
8. Ikeuchi, Y., Shigi, N., Kato, J., Nishimura, A. & Suzuki, T. Mechanistic insights into sulfur relay by multiple sulfur mediators involved in thiouridine biosynthesis at tRNA wobble positions. *Mol. Cell* **21**, 97–108 (2006).
9. Miles, Z.D., McCarty, R.M., Molnar, G. & Bandarian, V. Discovery of epoxyqueuosine (oQ) reductase reveals parallels between halorespiration and tRNA modification. *Proc. Natl. Acad. Sci. USA* **108**, 7368–7372 (2011).
10. Noma, A., Kirino, Y., Ikeuchi, Y. & Suzuki, T. Biosynthesis of wybutosine, a hyper-modified nucleoside in eukaryotic phenylalanine tRNA. *EMBO J.* **25**, 2142–2154 (2006).
11. Kowtoniuk, W.E., Shen, Y., Heemstra, J.M., Agarwal, I. & Liu, D.R. A Chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA. *Proc. Natl. Acad. Sci. USA* **106**, 7768–7773 (2009).
12. Chen, Y.G., Kowtoniuk, W.E., Agarwal, I., Shen, Y. & Liu, D.R. LC/MS analysis of cellular RNA reveals NAD-linked RNA. *Nat. Chem. Biol.* **5**, 879–881 (2009).
13. Scott, A.I. How were porphyrins and lipids synthesized in the RNA world? *Tetrahedr. Lett.* **38**, 4961–4964 (1997).

14. Hagervall, T.G., Edmonds, C.G., McCloskey, J.A. & Bjork, G.R. Transfer RNA(5-methylaminomethyl-2-thiouridine)-methyltransferase from *Escherichia coli* K-12 has two enzymatic activities. *J. Biol. Chem.* **262**, 8488–8495 (1987).
15. Kambampati, R. & Lauhon, C.T. MnmA and IscS are required for *in vitro* 2-thiouridine biosynthesis in *Escherichia coli*. *Biochemistry* **42**, 1109–1117 (2003).
16. Nicolas, E.C. & Scholz, T.H. Active drug substance impurity profiling part II. LC/MS/MS fingerprinting. *J. Pharm. Biomed. Anal.* **16**, 825–836 (1998).
17. Chen, P., Crain, P.F., Nasvall, S.J., Pomerantz, S.C. & Bjork, G.R.A. "Gain of function" mutation in a protein mediates production of novel modified nucleosides. *EMBO J.* **24**, 1842–1851 (2005).
18. Jühling, F. *et al.* tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.* **37**, D159–D162 (2009).
19. Yaniv, M. & Folk, W.R. The nucleotide sequences of the two glutamine transfer ribonucleic acids from *Escherichia coli*. *J. Biol. Chem.* **250**, 3243–3253 (1975).
20. Yokogawa, T., Kitamura, Y., Nakamura, D., Ohno, S. & Nishikawa, K. Optimization of the hybridization-based method for purification of thermostable tRNAs in the presence of tetraalkylammonium salts. *Nucleic Acids Res.* **38**, e89 (2010).
21. Volkin, E. & Cohn, W.E. On the structure of ribonucleic acids. 2. The products of ribonuclease action. *J. Biol. Chem.* **205**, 767–782 (1953).
22. Mcluckey, S.A., Vanberkel, G.J. & Glish, G.L. Tandem mass spectrometry of small, multiply charged oligonucleotides. *J. Am. Soc. Mass Spectrom.* **3**, 60–70 (1992).
23. Dong, H., Nilsson, L. & Kurland, C.G. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.* **260**, 649–663 (1996).
24. Jakubowski, H. & Goldman, E. Quantities of individual aminoacyl-tRNA families and their turnover in *Escherichia coli*. *J. Bacteriol.* **158**, 769–776 (1984).
25. Wolfe, M.D. *et al.* Functional diversity of the rhodanese homology domain: the *Escherichia coli ybbB* gene encodes a selenophosphate-dependent tRNA 2-selenouridine synthase. *J. Biol. Chem.* **279**, 1801–1809 (2004).
26. Looman, A.C. *et al.* Influence of the codon following the AUG initiation codon on the expression of a modified *lacZ* gene in *Escherichia coli*. *EMBO J.* **6**, 2489–2492 (1987).
27. Sørensen, M.A. & Pedersen, S. Absolute *in vivo* translation rates of individual codons in *Escherichia coli*. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. *J. Mol. Biol.* **222**, 265–280 (1991).
28. Wittwer, A.J. & Ching, W.M. Selenium-containing tRNA(Glu) and tRNA(Lys) from *Escherichia coli*: purification, codon specificity and translational activity. *Biofactors* **2**, 27–34 (1989).
29. Gurvich, O.L. *et al.* Sequences that direct significant levels of frameshifting are frequent in coding regions of *Escherichia coli*. *EMBO J.* **22**, 5941–5950 (2003).
30. Lindsley, D. & Gallant, J. On the directional specificity of ribosome frameshifting at a "hungry" codon. *Proc. Natl. Acad. Sci. USA* **90**, 5469–5473 (1993).
31. Suzuki, T. Biosynthesis and function of tRNA wobble modifications. in *Fine-Tuning of RNA Functions by Modification and Editing*, Vol. 12 (ed. Grosjean, H.) 23–69 (Springer, Berlin/Heidelberg, 2005).
32. Agris, P.F., Vendeix, F.A. & Graham, W.D. tRNA's wobble decoding of the genome: 40 years of modification. *J. Mol. Biol.* **366**, 1–13 (2007).
33. Krüger, M.K., Pedersen, S., Hagervall, T.G. & Sorensen, M.A. The modification of the wobble base of $tRNA^{Glu}$ modulates the translation rate of glutamic acid codons *in vivo*. *J. Mol. Biol.* **284**, 621–631 (1998).
34. Yarian, C. *et al.* Accurate translation of the genetic code depends on tRNA modified nucleosides. *J. Biol. Chem.* **277**, 16391–16395 (2002).
35. Sylvers, L.A., Rogers, K.C., Shimizu, M., Ohtsuka, E. & Soll, D. A 2-thiouridine derivative in $tRNA^{Glu}$ is a positive determinant for aminoacylation by *Escherichia coli* glutamyl-tRNA synthetase. *Biochemistry* **32**, 3836–3841 (1993).
36. Ashraf, S.S. *et al.* Single atom modification (O→S) of tRNA confers ribosome binding. *RNA* **5**, 188–194 (1999).
37. Yarian, C. *et al.* Modified nucleoside dependent Watson-Crick and wobble codon binding by $tRNA^{Lys}UUU$ species. *Biochemistry* **39**, 13390–13395 (2000).
38. Lane, B.G. Historical Perspectives on RNA Nucleoside Modifications. in *Modification and Editing of RNA* (eds. Grosjean, H. & Benne, R.) 1–20 (ASM Press, 1998).
39. Tsuchihashi, Z. & Brown, P.O. Sequence requirements for efficient translational frameshifting in the *Escherichia coli dnaX* gene and the role of an unstable interaction between tRNA(Lys) and an AAG lysine codon. *Genes Dev.* **6**, 511–519 (1992).
40. Licznar, P. *et al.* Programmed translational −1 frameshifting on hexanucleotide motifs and the wobble properties of tRNAs. *EMBO J.* **22**, 4770–4778 (2003).
41. Begley, U. *et al.* Trm9-catalyzed tRNA modifications link translation to the DNA damage response. *Mol. Cell* **28**, 860–870 (2007).
42. Chan, C.T. *et al.* A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS Genet.* **6**, e1001247 (2010).
43. Keiler, K.C. RNA localization in bacteria. *Curr. Opin. Microbiol.* **14**, 155–159 (2011).
44. Nevo-Dinur, K., Nussbaum-Shochat, A., Ben-Yehuda, S. & Amster-Choder, O. Translation-independent localization of mRNA in *E. coli*. *Science* **331**, 1081–1084 (2011).
45. Chan, Y.H., van Lengerich, B. & Boxer, S.G. Effects of linker sequences on vesicle fusion mediated by lipid-anchored DNA oligonucleotides. *Proc. Natl. Acad. Sci. USA* **106**, 979–984 (2009).

Acknowledgments

This work was supported by the Howard Hughes Medical Institute and the US National Institutes of Health (NIH)-National Institute of General Medical Sciences (NIGMS) (R01GM065865). C.E.D. acknowledges fellowship from the Novartis Foundation. A.M.L. was supported by a NIH National Research Service Award Postdoctoral Fellowship (F32GM095028). We thank S.-L. Zheng for his help with X-ray diffraction and structural determination of synthetic geranyl-2-thiouracil. We thank A. Saghatelian (Harvard University) for providing *P. aeruginosa* and I. Chen (Harvard University) for providing *S.* Typhimurium and BL2 facilities. We are also grateful to J. Carlson for his assistance and M. Ibba for helpful discussions.

Author contributions

C.E.D., Y.C., A.M.L. and Y.G.C. designed and performed the experiments. All authors analyzed the results and wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information and chemical compound information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to D.R.L.