Decoding the *Escherichia coli* epitranscriptome

Sebastian Riquelme Barrios[1]; Leonardo Vasquez Camus[1]; Siobhan A. Cusack[1]; Korinna Burdack[1], Dimitar Plamenov Petrov[1], G. Nur Yeşiltaç[2]; Stefanie Kaiser[2]; Pascal Giehr[3]*; Kirsten Jung[1]*

[1]Faculty of Biology, Microbiology, Ludwig-Maximilians-Universität München, 82152 Martinsried, Germany

[2]Goethe University Frankfurt, Faculty 14, Institute of Pharmaceutical Chemistry, 60438 Frankfurt

[3]Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 München, Germany

*To whom correspondence should be addressed:

Prof. Dr. Kirsten Jung

Tel: + 49 89/2180-74500 21 Fax: + 49 89/2180-74520

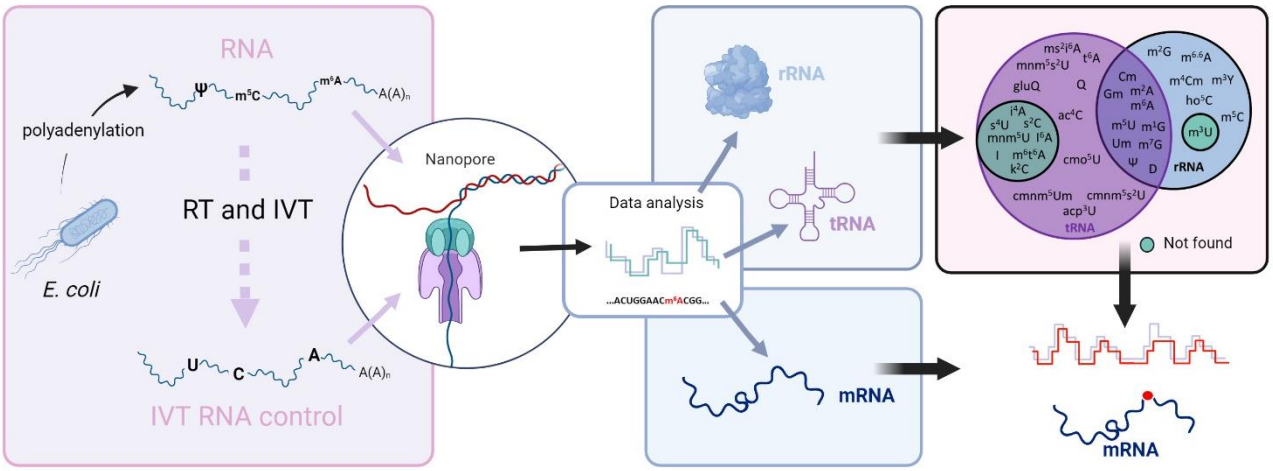
Email: jung@lmu.de

Dr. Pascal Giehr

Tel: +49 89/2180-73488

Email: p.giehr@lmu.de

GRAPHICAL ABSTRACT



ABSTRACT

Modifications of RNA, known as the epitranscriptome, affect mRNA stability, translation, and splicing in eukaryotes and have implications for developmental processes, cancer, and viral infections. In prokaryotes, however, the landscape of the epitranscriptome is still poorly understood. To address this knowledge gap, we used direct RNA sequencing with Nanopore technology to study RNA modifications in the model bacterium *Escherichia coli*. With a single sequencing reaction, we were able to simultaneously identify and map most of the known modification types in rRNA, tRNA, and mRNA. Subsequently, a multifaceted approach integrating different algorithms for data analysis, deletion mutants, mass spectrometry, qPCR, and *in vitro* methylation was implemented to evaluate the presence of m^5C and m^6A in *E. coli*. Known m^5C and m^6A sites in rRNA were confirmed, but these modifications could not be localized in the mRNA. Nevertheless, based on the sequencing data, modifications were found to be enriched in the coding regions of genes associated with general metabolism and RNA processing. This study provides a useful resource for experimental and bioinformatic approaches to gain new insights into post-transcriptional regulation in a prokaryotic model.

INTRODUCTION

RNA molecules are subject to chemical modifications similar to those found in proteins and DNA. In prokaryotes, modifications in the ribosomal RNA (rRNA) and transfer RNA (tRNA) have already been characterized in detail. For example, rRNA modifications are known to play a key role in the catalytic activity and assembly of ribosomes, and the absence of such modifications can alter translation accuracy [1-4]. tRNAs are also highly modified, particularly in the D- and TΨ-loops and in the anticodon region. These modifications affect the tRNA structure [5], mediate interactions with aminoacyl tRNA synthetases and elongation factor thermo unstable (EF-Tu) [6, 7], and influence the affinity of the anticodon region for specific codons [8]. Previous results from our laboratory have demonstrated the presence of numerous modification types in *E. coli* mRNA, which vary between growth stages and in response to stress conditions [9]. Despite these earlier findings, bacterial mRNA modifications remain to be explored in detail; their roles, abundance, and distribution are still largely unknown.

Over the past decade, several platforms have been developed with the capacity to directly sequence native RNA molecules, allowing new approaches to elucidate the epitranscriptome. Direct RNA

sequencing (DRS) using Oxford Nanopore Technologies (ONT) platforms not only provides information about nucleotide (nt) sequences, but also about nt modifications. Modified bases produce slightly different electrical signals than canonical (i.e., unmodified) bases when the RNA strand translocates through the nanopore. Multiple experimental strategies and bioinformatics tools have been developed to process the resulting data. Some tools compare DRS signals between pairs of samples (e.g., wild-type [WT] vs. gene knockout [KO] or WT vs. *in vitro* transcribed [IVT] RNA) to identify signals corresponding to modifications [10-13], whereas other approaches are based on the evaluation of individual sample types [14, 15]. Although protocols have been developed to evaluate DRS data from bacterial samples [16] and the signals produced by bacterial rRNA modifications have been analyzed [17], a comprehensive epitranscriptomic study in *E. coli* is still pending.

In *E. coli*, detailed studies have been conducted for the detection of pseudouridine (Ψ) and N^6-methyladenosine (m^6A) [18, 19]. A protocol called Pseudo-Seq that combines carbodiimide N-cyclohexyl-N'-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate (CMC) treatment, size selection for truncated reverse-transcribed cDNA, and high-throughput sequencing revealed that RluA, previously known as a tRNA- and 23S rRNA-specific Ψ-synthase, is responsible for modifications of at least 31 different mRNA positions [18]. Other rRNA Ψ-synthases, such as RluC and RluD, also modify some mRNA positions [18]. Notably, *fimA* transcripts (which encode the major subunit of the pilus) are less abundant in a Δ*rluA* mutant.

m^6A is the most prevalent mRNA modification in eukaryotes, where it plays roles in translation, degradation, splicing, nuclear export, and mRNA turnover [20]. In Gram-negative bacteria, such as *E. coli* and *Pseudomonas aeruginosa*, the ratio of m^6A to unmodified A bases has been calculated as ~0.2%, similar to the ratios in eukaryotic cells [19]. Furthermore, immunoprecipitation of m^6A followed by Illumina sequencing has revealed m^6A modifications in the mRNA of bacterial genes related to metabolism, transport, gene regulation, the cell envelope, ribosomes, stress responses, and nucleic acid synthesis [19]. Although m^6A consensus motifs have been proposed in *E. coli* and *P. aeruginosa* (UGCCAG and GGCCAG, respectively), no m^6A methyltransferase for mRNA is currently known [19].

Several *in vitro* studies have been conducted to investigate the role of m^6A in prokaryotic mRNA. It has been shown that m^6A in mRNA impedes tRNA interactions, likely due to destabilization of the m^6A–U pairing, which affects translation elongation *in vitro* [21]. Moreover, m^6A methylation has been associated with an increase in amino acid substitution errors and a decrease in reading accuracy at stop codons [22]. Finally, the presence of m^6A in mRNA inhibits completion of the decoding process by the ribosome, which is attributed to decreased stability of the ribosome complexes [23]. These *in vitro* studies provide compelling evidence that mRNA modifications significantly impact the translation machinery in *E. coli.* However, mRNA modifications have yet to be comprehensively characterized in *E. coli*, hindering a broad understanding of their mechanisms and effects.

To address gaps in knowledge of prokaryotic RNA modifications, we here established a systematic DRS approach to thoroughly and simultaneously study numerous modification types in the rRNA, tRNA, and mRNA of *E. coli* K-12 (WT), an IVT RNA control, and an m^6A methyltransferase KO mutant. This approach was complemented by mass spectrometry, qPCR, and *in vitro* methylation to localize m^5C and m^6A and other modifications in the mRNA of *E. coli*, yielding a uniquely robust dataset that enabled the first comprehensive analysis of the epitranscriptomic landscape in *E. coli*. Crucially, our study not only advances understanding of RNA modifications in this model species, but establishes a blueprint for future epitranscriptomic analyses in prokaryotes.

MATERIALS AND METHODS

Strains and growth conditions

Experiments were conducted with the WT *E. coli* strain MG1655. Δ*rlmF*, Δ*rlmJ*, and the double mutant Δ*rlmF*Δ*rlmJ* were constructed using double homologous recombination for in-frame deletion with neomycin acetyltransferase and SacB as the selection and counterselection markers, respectively, as described by [24]. Δ*trmM* mutants were generated using the Quick & Easy *E. coli* Gene Deletion Kit (Gene Bridges GmbH, Heidelberg, Germany), which uses the RED®/ET® recombinase system [25]. All strains were grown in LB at 37 °C with 200 rpm shaking to an optical density at 600 nm (OD_{600}) of ~0.8–1. Primers used for mutant construction are listed in Supplementary Table 1.

Total RNA isolation

RNA was isolated using the PCI protocol [26] with some modifications (Petrov et al., 2022). Specifically, bacterial cultures were mixed with phenol and ethanol to final concentrations of 1% (v/v) and 20% (v/v), respectively, then flash-frozen in liquid nitrogen. Thawed samples were centrifuged at 16,000 × g for 10 min and the resulting pellets were resuspended in 500 µL of ice-cold sodium acetate buffer (20 mM, pH 5.2) with 1 mM EDTA (AE buffer). Extracted RNA was treated with RNase-free DNase I (New England Biolabs [NEB], Ipswich, MA, USA) following the manufacturer's protocol. RNA quality and integrity were assessed via chip gel electrophoresis with a 2100 Bioanalyzer and an RNA Nano chip kit (Agilent Technologies, Santa Clara, CA, USA).

tRNA/rRNA removal and mRNA enrichment

To remove tRNAs, RNA samples were purified using the RNA Clean & Concentrator Kit (Zymo Research, Irvine, CA, USA) following the manufacturer's protocol for exclusion of RNA below 200 nt. RNA quality was confirmed by electrophoresis using the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA,

USA). Then, oligo-based rRNA depletion was performed with the pan-riboPOOL rRNA Depletion Kit (siTOOLS, Planegg, Germany) according to the manufacturer's instructions. Depletion efficiency was assessed using Bioanalyzer mRNA chip electrophoresis (with Pico and Nano chips) and reverse transcription quantitative PCR (RT-qPCR) to amplify 23S and 16S rRNA and the mRNA marker *recA*. Ribosome depletion was calculated as the proportion of 23S or 16S to *recA* transcripts before and after the depletion was performed (Supplementary Table 2).

Mass spectrometry (MS) analysis of RNA methylation

Absolute quantitative analyses using isotope dilution MS were performed as previously described [9]. Briefly, RNAs (100–300 ng per sample) were enzymatically hydrolyzed to nucleosides by incubation with a mixture of benzonase (2 U), snake venom phosphodiesterase (0.2 U), and calf intestine phosphatase (2 U) for 2 h at 37 °C in a buffer containing Tris-HCl (5 mM, pH 8), deoxycoformycine (1 µg), and tetrahydrouridine (5 µg). For calibration, the nucleosides cytidine (C), uridine (U), guanosine (G), adenosine (A) (Sigma-Aldrich, St. Louis, MO, USA), and N6-methyladenosine (m^6A) (Carbosynth, Staad, Switzerland) were each mixed to a final concentration of 100 nM (canonical nucleosides) or 5 nM (m^6A). This stock calibration solution was serially diluted 1:1 to a total of 12 calibration levels. Prior to analysis, the improved gen $^{13}C/^{15}N$ stable isotope-labeled internal standard (SILIS) from *Saccharomyces cerevisiae* tRNA (PMID: 34085252) was added for absolute quantitative analysis. The resulting ribonucleotide mixture was separated using a Synergy Fusion RP column with a 2.5-µm particle size, 100-Å pore size, 100-mm length, and 2-mm inner diameter (Phenomenex, Torrance, CA, USA) on an Agilent 1290 Infinity II series ultra-high-performance liquid chromatography (UHPLC) instrument. Mobile phase A was 5 mM ammonium acetate adjusted to pH 5.3 with glacial acetic acid and mobile phase B was pure acetonitrile. Gradient elution began with 100% A for 1 min, increased to 10% B at 4 min and 40% B at 7–8 min, then returned to the starting condition of 100% A for an additional 2.5 min. The flow rate was 0.35 mL/min and the column temperature was 35 °C. For MS measurements, an Agilent 6470 Triple Quadrupole MS instrument set to dynamic multiple reaction monitoring mode was used. MS was conducted in positive ion mode with the following parameters: skimmer voltage of 15 V, cell accelerator voltage of 5 V, N_2 gas temperature of 230 °C, N_2 gas flow of 6 L/min, sheath gas (N_2) temperature of 400 °C with a flow of 12 L/min, capillary voltage of 2500 V, nozzle voltage of 0 V, and nebulizer at 40 psi. A list of monitored mass transitions is given in Supplementary Table 6. Data analysis was performed using MassHunter Quantitative software to yield molar amounts of analytes. Absolute modification abundance was calculated as the abundance of modified nucleosides per 1000 As.

Purification of RlmJ and RlmF

rlmJ and *rlmF* were cloned into the pET28 vector (bearing a C-terminal His6-tag) and the sequences were verified by sequencing. *E. coli* BL21(DE3) pLysS competent cells (Novagen, Darmstadt, Germany) containing the pET28-based expression system were cultured in LB at 37 °C to an OD_{600} of 0.5, then gene expression was induced by addition of 0.5 mM isopropyl β-D-thiogalactoside (IPTG) for 3 h at 37 °C to a final OD_{600} of ~2. Cell pellets were lysed by sonication in buffer containing 20 mM Tris-HCl (pH 7.0), 200 mM NaCl, 5% glycerol (v/v), and 5 mM β-mercaptoethanol as previously reported [27]. The lysate was centrifuged for 10 min at 5,000 x g, then for 30 min at 70,000 x g. The resulting supernatant was incubated with Ni-NTA Agarose (QIAGEN, Hilden, Germany) for 45 min at 4 °C, then washed with 30 volumes of 20 mM Tris-HCl (pH 7.0), 500 mM NaCl, 5% glycerol (v/v), 5 mM β-mercaptoethanol, and 10 mM imidazole. Enzymes were eluted with 20 mM Tris-HCl (pH 7.0), 500 mM NaCl, 5% glycerol (v/v), 5 mM β-mercaptoethanol, and 100 mM EDTA. Proteins were concentrated using Amicon® (Merck, Darmstadt, Germany) and stored at 4 °C for up to 1 month.

In vitro transcription and methyltransferase activity assays

Following the manufacturer's protocols, RNA was synthesized from a DNA template (Supplementary Table 3) using the HiScribe T7 High Yield RNA Synthesis Kit (NEB, Ipswich, MA, USA) and treated with DNase I (NEB, Ipswich, MA, USA) to eliminate DNA. The resulting RNA was purified and concentrated with a Monarch RNA Clean-up Kit (NEB, Ipswich, MA, USA), quantified, and stored at -80 °C prior to further use. Before the methylation reaction, IVT RNA (50 to 200 pmol) was heated for 1 min at 95 °C, then cooled on ice. The reaction was carried out for 10 min at 37 °C in a buffer containing 20 mM potassium phosphate (pH 6.5), 50 mM NaCl, 5 mM β-mercaptoethanol, 1 U/µl murine RNase inhibitor (NEB, Ipswich, MA, USA), and 6 µg of pure enzyme (RlmJ/RlmF) or BSA (NEB, Ipswich, MA, USA) as a control [27]. Finally, 0.5 µM S-adenosyl-L-[methyl-^{3}H]-methionine (SAM) (5–15 Ci/mmol; American Radiolabeled Chemicals, Inc., Saint Louis, MO, USA) was added to a total volume of 50 µL to start the reaction. After 10 min, the reaction was stopped using RNA binding buffer from the RNA Clean & Concentrator Kit (Zymo Research, Irvine, CA, USA) plus ethanol following the manufacturer's protocol. Purified RNA was diluted in scintillation liquid, homogenized, and quantified in a Tri-Carb® 2910 TR scintillation counter (PerkinElmer, Waltham, MA, USA). Previously reported methylated sites in the 23S were used as the positive controls. For RlmJ, the positive control was a 28-nt fragment containing the A2030 position of the 23S rRNA with an additional 5'-GG sequence [28]. For RlmF, a 30-nt fragment from 1601 to 1631 of the *E. coli* 23S rRNA containing A1618 served as the positive control.

SELECT method for m^{6}A detection

The SELECT method was conducted as previously described [29]. Briefly, total RNA was hybridized with 40 nM up/down primers (Supplementary Table 4), then combined with 5 µM dNTPs and 2 µL of 10× rCutSmart buffer (NEB, Ipswich, MA, USA) in a total reaction volume of 17 µL. The mixture was

annealed sequentially at 90, 80, 70, 60, and 50 °C for 1 min each, then at 40 °C for 6 min. Subsequently, 3 µL of a ligation mixture containing 0.01 U of Bst 2.0 DNA polymerase (NEB, Ipswich, MA, USA), 0.5 U of SplintR ligase (NEB, Ipswich, MA, USA), and 10 nmol ATP was added. The mixture was incubated at 40 °C for 20 min, then denatured at 80 °C for 20 min. RT-qPCR was conducted with 1 µL of the product using 200 nM of SELECT specific forward and reverse primers.

IVT RNA control preparation

The IVT RNA control was prepared as previously described [30, 31] with some modifications. Briefly, 50 ng of polyadenylated rRNA-depleted *E. coli* RNA was used as the starting material. VN and Strand-Switching primers from the cDNA-PCR Sequencing Kit (ONT #SQK-PCS109) were used with Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific, Waltham, MA, USA) for reverse transcription and strand switching with the following parameters: 90 min at 42 °C (one cycle), heat inactivation for 5 min at 85 °C (one cycle), then holding at 4 °C. The RT RNA sample (5 µL) was amplified using the LongAmp Taq 2× Master Mix (NEB, Ipswich, MA, USA) with IVT Nanopore T7 Fw and IVT Nanopore T7 Rv primers (Supplementary Table 5). The thermocycling conditions were as follows: initial denaturation for 30 s at 95 °C (one cycle), denaturation for 15 s at 95 °C (11 cycles), annealing for 15 s at 62 °C (11 cycles), extension for 5 min at 65 °C (11 cycles), and a final extension for 6 min at 65 °C (one cycle), followed by holding at 10 °C. The PCR product was treated with Exonuclease I (NEB, Ipswich, MA, USA) for 15 min at 37 °C, followed by 15 min at 80 °C to eliminate single-stranded products. DNA was purified using AMPure XP (Beckman Coulter, Brea, CA, USA) following the manufacturer's instructions for the cDNA-PCR Sequencing Kit (ONT #SQK-PCS109). The PCR product was used as the template for IVT with the HiScribe T7 High Yield RNA Synthesis Kit (NEB, Ipswich, MA, USA). Samples were then treated with DNase I and cleaned with the Monarch RNA Cleanup Kit (NEB, Ipswich, MA, USA) following the manufacturer's protocol. RNA was polyadenylated with *E. coli* Poly(A) Polymerase (NEB, Ipswich, MA, USA) and cleaned with the Monarch RNA Cleanup Kit (NEB, Ipswich, MA, USA). Ribosomal 23S and 16S samples were prepared by amplifying the appropriate sequences from genomic *E. coli* DNA (Supplementary Table 5) and adding the T7 sequence to the forward primer. IVT was performed following the manufacturer's protocol for the HiScribe T7 High Yield RNA Synthesis Kit (NEB, Ipswich, MA, USA). Polyadenylated RNA was added to the IVT RNA control at a ratio of 1:1000.

Nanopore sequencing

mRNA-enriched samples and IVT RNA controls were analyzed via DRS using the MinION platform (ONT, Oxford, UK). The DNase-treated and tRNA-depleted RNA samples were polyadenylated with *E. coli* Poly(A) Polymerase for 30 min following the manufacturer's protocol (NEB, Ipswich, MA, USA). Polyadenylation levels were assessed with Bioanalyzer chip gel electrophoresis and through evaluation

of a size shift increase of ~500 bp for the 23S and 16S rRNAs. rRNA was depleted as described above. Nanopore libraries were prepared with ~200 ng of polyadenylated *E. coli* mRNA or IVT RNA following the manufacturer's protocol (ONT #SQK-RNA002). Each sample was sequenced on an individual MinION R9.4.1 flow cell.

ONT data processing

Reads from ONT sequencing were basecalled using Guppy v6.5.7 (ONT, Oxford, UK). With the two biological replicates of each sample type combined to form a single sample each, Minimap2 (v2.26) was used for alignment to a modified version of the ensemble *E. coli* reference genome (Escherichia_coli_str_k_12_substr_mg1655_gca_000005845.ASM584v2) with duplicate tRNA and rRNA genes removed and to a transcriptome generated using the coordinates of transcriptional units, including 5' and 3' untranslated regions (UTRs) [32]. Aligned reads were filtered using SAMtools (v1.17) [33] to remove secondary alignments, then quantified using Salmon [34] and analyzed with ELIGOS [11] to identify putative modifications of all types. Aligned reads were also indexed and preprocessed with Nanopolish (v0.14) [35] before analysis with CHEUI [36] for detection of only m^6A and m^5C and analysis with Nanocompore [37] for detection of all modification types. With the exception of a minimum read count of 20 in Nanocompore, all programs were run with default parameters.

Putative modification site analyses

Biotype annotations. Putative modification sites were classified as present within the tRNA, rRNA, ncRNA, coding sequence (CDS), 5' UTR, 3' UTR, or intergenic region using the annotated *E. coli* K-12 substr. MG1655 ASM584v2 reference genome (RefSeq #GCF_000005845.2) and the genomic coordinates of UTRs [32].

Thresholding. Precision–recall curves were generated for modifications in the rRNA and tRNA using a range of thresholds for each of several values generated by ELIGOS: the percent error of specific bases (% ESB) in the WT sample (ESB test), ESB in the IVT control sample (ESB control), the odds ratio, and the adjusted *p*-value. Precision and recall were also calculated for all possible combinations of the threshold values tested for these four parameters. For these analyses, recall was calculated using known modification sites only. Precision was calculated considering known modification sites ± 10 nt within continuous regions of putative modification sites as true signals due to the known noise caused by modifications in Nanopore data. The final thresholds for the ESB test, ESB control, odds ratio, and *p*-value were selected by taking the combination of threshold values that resulted in the highest recall at a precision of ≥ 85% (i.e., error ≤ 15%). Where multiple values of a single parameter resulted in identical precision and recall values, the least restrictive threshold value was selected.

Modified region identification and characterization. The threshold values established as described above were applied to modification sites called by ELIGOS in all biotypes other than the tRNA and rRNA (namely the CDS, ncRNAs, intergenic regions, and 5′ and 3′ UTRs). Due to the presence of known modification sites in the tRNA and rRNA, true modification sites could be distinguished from noise within regions of adjacent nt in which all sites were called as modified by ELIGOS; because this was not possible for the other gene biotypes, each group of adjacent nt in which every site was called as modified by ELIGOS was considered a single "signal region" in which one or more modifications were present. The relative abundance of modifications in each gene biotype was visualized with the R packages 'ggplot2' [38] and 'treemapify' (https://CRAN.R-project.org/package=treemapify). The distribution of modifications across a model gene transcript was determined by calculating the normalized relative location of each modification within the 5′ UTR, CDS, or 3′ UTR.

Statistical analyses

Differences between pairs of samples were assessed with an unpaired Student's *t*-test. Gene Ontology (GO) enrichment was calculated with FUNAGE-Pro v2 [39].

Data availability

Sequencing data have been deposited at NCBI Gene Expression Omnibus (GEO) under accession number XXX.

This study does not describe any novel programs, software, or algorithms.

RESULTS

Direct RNA Sequencing (DRS) using ONT

RNA was extracted from two independent biological replicates of *E. coli* MG1655 cells (WT) grown to mid-exponential growth phase in LB medium. Due to the high abundance of tRNA and rRNA molecules, a tRNA/rRNA depletion strategy was applied (Figure 1A). Specifically, tRNA was depleted using an RNA purification column, which can eliminate RNAs of < 200 nt. The RNA was then polyadenylated, and 16S and 23S rRNAs were depleted using a specific oligo-based rRNA depletion kit (Figure 1A). At each step, the quantity and quality of RNA were assessed by RNA chip electrophoresis. tRNA depletion was evaluated based on the signal at 100 nt and rRNA depletion was assessed based on the presence of peaks at ~1,500 nt (16S) and ~2,900 nt (23S) (Figure 1A). The presence of 16S and 23S rRNA was also quantified in the mRNA-enriched samples via RT-qPCR; these samples contained 0.4–0.6% and 0.1% of

the 16S and 23S rRNA levels of the non-depleted samples, respectively (Supplementary Figure 1A). The tRNA/rRNA-depleted samples are referred to as "WT mRNA".

As an unmodified RNA control, IVT RNA was prepared by adapting available protocols [30, 31] for prokaryotic samples (Figure 1B). Specifically, polyadenylated (tRNA-/rRNA-depleted) RNA (Supplementary Figure 1B) was prepared and a single reverse primer was used to initiate reverse transcription of all molecules. A second primer was used to switch the strands and produce cDNA from *E. coli* RNA. The DNA was then amplified by PCR, incorporating the T7 promoter into the forward primer (Figure 1B). The resulting PCR product was used as a template for IVT, yielding unmodified RNAs (Supplementary Figure 1C; Figure 1B). There was a notable difference in the relative abundance of large RNAs (~1000–3000 nt) between the WT and IVT samples at this stage, likely due to the extremely high abundance of the transfer-messenger RNA (tmRNA) *ssrA* in the IVT samples as discussed below. After a final polyadenylation step (Supplementary Figure 1D; Figure 1B), the samples were prepared for DRS on the ONT MinION platform (Figure 1C).

Two independent biological replicates of the WT mRNA and the IVT RNA control were sequenced individually, then the replicate samples were combined for further analyses. The *E. coli* genome contains duplicate tRNA and rRNA genes encoding identical molecules that cannot be effectively distinguished from each another. Therefore, to ensure accuracy, reads were mapped to a modified version of the *E. coli* transcriptome [40] containing a single representative copy of each duplicate tRNA and rRNA gene. For example, only the 16S and 23S genes *rrsH* and *rrlH*, respectively, were retained. Sequencing produced 1.3 million and 1 million reads from the WT and IVT RNA control samples, respectively, that could be mapped to this version of the *E. coli* transcriptome (Figure 1D).

Per-gene sequencing read depths were quantified with Salmon [34]. In the WT sample, 45% of the sequencing reads mapped to the mRNA (Figure 1D). Approximately 2,154 transcripts (49.4% of all *E. coli* mRNAs) were present at a depth of 20 or more reads, which has been reported as the minimum depth required to accurately analyze RNA modifications [11] (Supplementary Figure 1E). Among the ncRNAs, the tmRNA *ssrA* was the most highly abundant, accounting for ~25% of all reads in the WT samples. This was significantly more than the ~10–15% previously detected with Illumina sequencing [41]. rRNAs and tRNAs accounted for ~15% and ~1% of all reads in the WT, respectively. In the IVT RNA control, 21.7% of reads mapped to mRNA sequences, corresponding to 1,379 transcripts (31.6%) that were present at a sufficient sequencing depth for further analyses (Supplementary Figure 1F). Of the RNAs that were not present at a sufficient sequencing depth, there was no statistically significant difference in RNA size in the IVT compared to the WT sample (Supplementary Figure 1G). Notably, the IVT RNA control sample had a very high level of tmRNA (~70% of all reads), a serious distortion of the true abundance of this RNA *in vivo*. The increased abundance in the IVT samples was likely due to the

presence of a polyA tail in the tmRNA. rRNAs and tRNAs accounted for 0.13% and 0.78%, respectively, in the IVT RNA control sample; the rRNA abundance in the final sample was dependent on the amount spiked in after IVT (see Methods). Even at this low read depth, at least 33 (68.8%) of the 48 unique *E. coli* tRNA molecules were detected at a read depth of at least 20 in both the WT and IVT RNA control samples (Figure 1D), indicating that the numerous known modifications in tRNA and rRNA (in addition to those in other regions) could be used to validate our sequencing results.

Detection of known RNA modifications in *E. coli*

We first searched for known modifications in the mRNA using ELIGOS [11], which identifies modifications by comparing the percent error of specific bases (%ESB) produced during Nanopore sequencing of a modified WT RNA sample with those in an unmodified IVT control; the proportion of %ESB in the WT compared to the IVT sample is referred to as the odds ratio (OddR). A recent study using Pseudo-Seq demonstrated the presence of 44 Ψ sites in *E. coli* mRNAs [18]. Using ELIGOS, we were able to identify nine of these sites, in the CDSs of *rplA*, *tufB*, *acnB*, *rpsQ*, *tufA, secY, rpsA, ptsI*, and *fimA* (Figure 2A). Moreover, the known tmRNA modifications at positions 341 (m^5U) and 342 (Ψ) [42, 43] were detected in our dataset (Figure 2B).

Depletion protocols are not 100% effective in removing tRNAs and rRNAs. We therefore took advantage of the presence of tRNAs and rRNAs remaining in our samples by analyzing known modifications in these reads. Importantly, the tRNA read depths were comparable between our samples and those from a prior publication describing analysis of solely tRNAs (Wang et al., 2021). Furthermore, coverage across the entire rRNA was more than sufficient for accurate detection of modifications. Comparing the WT signal to the systematic sequencing noise as determined with the IVT RNA control, 59.7% of all known modifications in the rRNA and tRNA produced signals that were detectable with ELIGOS (Supplementary Figure 2A). In a previous study, it was shown that the modification types m^6A, 1-methyladenosine (m^1A), 5-methoxyuridine (5moU), Ψ, 7-methylguanosine (m^7G), inosine (Ino), 5-hydroxymethylcytidine (hm^5C), and 5-formylcytidine (f^5C) can be detected with ELIGOS, but m^5C cannot [11]. Similarly, we were able to detect signals for 24 of the 34 different types of modifications described in *E. coli* rRNA and tRNA, but the modification types 4-thiouridine (s^4U), 2-thiocytidine (s^2C), m^5C, 5-methylaminomethyluridine (mnm^5U), 2-lysidine (k^2C), N^6-methyl-N^6-threonylcarbamoyladenosine (m^6t^6A), I, N^6-isopentenyladenosine (i^6A), 3-methyluridine (m^3U), and N4,2′-O-dimethylcytidine (m^4Cm) were not detected.

After controlling for sequencing noise based on the IVT sample, we identified 30 out of 36 known rRNA modifications (83.3%) in the WT mRNA sample. However, excluding the types of modifications that are not known to be detectable with ELIGOS, there were 31 known rRNA modifications; i.e., 96.8% of the expected modifications were detected (Figure 3A). m^6A could be detected at position 2030 but not at

position 1618 of the 23S rRNA, which is consistent with a previous report [17] (Figure 3A, B). It was not possible to identify m^5C at C967/C1407 of the 16S rRNA or C1962 of the 23S rRNA, or to detect m^4Cm at 1402 of the 16S rRNA (Figure 3A–C), presumably because the signal alterations produced by m^5C and m^4Cm were not strong enough to be classified as errors during basecalling [11]. Notably, the m^3U modification of the 16S rRNA did not produce a signal at the known modification site (position 1498), but did produce a signal 5 nt upstream (Figure 3A, C).

In the tRNA, we were able to detect signals in 198 out of all 346 known modification sites (57.2%) (Figure 3D, E). Considering only the modifications that are known to be detectable with ELIGOS and the tRNAs that were present at a sufficient sequencing depth in our dataset (72.7% of the unique tRNA molecules with known modifications), we were able to detect 83.2% of known tRNA modifications (Figure 3D). The signals produced by modifications in the tRNAs *hisR*, *cysT*, and *ileT* are shown as examples (Figure 3E).

Due to the previously demonstrated effects of modifications on neighboring nt in Nanopore sequencing [11], the signals surrounding known modifications in the tRNA and rRNA were analyzed. We found that modifications consistently produced detectable signals up to 10 nt upstream and downstream of known modification sites (Supplementary Figure 2B–D). This made it difficult to determine whether the signal at each individual nt, especially in regions without well-characterized modification sites, originated from one or more closely placed modifications.

Detection of m^5C in *E. coli* RNA

Previous studies have evaluated the presence of m^5C in *E. coli* RNA using bisulfite treatment coupled with reverse transcription and DNA sequencing. Using this strategy, known m^5C sites have been detected at positions 967 and 1407 of the 16S rRNA and 1962 of the 23S rRNA, but due to a lack of sequencing depth in the CDS, the presence of m^5C in *E. coli* mRNA could neither be confirmed nor refuted [44]. Although ELIGOS is also unable to detect m^5C modifications, another program developed for the analysis of Nanopore data, CHEUI, does not have this limitation [36]. CHEUI can be run in CHEUI-diff mode to evaluate the differences between paired samples (e.g., WT compared to IVT RNA or a mutant sample) or in CHEUI-solo mode to calculate the probability that a given 9-mer is modified. Here, a comparison of the WT mRNA to the IVT RNA control was performed with CHEUI-diff using thresholds of stoichiometry difference ≥ 0.1 and *p*-value ≤ 0.05 (Figure 4A). Samples were also processed individually with CHEUI-solo, considering only sites with a probability ≥ 0.9999 in the WT and < 0.9999 in the IVT RNA control. Ten 9-mers were detected using this approach (Figure 4B): the known m^5C modification at position 967 of the 16S rRNA; the known m^5C modification at 1962 of the 23S rRNA and the adjacent cytosine (Figure 4A); 5-hydroxycytidine (ho^5C) at site C2501 of the 23S

rRNA; 2'-O-ribose (Cm) at site C2498 of the 23S rRNA and the adjacent cytosine (Figure 4A); and putative modifications in the mRNA *ftsk* (genomic positions 933138 and 933139), the small RNA *sdhX* (genomic position 765142), and the tRNA *thrU* (genomic position 4175443) (Figure 4A). Using less stringent thresholds, the m^4Cm at position 1402 of the 16S could be distinguished in the CHEUI m^5C analysis (stoichiometry difference 0.51, *p*-value 9.97×10^{-35}, and probability 0.999), as could the m^5C at site C1407 of the 16S (stoichiometry difference 0.42, *p*-value 2.53×10^{-16}, and probability 0.989).

The stoichiometry values predicted from Nanopore sequencing of the WT and IVT RNA control samples and the calculated stoichiometry differences were plotted for the regions surrounding known m^5C sites in *E. coli* at positions 967 (Figure 4C) and 1407 (Figure 4D) of the 16S rRNA and 1962 of the 23S rRNA (Figure 4E). These well-characterized modifications could be clearly distinguished from background noise using the stoichiometry predictions generated with CHEUI, validating this approach. Consistent with previous results from bisulfite sequencing [44], m^5C methylation levels were lower at the 16S rRNA position 1407 than at the 16S position 967 or the 23S rRNA position 1962 (Figure 4D).

m^6A characterization in *E. coli* RNA

To further characterize m^6A in *E. coli* mRNA, mutants lacking the previously identified *E. coli* m^6A rRNA methyltransferase genes *rlmF* [45] and *rlmJ* [28] were generated. We also examined the mutant *trmM,* which lacks the gene encoding the methyltransferase responsible for m^6A methylation of $tRNA^{Val}$ position A37 [46]. Mass spectrometry was used to quantify m^6A contents in rRNA + mRNA (tRNA-depleted) samples of WT *E. coli*, the single mutants Δ*rlmF*, Δ*rlmJ*, and Δ*trmM*, and the double mutant Δ*rlmF*Δ*rlmJ* (Figure 5A). In addition, the double deletion mutant was complemented *in cis* with *rlmF* and *rlmJ*. The WT strain and the complemented mutant had similar m^6A levels (Figure 5A). The single deletion mutants Δ*rlmF* and Δ*rlmJ* showed half the methylation levels of the WT strain, while m^6A was barely detectable in the double deletion mutant. The Δ*trmM* mutant had similar m^6A contents as the WT, confirming that TrmM does not play a role in rRNA or mRNA methylation. To confirm the functionality of TrmM, tRNAs (< 200 nt) were purified from WT and Δ*trmM E. coli*; the mutant showed an eight-fold decrease in m^6A levels compared to the WT strain, confirming its role as a tRNA m^6A methyltransferase (Supplementary Figure 3A).

METTL16 is a homolog of RlmF [47] that has been shown to methylate MAT2A mRNA and U6 small nuclear RNA (U6 snRNA) in eukaryotes [48]. Because transcription and translation occur in a single compartment in prokaryotes, we tested whether one of these methyltransferases could also methylate mRNA in *E. coli*. As described above, mRNA was enriched through tRNA/rRNA depletion (Figure 1A); rRNA was reduced to maximum levels of 1% (16S rRNA) and 0.3% (23S rRNA) of the total RNA (Supplementary Figure 3B, C), indicating successful mRNA enrichment. Samples from WT *E. coli*, the

complemented Δ*rlmF*Δ*rlmJ* mutant, and the Δ*trmM* mutant showed comparable m^6A levels (Figure 5A). In the mRNA of the Δ*rlmF*Δ*rlmJ* double deletion mutant, we detected 0.078 m^6A per 1000 A, compared to 0.147 m^6A per 1000 A in WT mRNA (Figure 5B). This suggested that the rRNA methyltransferases RlmF and RlmJ, but not TrmM, were partially responsible for mRNA methylation in *E. coli*.

Next, *in vitro* methylation assays were performed to evaluate whether these methyltransferases can methylate additional sequences beyond those previously reported in the rRNA. First, the *in vitro* activities of purified RlmF and RlmJ were measured using fragments of 23S rRNA containing known methylation sites as positive controls. Specifically, a previously reported 28-nt fragment containing A2030 of the 23S RNA with an additional 5'-GG sequence [28] was used to test RlmJ activity. Similarly, a 30-nt fragment from 1601 to 1631 of the *E. coli* 23S rRNA containing A1618 was used as the positive control for RlmF activity. After confirming that the two enzymes could methylate the positive controls (Figure 5C–E), we examined their activity on various *in vitro*-synthesized RNA oligos from genes containing the sequences CACAGG or GUGAAGA (the consensus sequences for RlmF and RlmJ, respectively). The results indicated that RlmF was indeed able to methylate adenosine in sequences containing the consensus motif: *cysJ* (214 nt) and *phoP* (251 nt) were methylated at levels comparable to the positive control A1618 (Figure 5D). A sequence derived from *rpoH* (239 nt) was methylated, albeit at a lower level than the rRNA-derived sequence (Figure 5D). RlmJ was able to methylate the *uvrY* (241 nt)-derived sequence at intermediate levels and the *ppX* (188 nt) and *dnaK* (262 nt) sequences at a low level (Figure 5E). Fragments derived from *cas3* (243 nt) and *mazE* (177 nt) were not methylated, although they contained the appropriate motifs (Figure 5D, E). We also found that RlmF had greater and faster methylation activity than RlmJ (Supplementary Figure 3D, E). Overall, these results indicate that RlmJ and RlmF can methylate *E. coli* mRNA sequences *in vitro*.

Detection of m^6A in *E. coli* RNA using DRS

To evaluate the presence of m^6A in *E. coli*, WT and IVT RNA control samples were compared to identify m^6A positions throughout the *E. coli* transcriptome. Several algorithms have been developed to detect m^6A in RNA sequenced with ONT [49]. We compared three programs that are able to distinguish m^6A modifications independent of sequence context using different features: ELIGOS, which is based on modification-induced basecalling errors [11]; Nanocompore, which identifies modifications based on the electric currents generated during Nanopore sequencing [37]; and CHEUI-diff m^6A, which implements a two-stage neural network [36].

The three data sets were compared using the parameters suggested by the authors, namely an OddR ≥ 2.5 and adjusted *p*-value ≤ 0.00001 for ELIGOS, a Gaussian mixture model (GMM) logit *p*-value ≤ 0.01

for Nanocompore, and a stoichiometry difference ≥ 0.1 and *p*-value ≤ 0.05 for CHEUI-diff m^6A. A total of 13 positions in the tRNA and rRNA were identified by all three algorithms (referred to as shared sites). The shared sites included the known m^6A position A2030 of 23S rRNA (Figure 6A). Other shared sites were 1916 and 1919 of the 23S rRNA, which are near the known modification sites Ψ1915 and 1917, respectively. Similarly, the shared site A2450 is close to the known m^2G site 2445; the shared sites A2459 and A2460 are near the known modification site Ψ2457; and the shared sites A2600 and A2602 are close to the known Ψ sites 2604 and 2605 of the 23S. All three algorithms also detected A38 in *thrU*, which is adjacent to the t^6A modification at site 37; m^2A38 of *hisR*; A37 and A39 of *tyrU*, which surround the ms^2i^6A modification at site 38; and A44 of *thrT*, which is near the m^7G modification at site 46 (Figure 6A).

To identify positions methylated by RlmF or RlmJ, we isolated RNA from the Δ*rlmF*Δ*rlmJ* mutant, depleted the sample of tRNA and rRNA, and processed it with Nanopore sequencing. The data were then analyzed with ELIGOS, Nanocompore, and CHEUI-diff m^6A, comparing the Δ*rlmF*Δ*rlmJ* sample to the WT (Figure 6B). This showed only the presence of the known modification at A2030 of the 23S rRNA. An *in vivo* assay, the "single-base elongation- and ligation-based qPCR amplification method" (SELECT) [29], was then performed to independently confirm the presence of two m^6A positions in the 23S rRNA. This approach relies on the ability of m^6A to hinder the retrotranscriptional activity of BstII. The method was used to detect reported m^6A sites in total RNA samples (Figure 6D). Comparison of threshold cycle (C_t) values between the WT and Δ*rlmF*Δ*rlmJ* strains for each m^6A site indicated that SELECT was able to detect the known modified 23S rRNA positions A1618 and A2030, validating the sequencing results.

Putative novel modification sites in *E. coli* RNAs and their functional implications

Finally, we assessed the landscape of RNA modifications in *E. coli* outside of the tRNA and rRNA. The DRS data were processed with ELIGOS [11] and Nanocompore [37] pairing the WT and IVT RNA control samples. As with ELIGOS (Figure 3A, D), we benchmarked the performance of Nanocompore using the set of all known modification sites in the *E. coli* tRNA and rRNA [50]. Putative modification sites were retained only if they had a *p*-value < 0.05 from the GMM logit test or from the Kolmogorov-Smirnov test for current intensity or dwell time. In total, Nanocompore detected 24.9% of the expected tRNA and rRNA modifications (Supplementary Figure 5). After exclusion of RNAs that were not present at a depth of ≥ 20 reads, Nanocompore could identify 32.4% of the known tRNA and rRNA modifications. ELIGOS thus showed superior performance (Supplementary Figure 2A) and was used in further analyses.

To minimize error and maximize the detection of truly modified sites, we analyzed precision and recall among tRNA and rRNA regions using thresholds of varying stringency. Precision was calculated as the percentage of truly modified positions out of the total number of positions classified as modified, and recall was calculated as the percentage of known modified sites detected in our dataset out of the total number of known modified sites. Thresholds were tested for several parameters output by ELIGOS: the percent error at each base in the *E. coli* RNA sample (%ESB test) (Supplementary Figure 4A–C), the percent error at each base in the IVT RNA control sample (%ESB control) (Supplementary Figure 4D–F), the OddR (%ESB test divided by %ESB control) (Supplementary Figure 4G–I), and the adjusted *p*-value (Supplementary Figure 4J–L). In total, we analyzed precision and recall for 32,000 combinations of these four parameters (Figure 7A). Thresholding with the optimal values (%ESB test ≥ 0.1, %ESB control ≤ 1, OddR ≥ 3, and adjusted $p \leq 0.05$) yielded a precision of 85.9% (i.e., an error rate of 14.1%) and a recall of 58.8% in the combined tRNA/rRNA dataset.

These high-confidence thresholds were then applied to the remaining RNA biotypes: mRNAs, including the 5′ UTRs, CDSs, and 3′ UTRs; ncRNAs; and intergenic regions. Due to the inherent noisiness of DRS data, which consistently produced signals extending ~10 nt beyond known modification sites in the tRNA and rRNA (Supplementary Figure 2B–D), regions of adjacent nt in which all sites were classified as modified were each considered a single "modification region". After thresholding with the values established using the tRNA and rRNA data, we identified 283 putative novel modification sites/regions in CDSs, 56 in 5′ UTRs, 24 in intergenic regions, 14 in 3′ UTRs, and 10 in ncRNAs (Figure 7B).

To determine the general distribution of modifications across each transcript, the putative modification sites or regions were plotted based on their relative locations within each mRNA. This revealed enrichment of putative RNA modifications toward the beginning and end of the CDS (Figure 7C). Nucleotide-specific analysis of the modification distributions showed that As were the most abundantly modified base in the 5′ UTR and that modified Cs, Us, and Gs were most strongly enriched toward the beginning of the CDS (Figure 7C). Gene Ontology (GO) enrichment analysis was then performed for the putative modifications identified in mRNAs. Significantly enriched GO terms included those related to functions in general metabolism and RNA processing (Figure 7D).

DISCUSSION

Numerous epitranscriptomic studies have been conducted in eukaryotes from yeast to humans, and the machineries responsible for m^6A and m^5C modifications are well characterized. In comparison, mRNA modifications in bacteria remain poorly understood. One of the main reasons for this gap in epitranscriptomic knowledge is the difficulty in isolating mRNA from bacteria. The absence of poly(A)

tails and the high abundance of rRNA make the mRNA isolation process a challenging endeavor. Currently, the most widely used protocols for ribosome depletion use specific biotinylated probes to eliminate rRNA. Using biotinylated probes, we were able to reduce the 16S and 23S rRNA levels to 1% and 0.3% of the total RNA, respectively. Although mRNA isolation methods have been greatly improved in recent years, complete rRNA removal is still not achievable [51]. Importantly, the requirement for polyadenylation of bacterial RNA prior to sequencing is actually beneficial as it allows simultaneous evaluation of all RNA types (including mRNAs, ncRNAs, tRNAs, and rRNAs). Moreover, third-generation sequencing approaches provide key advantages over earlier methods: they are independent of rRNA depletion and facilitate the study of native-form RNA modifications with single-molecule resolution without the need for potentially harmful chemicals. However, DRS data are extremely noisy, necessitating the analysis of unmodified controls (e.g., IVT *E. coli* RNA) to eliminate systematic error signals associated with the sequencing process itself. In addition to an IVT control, to minimize errors while maximizing the detection of truly modified sites, we used the set of known tRNA and rRNA modifications to establish an optimal combination of thresholds for several parameters associated with modification abundance and confidence. Overall, this approach proved to be suitable to distinguish ~60% of all known tRNA and rRNA modifications in *E. coli*. Of the 34 different types of modifications known in *E. coli* tRNA and rRNA, we were able to detect 24 (including D, Ψ, m^5U, m^7G, and Gm) using ELIGOS (Supplementary Figure 2A). m^5C modifications were not detectable using ELIGOS (Supplementary Figure 2A), but were detected using another program, CHEUI (Figure 4A).

Although the DRS signals derived from tRNAs were noisier than those from rRNAs, it was possible to detect and evaluate many (~70%) of the total number of bacterial tRNAs. Our approach outperformed previous studies that used methods based on the interference of base modifications with RT enzymes, likely because some modifications (such as Ψ and 5-methyluridine) that produce little or no RT impairment [52, 53] can be easily identified with DRS. Thus, the approach used here enabled the observation of modifications in mRNA beyond Ψ, the only type of mRNA modification in *E. coli* that has been validated to date [18].

Without a separate treatment such as bisulfite [44], the dataset generated in the present study was suitable for the assessment of m^5C abundance in bacterial RNA. Although we successfully detected known m^5C positions in addition to Cm, ho^5C, and m^4Cm modifications in the 16S and 23S rRNAs, the presence of m^5C in *E. coli* mRNA was not supported. However, despite the advances of our approach compared to previous methods, the sequencing depth was not sufficient to evaluate all coding genes in the *E. coli* genome. Therefore, our results do not conclusively prove the absence of m^5C in *E. coli* mRNA.

Three different programs (ELIGOS, Nanocompore, and CHEUI-diff m^6A) were used to assess the presence of m^6A modifications in *E. coli* RNA [11, 36, 37]. We compared DRS data from the WT with both the IVT RNA control and the Δ*rlmF*Δ*rlmJ* mutant, which lacks the genes encoding m^6A methyltransferases for 23S rRNA. We detected position A2030 but not A1618 of the 23S rRNA, similar to a previous report [17]. Using DRS data, we could not detect previously reported methylated positions obtained using PA-m^6A-seq, a sequencing method that relies on antibody recognition of m^6A [19]. Notably, the m^6A abundance in the mRNA samples prepared according to our protocol was ~0.15 per 1000 A, which is less than 10% of the level previously reported [19]. Nevertheless, our MS data support the idea that RlmF (an ortholog of METTL16) and RlmJ are at least partially responsible for m^6A modifications in mRNA. This is conceivable as bacteria do not have a nucleus and all reactions take place in one compartment. Using the DRS approach, we were not able to detect any specific m^6A site in *E. coli*. This could be due to limitations in the sensitivity of the detection, the presence of m^6A modifications in mRNAs with low abundance, low levels of modification, or heterogeneity within the population. It should be noted that the presence of m^6A modifications in viral RNA genomes has recently been questioned [54]. Previously, using immunoprecipitation and sequencing approaches, Chikungunya virus (CHIKV) and dengue virus (DENV) were shown to contain m^6A modifications [55, 56]. A renewed investigation with a combination of techniques, such as SELECT, mass spectrometry, and DRS could not confirm these earlier findings [54]. Our work similarly emphasizes the need to use orthogonal methods to validate putative m^6A positions.

Using a combination of thresholds designed to maximize the detection of modified sites while keeping the error rate below 15%, we assessed the overall distribution of modifications throughout the *E. coli* transcriptome. This revealed the presence of 387 putative modifications outside the tRNAs and rRNAs (i.e., in the mRNAs, ncRNAs, and intergenic regions). Many of the putative mRNA modification sites were present in genes associated with general metabolism and RNA processing, suggesting that a primary function of mRNA modifications in *E. coli* may be related to the regulation of RNA itself.

In summary, we present here the first systematic study of RNA modifications in *E. coli.* Decoding the epitranscriptome of bacteria required a combination of complementary techniques, including DRS of high-quality RNA, mass spectrometry, single-nucleotide PCR, and methyltransferase assays. This allowed us to simultaneously detect numerous modification types across RNA biotypes, including rRNAs, tRNAs, and mRNAs. On this basis, two very important modifications for eukaryotic mRNA, m^5C and m^6A, could be analyzed in more detail. Our data indicate that these modifications are only present at very low levels in *E. coli* mRNA. Furthermore, putative mRNA modification sites of all types tend to be localized to the beginning and end of the CDS and in the 5' UTR in *E. coli*. This study provides a useful

resource for experimental and bioinformatic approaches to gain new insights into post-transcriptional regulation in a prokaryotic model.

ACKNOWLEDGEMENTS

This work was financially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): Project numbers 325871075 (SFB 1309) (to KJ and PG) and 464582101 (JU270/21-1 to KJ). SR acknowledges financial support by the Alexander von Humboldt foundation. L.V. acknowledges the support of the Graduate School Life Science Munich (LSM).

REFERENCES

1. Cech, T.R., *Structural biology. The ribosome is a ribozyme.* Science, 2000. **289**(5481): p. 878-9.
2. Cunningham, P.R., et al., *The absence of modified nucleotides affects both in vitro assembly and in vitro function of the 30S ribosomal subunit of Escherichia coli.* Biochimie, 1991. **73**(6): p. 789-96.
3. Green, R. and H.F. Noller, *In vitro complementation analysis localizes 23S rRNA posttranscriptional modifications that are required for Escherichia coli 50S ribosomal subunit assembly and function.* RNA, 1996. **2**(10): p. 1011-21.
4. Antoine, L., et al., *RNA Modifications in Pathogenic Bacteria: Impact on Host Adaptation and Virulence.* Genes (Basel), 2021. **12**(8).
5. Lorenz, C., C.E. Lunse, and M. Morl, *tRNA Modifications: Impact on Structure and Thermal Adaptation.* Biomolecules, 2017. **7**(2).
6. Eargle, J., et al., *Dynamics of Recognition between tRNA and elongation factor Tu.* J Mol Biol, 2008. **377**(5): p. 1382-405.
7. Giege, R. and M. Springer, *Aminoacyl-tRNA Synthetases in the Bacterial World.* EcoSal Plus, 2016. **7**(1).
8. Olejniczak, M. and O.C. Uhlenbeck, *tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition.* Biochimie, 2006. **88**(8): p. 943-50.
9. Petrov, D.P., et al., *Opportunities and Challenges to Profile mRNA Modifications in Escherichia coli.* Chembiochem, 2022. **23**(18): p. e202200270.
10. Pratanwanich, P.N., et al., *Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore.* Nat Biotechnol, 2021. **39**(11): p. 1394-1402.
11. Jenjaroenpun, P., et al., *Decoding the epitranscriptional landscape from native RNA sequences.* Nucleic Acids Res, 2021. **49**(2): p. e7.
12. Liu, H., et al., *Accurate detection of m^6A RNA modifications in native RNA sequences.* Nature Communications, 2019. **10**(1).
13. Abebe, J.S., et al., *DRUMMER-rapid detection of RNA modifications through comparative nanopore sequencing.* Bioinformatics, 2022. **38**(11): p. 3113-3115.
14. Hendra, C., et al., *Detection of m^6A from direct RNA sequencing using a multiple instance learning framework.* Nat Methods, 2022. **19**(12): p. 1590-1598.

15. Gao, Y., et al., *Quantitative profiling of N6-methyladenosine at single-base resolution in stem-differentiating xylem of Populus trichocarpa using Nanopore direct RNA sequencing.* Genome Biol, 2021. **22**(1): p. 22.
16. Grünberger, F., S. Ferreira-Cerca, and D. Grohmann, *Nanopore sequencing of RNA and cDNA molecules in Escherichia coli.* RNA, 2021: p. rna.078937.121.
17. Fleming, A.M., et al., *Direct Nanopore Sequencing for the 17 RNA Modification Types in 36 Locations in the E. coli Ribosome Enables Monitoring of Stress-Dependent Changes.* ACS Chem Biol, 2023. **18**(10): p. 2211-2223.
18. Schaening-Burgos, C., G.-W. Li, and W. Gilbert, *RluA is the major mRNA pseudouridine synthase in Escherichia coli.* bioRxiv, 2023: p. 2023.12.08.570740.
19. Deng, X., et al., *Widespread occurrence of N6-methyladenosine in bacterial mRNA.* Nucleic Acids Research, 2015. **43**(13): p. 6557-6567.
20. Meyer, K.D. and S.R. Jaffrey, *Rethinking m^6A readers, writers, and erasers.* Annual Review of Cell and Developmental Biology, 2017. **33**(1): p. 319-342.
21. Choi, J., et al., *N6-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics.* Nat Struct Mol Biol, 2016. **23**(2): p. 110-5.
22. Ieong, K.W., et al., *N 6-Methyladenosines in mRNAs reduce the accuracy of codon reading by transfer RNAs and peptide release factors.* Nucleic Acids Res, 2021. **49**(5): p. 2684-2699.
23. Jain, S., et al., *Modulation of translational decoding by m^6A modification of mRNA.* Nat Commun, 2023. **14**(1): p. 4784.
24. Lassak, J., et al., *ArcS, the cognate sensor kinase in an atypical Arc system of Shewanella oneidensis MR-1.* Appl Environ Microbiol, 2010. **76**(10): p. 3263-74.
25. Heermann, R., T. Zeppenfeld, and K. Jung, *Simple generation of site-directed point mutations in the Escherichia coli chromosome using Red(R)/ET(R) Recombination.* Microb Cell Fact, 2008. **7**: p. 14.
26. Sambrook, J. and D.W. Russell, *Purification of nucleic acids by extraction with phenol:chloroform.* CSH Protoc, 2006. **2006**(1).
27. Meynier, V., et al., *Synthesis of RNA-cofactor conjugates and structural exploration of RNA recognition by an m^6A RNA methyltransferase.* Nucleic Acids Res, 2022. **50**(10): p. 5793-5806.
28. Punekar, A.S., et al., *Structural and functional insights into the molecular mechanism of rRNA m^6A methyltransferase RlmJ.* Nucleic Acids Research, 2013. **41**(20): p. 9537-9548.
29. Xiao, Y., et al., *An Elongation- and Ligation-Based qPCR Amplification Method for the Radiolabeling-Free Detection of Locus-Specific N^6-Methyladenosine Modification.* Angewandte Chemie International Edition, 2018. **57**(49): p. 15995-16000.
30. Tavakoli, S., et al., *Semi-quantitative detection of pseudouridine modifications and type I/II hypermodifications in human mRNAs using direct long-read sequencing.* Nat Commun, 2023. **14**(1): p. 334.
31. McCormick, C.A., et al., *Multicellular, IVT-derived, unmodified human transcriptome for nanopore direct RNA analysis.* bioRxiv, 2023: p. 2023.04.06.535889.
32. Salgado, H., et al., *RegulonDB v12.0: a comprehensive resource of transcriptional regulation in E. coli K-12.* Nucleic Acids Res, 2024. **52**(D1): p. D255-D264.
33. Danecek, P., et al., *Twelve years of SAMtools and BCFtools.* Gigascience, 2021. **10**(2).
34. Patro, R., et al., *Salmon provides fast and bias-aware quantification of transcript expression.* Nat Methods, 2017. **14**(4): p. 417-419.
35. Loman, N.J., J. Quick, and J.T. Simpson, *A complete bacterial genome assembled de novo using only nanopore sequencing data.* Nat Methods, 2015. **12**(8): p. 733-5.
36. Acera Mateos, P., et al., *Prediction of m^6A and m^5C at single-molecule resolution reveals a transcriptome-wide co-occurrence of RNA modifications.* Nat Commun, 2024. **15**(1): p. 3899.
37. Leger, A., et al., *RNA modifications detection by comparative Nanopore direct RNA sequencing.* Nat Commun, 2021. **12**(1): p. 7198.
38. Wickham, H., *ggplot2.* Use R!, 2016.

39. de Jong, A., O.P. Kuipers, and J. Kok, *FUNAGE-Pro: comprehensive web server for gene set enrichment analysis of prokaryotes.* Nucleic Acids Res, 2022. **50**(W1): p. W330-W336.
40. Tierrafria, V.H., et al., *RegulonDB 11.0: Comprehensive high-throughput datasets on transcriptional regulation in Escherichia coli K-12.* Microb Genom, 2022. **8**(5).
41. Schumacher, K., et al., *Ribosome profiling reveals the fine-tuned response of Escherichia coli to mild and severe acid stress.* mSystems, 2023. **8**(6): p. e0103723.
42. Giudice, E., K. Mace, and R. Gillet, *Trans-translation exposed: understanding the structures and functions of tmRNA-SmpB.* Front Microbiol, 2014. **5**: p. 113.
43. Felden, B., et al., *Presence and location of modified nucleotides in Escherichia coli tmRNA: structural mimicry with tRNA acceptor branches.* EMBO J, 1998. **17**(11): p. 3188-96.
44. Edelheit, S., et al., *Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m^5C within archaeal mRNAs.* PLoS Genet, 2013. **9**(6): p. e1003602.
45. Sergiev, P.V., et al., *The ybiN gene of Escherichia coli encodes adenine-N6 methyltransferase specific for modification of A1618 of 23 S ribosomal RNA, a methylated residue located close to the ribosomal exit tunnel.* J Mol Biol, 2008. **375**(1): p. 291-300.
46. Golovina, A.Y., et al., *The yfiC gene of E. coli encodes an adenine-N6 methyltransferase that specifically modifies A37 of tRNA1Val(cmo^5UAC).* RNA, 2009. **15**(6): p. 1134-41.
47. Ruszkowska, A., et al., *Structural insights into the RNA methyltransferase domain of METTL16.* Scientific Reports, 2018. **8**.
48. Pendleton, K.E., et al., *The U6 snRNA m^6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention.* Cell, 2017. **169**(5): p. 824-835 e14.
49. Zhong, Z.D., et al., *Systematic comparison of tools used for m^6A mapping from nanopore direct RNA sequencing.* Nat Commun, 2023. **14**(1): p. 1906.
50. Cappannini, A., et al., *MODOMICS: a database of RNA modifications and related information. 2023 update.* Nucleic Acids Res, 2024. **52**(D1): p. D239-D244.
51. Wahl, A., C. Huptas, and K. Neuhaus, *Comparison of rRNA depletion methods for efficient bacterial mRNA sequencing.* Scientific Reports, 2022. **12**(1).
52. Yamagami, R., et al., *Genome-wide analysis of the in vivo tRNA structurome reveals RNA structural and modification dynamics under heat stress.* Proc Natl Acad Sci U S A, 2022. **119**(25): p. e2201237119.
53. Wang, J., et al., *Accurate characterization of Escherichia coli tRNA modifications with a simple method of deep-sequencing library preparation.* RNA Biol, 2021. **18**(1): p. 33-46.
54. Baquero-Perez, B., et al., *N6-methyladenosine modification is not a general trait of viral RNA genomes.* Nat Commun, 2024. **15**(1): p. 1964.
55. Gokhale, N.S., et al., *N6-Methyladenosine in Flaviviridae Viral RNA Genomes Regulates Infection.* Cell Host Microbe, 2016. **20**(5): p. 654-665.
56. Lichinchi, G., et al., *Dynamics of Human and Viral RNA Methylation during Zika Virus Infection.* Cell Host Microbe, 2016. **20**(5): p. 666-673.

Figures

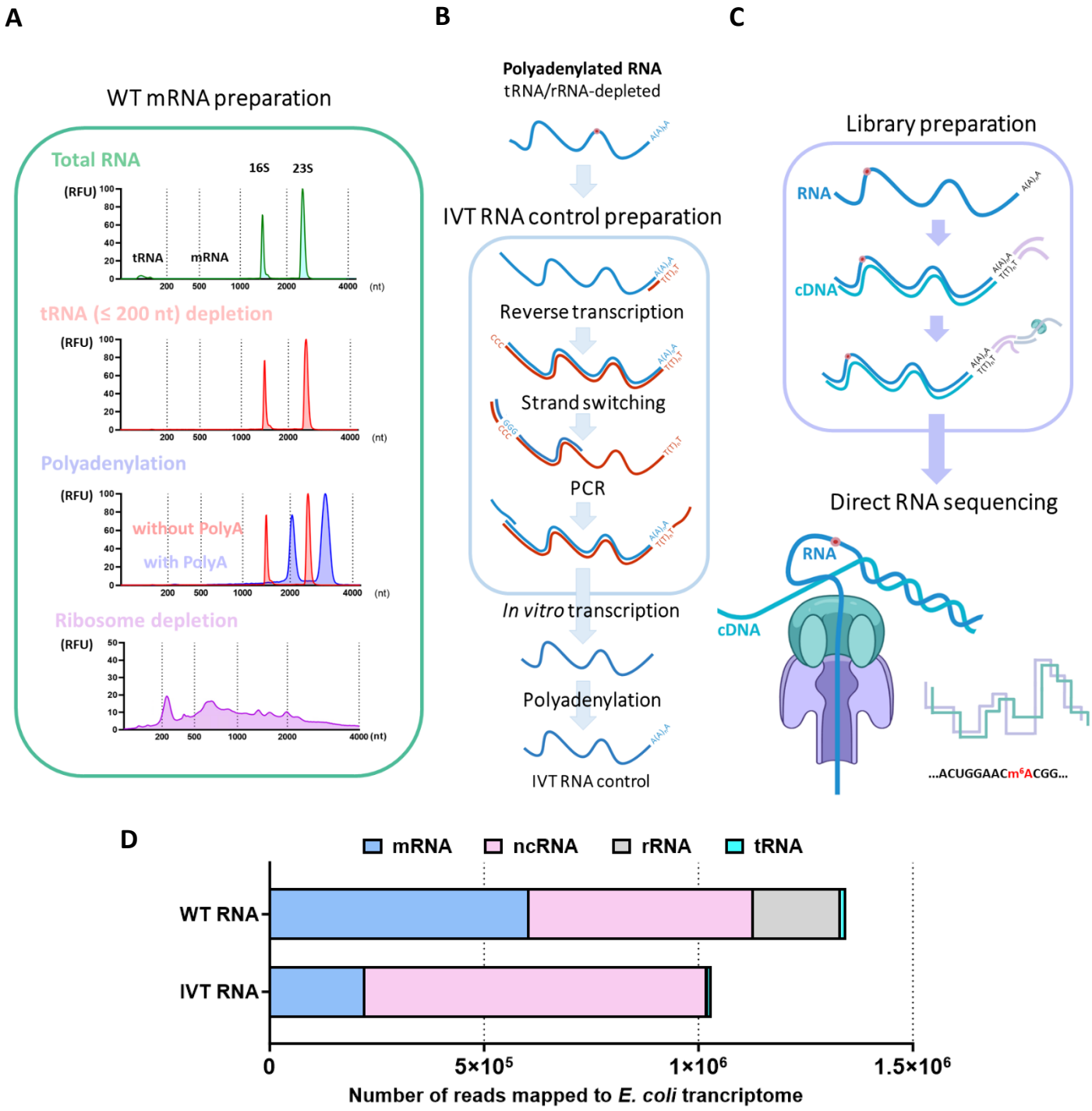


Figure 1. Graphical illustration of the protocol for direct RNA sequencing (DRS) of *Escherichia coli*. (A) Electropherograms from each step of the enriched mRNA preparation. **(B)** Preparation of the *E. coli in vitro* transcribed (IVT) RNA control. **(C)** Library preparation and Nanopore sequencing protocol. Red dots represent RNA modifications. **(D)** Total number of reads in the wild-type (WT) and IVT RNA control samples. Color indicates RNA biotype.

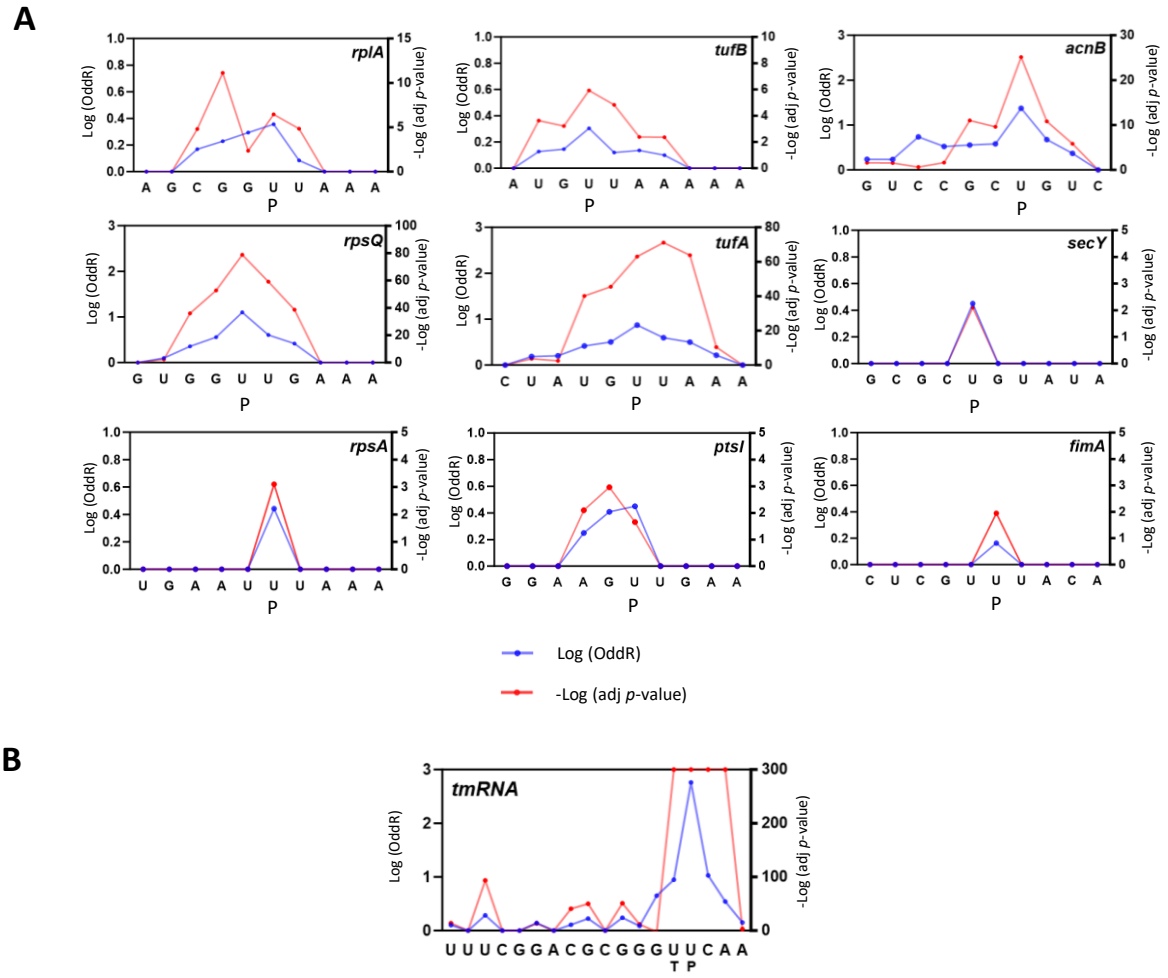


Figure 2. Detection of known messenger RNA (mRNA) and transfer-messenger RNA (tmRNA) modifications. (A, B) Log-transformed odds ratio and negative log-transformed adjusted *p*-values for **(A)** the nine known pseudouridine (P) positions previously identified in *E. coli* mRNA (Schaening-Burgos et al., 2023) and **(B)** positions 328–345 of the *E. coli* tmRNA *ssrA*. T, 5-methyluridine (m^5U).

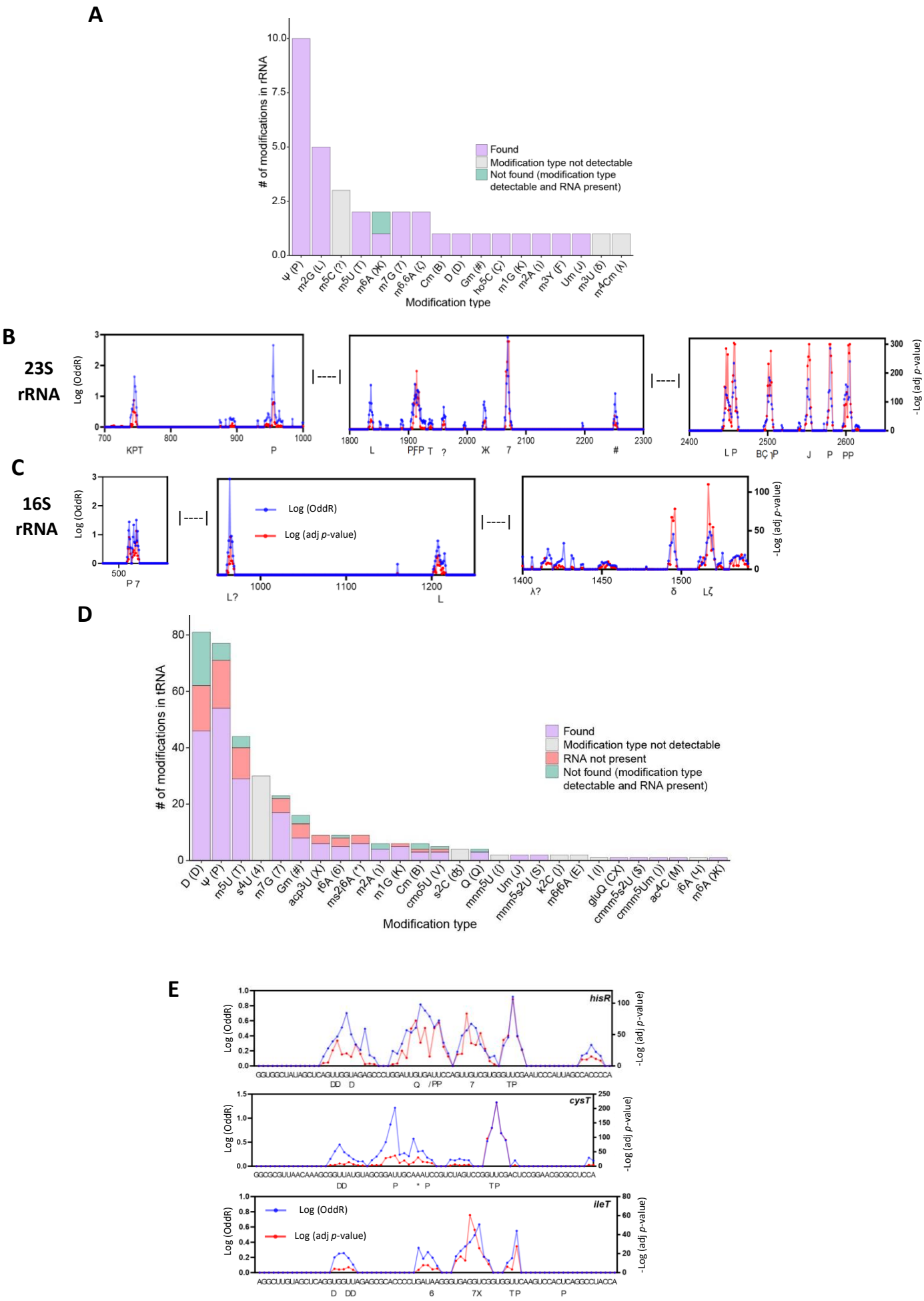

A
# of modifications in rRNA
Found
Modification type not detectable
Not found (modification type detectable and RNA present)
Modification type
B
23S rRNA
Log (OddR)
-Log (adj p-value)
C
16S rRNA
Log (OddR)
Log (adj p-value)
D
# of modifications in tRNA
Found
Modification type not detectable
RNA not present
Not found (modification type detectable and RNA present)
Modification type
E
hisR
cysT
ileT
Log (OddR)
Log (adj p-value)
-Log (adj p-value)

Figure 3. Detection of known transfer RNA (tRNA) and ribosomal RNA (rRNA) modifications. WT *E. coli* RNA and the IVT RNA control were processed with DRS and the resulting data were compared using ELIGOS. Putative modification sites were compared to the set of all known modifications in the *E. coli* tRNA and rRNA as published in MODOMICS (Cappannini et al., 2024). **(A)** Detection of known rRNA modifications by biotype. Light grey indicates a modification type that is not known to be detectable with ELIGOS; pink indicates a known modification that was detectable with ELIGOS and within an RNA that was present at a sufficient sequencing depth in our analysis but was not detected. **(B, C)** Negative log-transformed adjusted *p*-value (red) and log-transformed odds ratio (blue) values across the **(B)** 23S and **(C)** 16S rRNAs. Modification types are shown below the graph. **(D)** Detection of known tRNA modifications by biotype. Light grey indicates a modification type that is not known to be detectable with ELIGOS; pink indicates a modification in an RNA that was not present at a sufficient sequencing depth in our analysis; and teal indicates a known modification of a type that was detectable with ELIGOS and within an RNA that was present at a sufficient sequencing depth in our analysis but was not detected. **(E)** Negative log-transformed adjusted *p*-value (red) and log-transformed odds ratio (blue) values across positions in *hisR*, *cysT*, and *ileT.* Each point represents a specific nucleotide identified with ELIGOS. Modification types are shown below the graph. D, dihydrouridine; *, 2-methylthio-N6-isopentenyladenosine (ms^2i^6A); /, 2-methyladenosine (m^2A); K, 1-methylguanosine (m^1G); 6, N6-threonylcarbamoyladenosine (t^6A); P, pseudouridine (Ψ); T, 5-methyluridine (m^5U); L, N2-methylguanosine (m^2G); Ƒ, 3-methylpseudouridine (m^3Y); ?, 5-methylcytidine (m^5C); Ж, N6-methyladenosine (m^6A); 7, 7-methylguanosine (m^7G); #, 2'-O-methylguanosine (Gm); B, 2'O-methylcytidine (Cm); Ç, 5-hydroxycytidine (ho^5C); ı, 2-methyladenosine (m^2A); J, 2'-O-methyluridine (Um); λ, N4,2'-O-dimethylcytidine (m^4Cm); δ, 3-methyluridine (m^3U); ζ, N6,N6-dimethyladenosine ($m^{6,6}A$); X, 3-(3-amino-3-carboxypropyl)uridine (acp3U); Q, queuosine (Q); X, 3-(3-amino-3-carboxypropyl)uridine (acp^3U) (Cappannini et al., 2024).

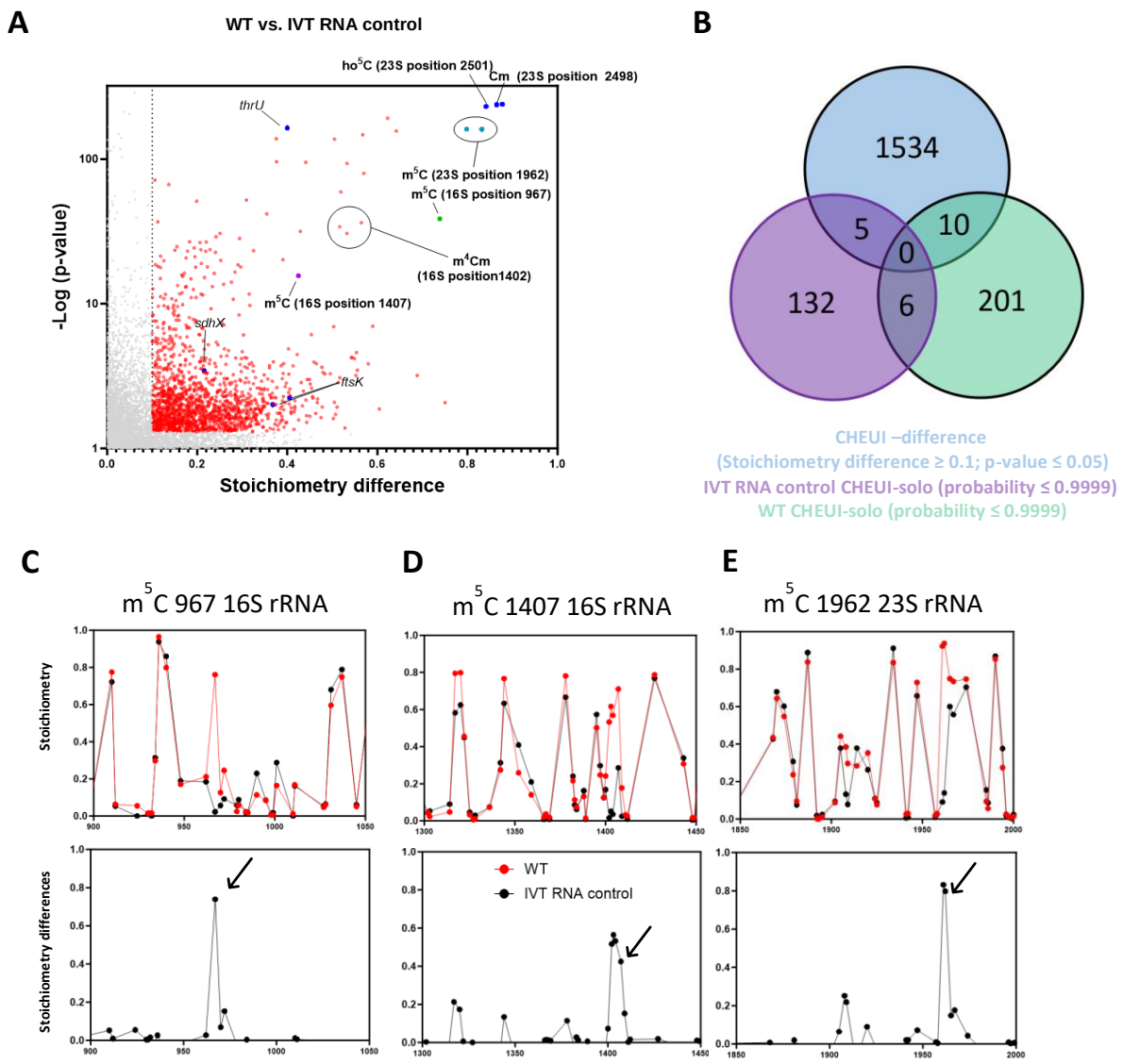


Figure 4. Detection of m^{5}C in *E. coli* RNA. (A, B) WT *E. coli* RNA and IVT RNA control samples were processed with DRS and the resulting data were analyzed using CHEUI-diff for m^{5}C. Each sample was also analyzed using CHEUI-solo. **(A)** CHEUI-diff data for m^{5}C in a comparison of WT *E. coli* RNA with the IVT RNA control. Each point represents a 9-mer. Red points correspond to 9-mers with significant differences in stoichiometry at a threshold of ≥ 0.1 and $p \leq 0.05$; blue points represent positions that were identified with CHEUI-solo at a probability ≥ 0.9999 and with CHEUI-diff at a stoichiometry difference ≥ 0.1 and $p \leq 0.05$. Points representing 9-mers corresponding to m^{5}C in the 967 or 1407 position of the 16S rRNA or m^{5}C in position 1962 of the 23S rRNA are indicated in the graph. **(B)** Total numbers of 9-mers identified with CHEUI-diff in the comparison of WT *E. coli* to the IVT RNA control, with CHEUI-solo in the IVT RNA control, and with CHEUI-solo in WT *E. coli* RNA. **(C–E)** Upper panels, CHEUI-diff stoichiometry values for m^{5}C in the comparison of WT *E. coli* with IVT RNA at each 9-mer in the 16S and 23S rRNA transcripts. Lower panels, stoichiometric differences at each position in the 16S and 23S rRNA transcripts. Graphs are shown for a region of 150 nt around three known modified positions: **(C)** m^{5}C at position 967 in the 16S rRNA, **(D)** m^{5}C at position 1407 in the 16S, and **(E)** m^{5}C at position 1962 in the 23S rRNA. Arrows indicate 9-mers containing known modified nucleotides.

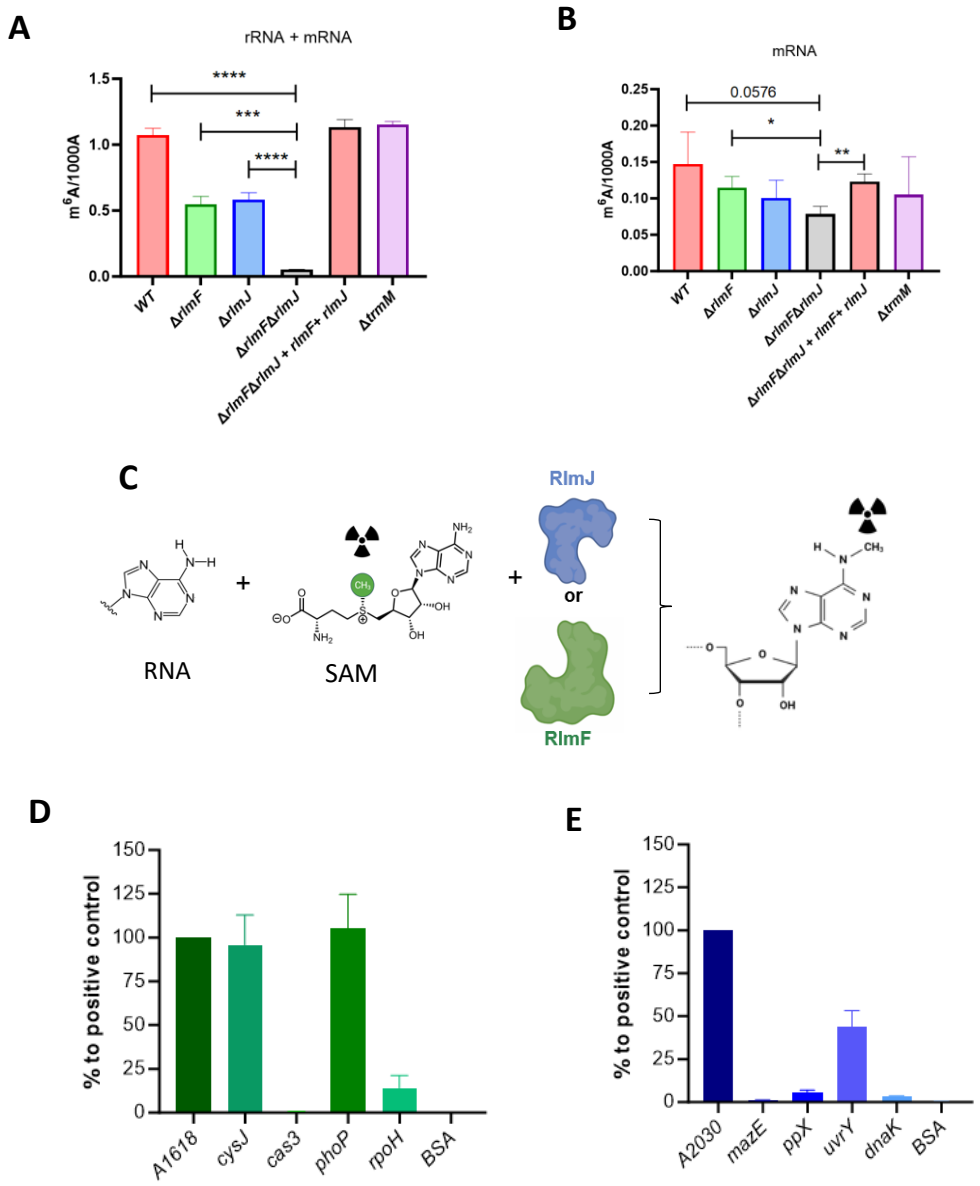


Figure 5. Detection of m^{6}A in *E. coli*. (A, B) Total RNA samples were purified from WT *E. coli*; the single mutants Δ*rlmF,* Δ*rlmJ,* and Δ*trmM*; the Δ*rlmF*Δ*rlmJ* double mutant; and the Δ*rlmF*Δ*rlmJ + rlmF + rlmJ* complementation strain. tRNA was depleted from total RNA to form rRNA + mRNA samples; tRNA and rRNA were depleted from total RNA to produce mRNA-enriched samples (mRNA). **(A, B)** m^{6}A levels were measured via mass spectrometry in **(A)** rRNA + mRNA (tRNA-depleted) samples and **(B)** mRNA-enriched (tRNA/rRNA-depleted) samples. Data are presented as the number of m^{6}As per 1000 non-modified As (1000A). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$ (unpaired Student's *t*-test). **(C)** Graphical illustration of the *in vitro* methylation protocol. **(D, E)** Methyltransferase activity of purified **(D)** RlmF and **(E)** RlmJ. IVT RNAs of 100–300 nt were generated for *cysJ, cas3, phoP, rpoH, mazE, ppX, uvrY*, and *dnaK*, each of which contained an expected modification site; 30-nt RNAs containing As in the appropriate sequence context were also generated for the 23S rRNA positions 2030 and 1618.

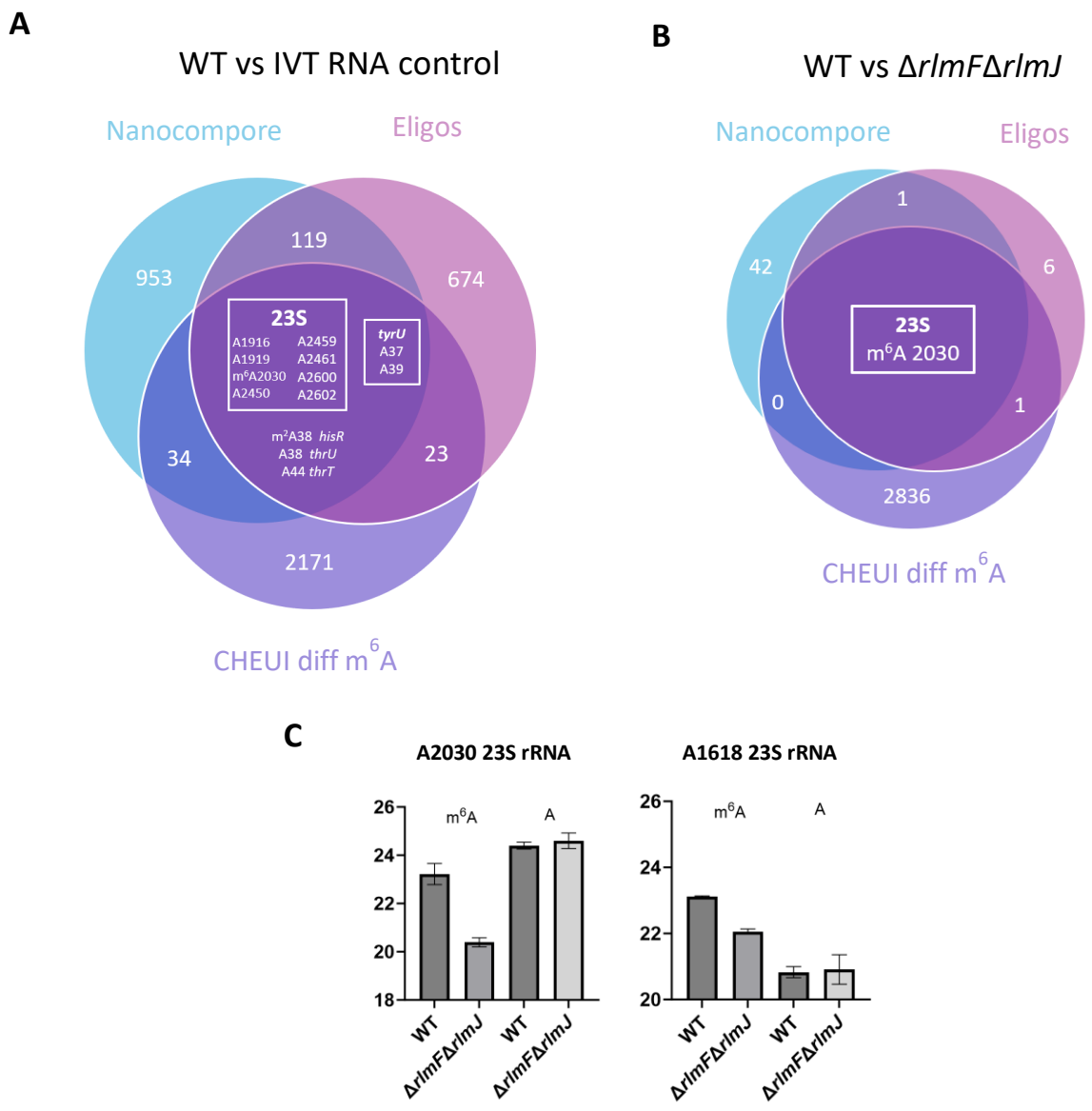


Figure 6. Detection of m^{6}A in *E. coli* using DRS. **(A, B)** Unique and overlapping putative modification sites identified in WT *E. coli* compared to **(A)** IVT RNA control and **(B)** and Δ*rlmF*Δ*rlmJ* samples using Nanocompore (blue), ELIGOS (pink), and CHEUI-diff for m^{6}A (purple). **(C)** Cycle threshold (C_t) values from quantitative PCR (qPCR) for detection of m^{6}A in WT and Δ*rlmF*Δ*rlmJ E. coli* with SELECT. Positions expected to show an m^{6}A modification are denoted "m^{6}A"; nearby As that were not expected to be modified are marked with "A". Data are presented as the mean ± standard deviation from two biological and two technical replicates.

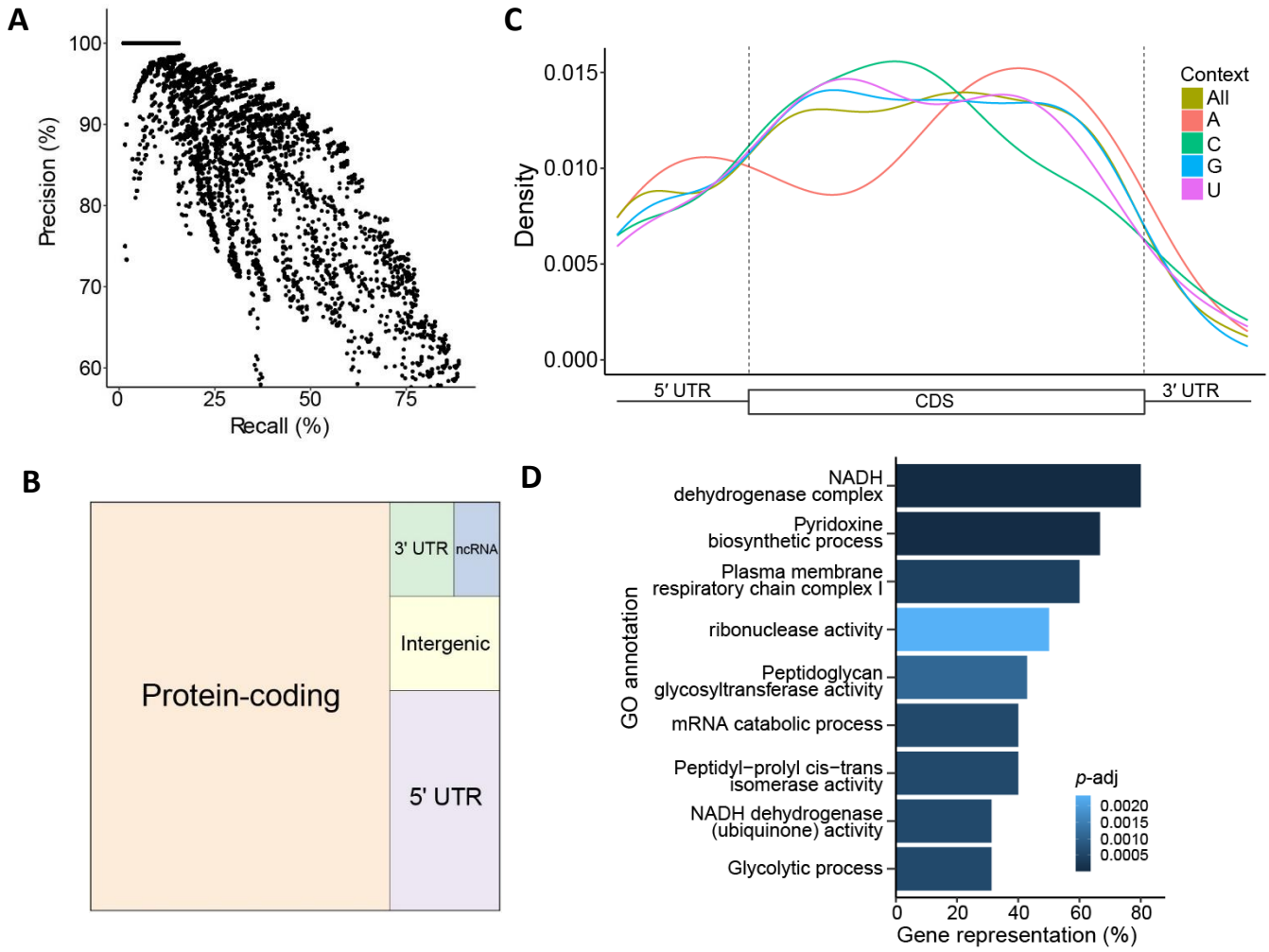


Figure 7. Characteristics and distribution of putative novel modification sites throughout the *E. coli* transcriptome. (A) Analysis of precision and recall in tRNA/rRNA sites called as modified with ELIGOS. Precision and recall were calculated for 32,000 combinations of four parameters: the percent Error of Specific Bases (%ESB) in the IVT control (ESB control), %ESB in the WT *E. coli* RNA sample (ESB test), the odds ratio (ESB test/ESB control), and the adjusted *p*-value. Precision was calculated as the percentage of putative modification sites corresponding to known modifications or within a ± 10-nt signal region surrounding a known modified site. Recall was calculated as the percentage of known tRNA/rRNA modification sites that were successfully detected. **(B)** Relative proportions of putative novel *E. coli* modification sites by genomic position, excluding tRNA and rRNA. A total of 387 sites were detected outside the tRNA and rRNA. **(C)** Normalized distribution of putative mRNA modifications in each nt across the 5′ untranslated region (UTR), coding sequence (CDS), and 3′ UTR. **(D)** Gene Ontology (GO) enrichment analysis of all genes with putative mRNA modifications. Adjusted *p*-values (*p*-adj) were derived from a hypergeometric probability test followed by false discovery rate (FDR) correction for multiple testing.