Sequencing-based quantitative mapping of the cellular small RNA landscape

Jennifer F. Hu[1], Daniel Yim[2], Sabrina M. Huber[2], Jo Marie Bacusmo[3], Duanduan Ma[4], Michael S. DeMott[2,6], Stuart S. Levine[4], Valerie de Crécy-Lagard[3], Peter C. Dedon[2,5,6*], Bo Cao[2,6,7*]

[1] Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA.

[2] Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA.

[3] Department of Microbiology & Cell Science, University of Florida, Gainsville, FL, USA.

[4] BioMicro Center, Massachusetts Institute of Technology, Cambridge, MA, USA.

[5] Singapore-MIT Alliance for Research and Technology Antimicrobial Resistance IRG, Singapore

[6] Center for Environ'l Health Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA.

[7] College of Life Sciences, Qufu Normal University, Qufu, Shandong 273165, China.

* Corresponding authors: B.C. (caobo@qfnu.edu.cn) (ORCID ID: 0000-0002-7011-1676) and P.C.D. (pcdedon@mit.edu) (ORCID ID: 0000-0003-0011-3067)

Abstract

Current next-generation RNA sequencing methods cannot provide accurate quantification of the population of small RNAs within a sample due to strong sequence-dependent biases in capture, ligation, and amplification during library preparation. We report the development of an RNA sequencing method – AQRNA-seq – that minimizes biases and enables absolute quantification of all small RNA species in a sample mixture. Validation of AQRNA-seq library preparation and data mining algorithms using a 963-member microRNA reference library, RNA oligonucleotide standards of varying lengths, and northern blots demonstrated a direct, linear correlation between sequencing read count and RNA abundance. Application of AQRNA-seq to bacterial tRNA pools, a traditionally hard-to-sequence class of RNAs, revealed 80-fold variation in tRNA isoacceptor copy numbers, patterns of site-specific tRNA fragmentation caused by stress, and

quantitative maps of ribonucleoside modifications, all in a single AQRNA-seq experiment. AQRNA-seq thus provides a means to quantitatively map the small RNA landscape in cells and tissues.

Next generation RNA sequencing platforms (RNA-seq) have advanced functional genomics by facilitating discoveries of new RNA species, gene annotations, and quantitative analysis of gene expression.[1, 2] RNA-seq has played a central role in unravelling the complexity of RNA function by facilitating quantitative profiling of the transcriptome as a function of cell state. Although a variety of RNA-seq methods provide precise and accurate analysis of changes in transcript abundance *between* samples or conditions, their major drawback is the inability to accurately quantify small RNA species *within* a sample. This is partly rooted in biased ligation of sequencing linkers to the 3'- and 5'-ends of RNAs, which varies by 10^3-fold depending upon on the identity of the terminal nucleotides[3-7] and causes 10^6-fold artifacts in sequencing read counts.[5, 8, 9] Highly structured and heavily modified RNA molecules, such as tRNAs, further challenge the quantitative accuracy of RNA-seq.[10, 11] These structural features cause polymerase fall-off during cDNA synthesis, which prevents detection of the transcripts.

A variety of specialized RNA-seq methods, many limited specifically to either miRNA or tRNA only,[12-14] attempt to minimize ligation bias using linkers with randomized terminal nucleotides and molecular crowding agents to enhance ligation[4] and to reduce polymerase fall-off during reverse transcription (RT) with two-step ligation[8] and removal of some methyl modifications with AlkB.[15-17] Although polymerase fall-off effects are minimized, residual ligation biases persist and lead to "jackpot" sequences.[8] For example, the template-switching reverse transcriptase TGIRT used in DM-tRNA-seq[16] is biased by the identity of the overhanging nucleotide in the adapter strand.[18]

The problem with existing RNA-seq techniques is that none have been systematically engineered to optimize ligation and amplification efficiencies or validated for quantitative accuracy and lack of bias artifacts. Furthermore, none are broadly applicable to all RNA species.

Here we describe AQRNA-seq, an RNA-seq method that enables absolute quantification of all RNA species in a sample by providing a direct, linear correlation between sequencing read count and RNA abundance. The library preparation and data mining algorithms have been rigorously validated by multiple orthogonal approaches. Application of AQRNA-seq to the dormancy response of *Mycobacterium bovis* BCG revealed large variations in tRNA copy numbers, tRNA fragmentation, and tRNA modification location and abundance within and among samples along a time course of stress-induced mycobacterial persistence.

Results

AQRNA-seq design and optimization

The AQRNA-seq workflow (**Figure 1a**) maximizes ligation capture of RNAs using novel adapters (linkers) and minimizes RT fall-off with two-step linker ligation and AlkB treatment. Adapter ligation ("linker 1") begins at the 3'-end, with two randomized nucleotides incorporated at the 5' end of linker 1 to maximize ligation efficiency based on known strong biases in T4 ligase activity.[19] Linker 1 is also a DNA oligo to facilitate removal of unligated linker with RecJ, a single-stranded-DNA-specific 5'→3' exonuclease, which leaves the hybrid RNA-DNA product intact.[20] A 50:1 excess of linker 1 resulted in >90% ligation efficiency (**Fig. 1b**).

The ligated RNA can then be treated with AlkB to reduce levels of RT-blocking methyl modifications.[15] Though this step is not essential for capturing all RNA sequences in a sample, it can provide information about the identities of polymerase-interfering modifications. Optimized buffer conditions and AlkB concentration allowed removal of m^1A (90%), m^1G (48%), and m^1I (96%). Only 12% of m^3C was removed, which is contrary to previous observations[16], and $m^{2,2}G$ was reduced by only 35%, apparently demethylated to AlkB-resistant m^2G (**Fig. 1c**).[15] After demethylation, unligated linker 1 is removed with RecJ. Optimized conditions resulted in >99% linker removal (**Fig. 1b**), obviating the need for HPLC purification of the RNA-DNA product.

RT is now accomplished with a DNA primer complementary to the 3' DNA linker 1 and the resulting cDNA is ligated at its 3'-end to a custom DNA adapter (linker 2) using T4 DNA ligase. Linker 2 possesses a hairpin, a random hexamer sequence, and a downstream primer binding site for subsequent amplification. The random hexamer in linker 2 acts as a splint to enhance ligation with the cDNA (**Fig. 1d**), which was optimized to nearly 100% with a 50:1 linker excess (**Fig. 1d**). Excess linker 2 is removed with RecJ and PCR amplification was performed with primers that complement linker 2 and incorporate standard Illumina anchor and barcode sequences for subsequent sequencing.

AQRNA-seq data processing workflow

Rigorous processing of the RNA sequencing data represents another step critical for quantitative accuracy. To this end, we developed a customized workflow for processing AQRNA-seq sequencing data, illustrating the main points here for bacterial tRNA sequences (**Fig. 1e**). Similar customization is necessary for other RNA species and for tRNAs from higher eukaryotes. Building on several commonly used software packages for traditional RNA-seq analyses (see Methods), workflow customization begins with curation of the reference sequences used for aligning sequencing reads. Here we focus on bacterial tRNAs, with considerations applying to all transcripts in any organism. Reference sequences are first culled of duplicate genes and pseudogenes that lead to ambiguous assignment of reads. Similar consideration must be given to post-transcriptional processing, such as trimming and processing of 5'- and 3'-termini of primary tRNA transcripts as well as tRNA modifications.[21] For mycobacteria, the 3'-CCA of each tRNA is variably genomically encoded or added post-transcriptionally.[22]

The resulting non-redundant reference library is then used to align forward and reverse sequencing reads, first separating uniquely aligned reads from reads matching multiple sequences in the library. Even with careful library curation, high sequence similarity or insufficient read length can result in multiply-mapped reads. We arbitrarily set a 10 nt read length

filter to maximize alignments. For closely-related reference sequences, multiply-mapped reads can be resolved by collapsing read assignments to broader "families" (**Fig. 1e**), with subsequent determination of the proportion of multiply-mapped reads and the cause of multiple mapping. These considerations rationalize a decision to discard, average, or sum the read counts from multiple, closely-related reference sequences.

Finally, the read count for full-length sequences and fragments is tabulated from the curated set of mapped reads. Data are normalized either to the total number of reads in each barcoded sample or to an internal RNA standard to account for sample-to-sample variation in input RNA as well variation in sample pooling prior to sequencing. For tRNAs, the resulting dataset can then be visualized using a graphical presentation that reveals abundance, fragmentation sites, and modification sites in a single view, as illustrated shortly for stress-induced bacterial persistence (**Figs. 3-7**).

Method validation shows a linear relationship between read count and RNA abundance

Three different approaches were used to test the precision and accuracy of AQRNA-seq. First, five synthetic RNA oligos with varying lengths (25-80 nt) were mixed at varying molar ratios and used as input RNA for library preparation. After sequencing, we found that read counts for each oligo varied directly and linearly with input RNA abundance (r^2 0.92-1; **Fig. 2a**), with an averaged sequencing response (slope) of ~300 ± 50 reads per fmol of input RNA. This demonstrates minimal sequencing bias for quantity or length of input RNA.

To test for sequence-dependent ligase or polymerase biases, we used AQRNA-seq to prepare libraries from the Miltenyi miRXplore Universal Reference consisting of 963 equimolar miRNA sequences (sourced from miRBase[23]), which ranged in length (16-28 nt) and possessed all 16 possible dinucleotide combinations at 3' and 5' termini. Resulting sequencing reads for each miRNA were counted and divided by total counts for all detected miRNAs to obtain a "normalized read count". Total counts were also divided by 963 (the number of detected miRNAs)

to obtain the “expected read count”, assuming all species are equimolar. A read ratio was then calculated by dividing “normalized” by “expected”. A plot of all 963 read ratios ranked from lowest to highest showed that ~75% fell within two-fold variation of expected abundance (**Fig. 2b**). The number of jackpot and dropout miRNAs (those with normalized read ratios >10-fold higher or lower than expected) was <3% of the total miRNA mixture. No other library preparation method resulted in greater than 40-50% of miRNAs falling within the two-fold threshold.[24] Based on this comparison and additional reports using the miRXplore reference,[5, 18, 25] AQRNA-seq is the most quantitatively accurate RNA sequencing workflow currently available.

The third validation study tested AQRNA-seq performance against the analysis of the *E. coli* K12 tRNA pool by two-dimensional gel electrophoresis and Northern blotting by Dong et al.[26] Applying AQRNA-seq to small RNAs (<200 nt) from *E. coli* K12 strain BW25113 (derived from W1485 used by Dong et al.[26]), the total expressed levels of 45 tRNAs (summed read counts for full-length and truncated sequences) were compared to the 46 tRNAs identified by Dong et al.[26] Excluding one outlier, there was strong agreement ($r^2 = 0.81$) between the two orthogonal approaches (**Fig. 2d**).

tRNA landscape dynamics in stress-induced bacterial persistence

AQRNA-seq was now applied to quantify the dynamics of a particularly challenging set of targets: the 45 tRNAs in the *Mycobacterium bovis* BCG model for stress-induced, non-replicative, antibiotic-resistant state of persistence in tuberculosis.[27-29] Total small RNA (<200 nt) was isolated along the time course of BCG persistence caused by complete nutrient deprivation (**Fig. 3a,b**), with ~1% of the bacterial population surviving as persisters after 20 days in PBS and restoration of growth with transfer to nutrient-rich medium (**Fig. 3a**). Processing of the AQRNA-seq read data for one BCG starvation study illustrates the workflow points made earlier. In this highly-multiplexed sequencing experiment involving triplicate samples from each of 5 time points during starvation and resuscitation, an average of 5 million raw sequencing reads was

obtained for each sample (**Fig. 3c**). After size-selection and adapter trimming, the majority (75%) of the remaining reads consisted of uniquely-mapped, paired reads when mapped to the full set of mature BCG RNA sequences (**Fig. 3c**). Of these, another 75% mapped to an 80-nt internal standard that was added in large excess in this experiment, while remaining reads mapped to reference library tRNA sequences (**Fig. 3c**). A much lower level of internal standard allows detection of rare RNA species, such as tRNA fragments illustrated in **Figure 4e** as discussed shortly. To account for variation introduced by input RNA and sample processing, reads originating from a single sample can be normalized to a spiked-in standard (80 nt here). Comparison between samples is facilitated by expressing RNA read counts as either a percentage of total aligned reads or total aligned tRNA reads within each sample. For example, the abundance of a single isoacceptor can be expressed as a proportion of the complete isoacceptor pool in a sample and this value can be tracked and compared across samples derived from various treatment conditions. This normalization process applies to any RNA species of interest.

The resulting BCG dataset was mined for information about starvation-induced changes in tRNA expression, tRNA fragmentation patterns, and the locations of modified nucleosides in individual tRNAs. These tRNA features are best visualized graphically in horizontally stacked alignment plots (**Fig. 4a**), in which the start and end positions of each sequencing read are aligned along an X-axis annotated using the Sprinzl coordinate system for tRNA structure (**Fig. 4b**).[30] This plot format readily informs about the state of each tRNA isoacceptor, with the 3' end of a read reflecting where the source RNA species was captured by linker 1 (**Fig. 4c**). The complete set of alignment plots for 45 unique, annotated tRNA species in BCG is shown in **Figure S.1.** As a first observation, the height off the stack is directly proportional to the total number of expressed transcripts. The graph can then be split into three sections. The bottommost portion consists of full-length reads – reads that span the entire tRNA sequence (**Fig. 4b**). These "type 3" reads originated from full-length tRNA molecules. The top section contains

reads that do not reach the 3' end of the tRNA ("type 1" reads; **Fig. 4b**) and correspond to 5' tRNA fragments, with the 3' linker ligated to the 5' end of the break site (**Fig. 4c**). The middle section contains reads that start at the 3' end of the tRNA but do not reach the 5' end ("type 2" reads; **Fig. 4b**). These reads correspond to tRNA molecules that were fragmented to leave behind a 3' portion, or to full-length tRNAs for which cDNA synthesis was prematurely truncated due to polymerase fall-off. Analysis of 3 tRNA isoacceptors by northern blotting using 5' and 3' probes (**Fig. S2**) suggests that the majority of the type 2 reads are in fact full-length tRNAs as the vast majority of bands detected using the 3' probe were full-length. This is consistent with the relatively low level of 5' tRNA fragments in stack plots for all of expressed tRNA molecules in BCG (**Fig. S1**).

Starvation-induced persistence remodels the tRNA landscape in BCG

AQRNA-seq provides a global view of changes in the RNA landscape caused by stress. In starved BCG, the abundance of individual tRNAs – here defined as the sum of all reads that aligned to a particular tRNA species – spanned a large range. In most samples, tRNA Lys-CTT and tRNA fMet-CAT were the most highly expressed tRNAs, together totaling ~20% of the tRNA pool during log growth, whereas tRNA Ser-GGA was ~80-times lower, comprising only 0.1-0.3% of the tRNA pool (**Figs. 5a, S1**). In the transition from rich medium to starvation over 20 days, the abundances of several tRNAs are significantly altered (**Fig. 5a,b; Fig. S1**). For example, tRNA Ser-CGA and tRNA His-GTG drop significantly in early starvation and rise again during late starvation and resuscitation. tRNA Leu-CAG and tRNA Thr-GGT show the opposite pattern, increasing during starvation and returning to S0 levels during resuscitation.

Starvation also induced significant shifts among the isoacceptors associated with a specific amino acid (**Fig. 5b**). In BCG, there are 7 amino acids for which there is a single isoacceptor (Trp, Asp, Asn, Cys, His, Phe, Tyr), 5 have 2 isoacceptors, and 8 have ≥3 tRNA isoacceptors. Illustrating the latter, Ser, Thr, and Leu are specified by 6, 4, and 6 synonymous

codons, respectively, and these codons are read by 4, 3, and 5 different tRNA species, respectively.[31] **Figure 5b** shows starvation-induced changes in fractional abundance of isoacceptors for these amino acids. Prior to starvation (S0), tRNA Ser-CGA, tRNA Ser-TGA and tRNA Ser-GCT comprise 36%, 34%, and 25% of the Ser isoacceptor pool, respectively. At 4 days of starvation (S4), the abundance of tRNA Ser-CGA drops to 7% while tRNA Ser-TGA and tRNA Ser-GCT surge to 47% and 44%, respectively. For Thr, tRNA Thr-CGT and tRNA Thr-GGT represent 51% and 31%, respectively, during log growth (S0), but flip to 40% and 47% at S10, and 40% and 48% at S20, before returning to pre-starvation levels during resuscitation. These data illustrate the dynamics of individual tRNAs resulting from stress-induced changes in gene expression or tRNA degradation.

AQRNA-seq also captures the dynamics of tRNA fragmentation and degradation, as occurs in tRNA maturation and quality control,[32-34] small RNA regulation of gene expression, and toxin-antitoxin systems.[35-40] It is difficult to differentiate 5'-degradation of full-length tRNA from polymerase fall-off as both generate a fragment that aligns at the 3' end of tRNA ("type 2" reads, **Fig. 4b**). However, 3'-degradation and endonucleolytic cleavage generate fragments with 3' ends positioned inside the reference sequence ("type 1" reads, **Fig. 4b**). An extreme case is illustrated by tRNA Glu-TTC. While ~80% of the reads aligned with the 3' end in log growth, ~80% of the reads at S4 had 3' ends located at position 58 in the TΨC arm of the tRNA (**Fig. 4b**). These tRNA fragments appeared as a gel band present in the S4 sample but much fainter in the log sample (**Fig. S3b**). The absence of a corresponding number of short (15-20 nt) tRNA fragments representing the 5'-side of the cleaved tRNA (type 2 reads, **Fig. 4e**) could result from degradation of the 3'-fragments. AQRNA-seq thus provides quantitative information about tRNA fragmentation in addition to total expressed tRNA copy numbers and, as discussed next, the locations of modified nucleosides and secondary structures in RNA.

Quantitative mapping of RNA modifications and secondary structure

Nearly all forms of RNA contain post-transcriptional modifications, with over 120 structures known.[41] tRNAs are particularly heavily modified to the extent of ~10% of the component nucleotides, on average.[10] In some cases, modifications interfere with the fidelity and processivity of the reverse transcriptase during RNA-seq library preparation, which can be exploited to map modification positions based the resulting sequence data. [15, 42, 43] AQRNA-seq detects RT defects as read pile ups at specific sites along the RNA sequence. We illustrate this behavior with small RNA isolated from log-growing *E. coli*, for which there is detailed information about the types and locations of tRNA modifications.[41] As shown in **Figure 6a**, several tRNAs had high levels of polymerase stops at positions 38 and 48. By overlaying the sequencing stop positions on tRNA modification maps for *E. coli* in the Modomics database,[41] we determined that these two positions are proximal to known tRNA modifications. One subset of 9 tRNAs had 31-83% of the mapped reads stop at position 48, which abuts the modification 3-(3-amino-3-carboxypropyl)uridine (acp^3U) at position 47 (orange boxes on right in **Fig. 6a**). With U-N^3 involved in base-pairing and N^3-methyluridine (m^3U) previously reported to block RT,[44] it makes sense that acp^3U would also prevent polymerase procession. tRNAs with NNA anticodons had 48-69% of the reads end at position 38, which is adjacent to N^6-isopentenyladenosine (i^6A) and its hypermodified derivatives (e.g., 2-methylthio-N^6-isopentenyladenosine, ms^2i^6A) in the anticodon loop (yellow boxes on left in **Fig. 6a**). Modifications involving the N^6 amine of adenosine are not known to block RT, but N^6-methyladenosine induces pausing.[44] It is also plausible that the bulky size of the isopentenyl modification combined with the sharp, three-dimensional turn of the anticodon hairpin interferes with polymerase activity.

The presence of modified nucleotides appears to block RT, leading to a read interruption 1-2 nucleotides away from the modification site. Further validation of this idea comes from a BCG library preparation lacking AlkB demethylase. In the absence of AlkB, up to 90% of the reads mapping to 23 of 45 of tRNA species were truncated at positions 59-60, as illustrated in

the heatmap of polymerase stops in **Figure 6b**. After AlkB treatment, most stops at positions 59-60 have disappeared and reads increased in length with a leftward shift toward the 5' end of the tRNA (**Fig. 6c**). The alignment profile of Glu-CTC illustrates the effect of AlkB on the read distribution along the tRNA sequence. The sharp "cliff" at position 60 in the untreated sample (**Fig. 6b**) corresponds to a predicted m^1A modification at position 58 (**Fig. 6b**). m^1A is a known substrate for AlkB (**Fig. 1c**) and blocks RT.[15, 43, 44] After AlkB treatment, the cliff disappears and the reads span the tRNA sequence (**Fig. 6c**). The presence of 5' cliffs in the alignment plots for nearly all tRNA species in BCG (**Fig. S.1**) points to the potential for quantitative mapping of RNA modifications by AQRNA-seq.

Modification mapping is also illustrated by the presence of signature mutations resulting from RT.[44] For example, inosine (I) is present at the wobble position of a single tRNA in mycobacteria[41] and tends to pair with C during RT.[42, 45] I is thus detected by a near stoichiometric T-to-C sequencing mutation, as illustrated in **Fig. 6d** for position 34 of BCG tRNA Arg-ACG. As reviewed extensively by Motorin and Helm, this kind of modification mapping could aid in the discovery of previously unannotated or unlocalized modifications in poorly characterized species. AQRNA-seq enables both modification mapping and mapping of cleavage sites while quantifying the landscape of the population of small RNA species in a single sample.

Discussion

Here we have developed, validated, and applied the first RNA sequencing method that provides precise and accurate absolute quantification of all small RNA species in a single sample. The inability of current RNA-seq methods to provide accurate information about the quantities of individual RNA molecules in a sample mixture has constrained our understanding of RNA regulation and RNA biology. Following a systematic deconstruction of the RNA-seq NGS library prep workflow, we identified several steps critical to the quantitative precision and accuracy of

RNA-seq results: (1) sequential ligation of the 5' linker after reverse transcription, (2) linker structures and biochemical conditions that provide >90% ligation efficiency, (3) a non-essential but informative AlkB demethylation step to reduce interference from polymerase-blocking alkyl modifications, and (4) a data mining workflow that minimizes loss of read information to improve quantitative accuracy. Within AQRNA-seq, adapter ligation steps have been rigorously optimized to maximize RNA capture, with ligation of the 5'-adapter scheduled after reverse transcription to ensure that all cDNA species, including truncated cDNAs, are captured and represented in the final library mixture. A non-essential AlkB demethylation step is employed to minimize premature cDNA truncation and to provide information about the identities and locations of polymerase-blocking or mutating modifications. Beyond the linear relationship between read count and RNA copy number, the RNA-seq method provides information about modification occupancy, tertiary structure, and fragmentation.

While our RNA-seq method was applied first to tRNA analysis, it is broadly applicable to any form of RNA. The validation against the miRNA library illustrates this general applicability. The 963 miRNAs in the Miltenyi miRXplore Universal Reference possess varying lengths (16-28 nt) and varying sequences (including all 16 possible 3'- and 5'-termini), which provides a robust test of the quantitative accuracy of the RNA-seq method. With the best-performing commercial kits able to quantify only 40-50% of the miRNAs within a two-fold range of accuracy, the fact that we were able to quantify ~75% of the miRNAs to within two-fold of expected abundance, without incidence of jackpot and a low incidence of dropouts, attests to the rigor of the RNA-seq method. The AQRNA-seq method is thus established to be the most accurate for application in miRNA profiling studies, with the recent miRNA method developed by Kim *et al*. lacking evidence of quantitative rigor.[14] However, while random priming RNA-seq methods provide relatively accurate quantification of mRNAs,[46] AQRNA-seq is also applicable to longer RNA species such as mRNA and rRNA as a means to map RNA modifications and cleavage sites, as illustrated in

Fig. S4. The addition of a Covaris fragmentation step after ligation of the first linker allows longer RNA species to be reduced to an acceptable length for library generation. The introduction of this step allows for quantification of (1) all expressed copies of an RNA (all molecules with a 3'-end that maps to the end of the transcribed or fully mature sequence), (2) polymerase fall-off due to modifications, secondary structure, or polymerase detachment, and (3) fragmentation sites within the RNA molecules (3'-ends that map within the full-length sequence).

Conclusion

Collectively, our results demonstrate a quantitatively accurate method for sequencing RNAs. We have rigorously validated the accuracy of the method to ensure ligation and amplification biases are minimized. In contrast to some currently available approaches, AQRNA-seq is generalizable to all classes of RNA – short noncoding RNAs such as microRNAs, highly-modified species such as tRNAs, and longer mRNAs, rRNAs, and other types of RNA.

Acknowledgments

We thank Sidney Vermeulen for technical assistance during method development and members of the Bartel Laboratory (Whitehead Institute and MIT Department of Biology), especially Asia Stefano and Prof. David Bartel, for assistance with northern blots. We thank Dr. Pavel Ivanov (Harvard Medical School) for sharing synthetic RNA standards. This work was supported by grants from the US National Science Foundation (MCB-1412379 to V.C.L.), the National Institutes of Health (ES002109), and the National Research Foundation of Singapore through the Singapore-MIT Alliance for Research and Technology Antimicrobial Resistance IRG. J.F.H. was supported by the MIT Toxicology Training Grant T32-ES007020 to J.F.H.), D.Y. by a postdoctoral fellowship from A*STAR, Singapore, and S.M.H. by a postdoctoral fellowship from the Swiss National Science Foundation.

Author Contributions

P.C.D., B.C., J.F.H. conceived the idea, designed the experiments, and wrote the first draft of the manuscript. P.C.D., B.C., D.Y. and J.F.H. developed the method and performed the sequencing experiments. J.F.H., B.C., D.M., and S.S.L. implemented and interpreted the computational analyses. D.Y. and J.F.H. performed mycobacterial culturing and RNA isolation. J.M.B. performed *E. coli* culturing and RNA isolation. S.M.H. optimized experimental conditions and characterized demethylation efficiencies by LC-MS. S.M.H. and J.F.H. performed northern blot analyses. M.S.D. contributed reagents. V.C.L. supervised *E. coli* experiments and contributed insights and analysis. All authors participated in the writing of the manuscript.

Competing Interests

B.C., J.F.H., D.Y., S.M.H., M.S.D., and P.C.D are co-inventors on two patents (PCT/US2019/013714, US 2019/0284624 A1) relating to the published work.

Data Availability

The complete sets of data for the RNA sequencing studies reported in **Figures 3-6** have been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject ID PRJNA579244. miRNA and standards data shown in **Figure 2** have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE139936.

Code Availability

The software used in the studies presented here is publicly available as follows. Blast version 2.6.0 (nucleotide BLAST) available at https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download. Peakfit.m version 9.0 available at Tom O'Haver, MATLAB Central File Exchange - https://terpconnect.umd.edu/~toh/spectrum/. fgrep (Linux command) available at

https://unix.stackexchange.com/questions/17949/what-is-the-difference-between-grep-egrep-and-fgrep. fastxtoolkit version 0.013 available at http://hannonlab.cshl.edu/fastx_toolkit/. Custom python scripts (.gitignore, blast_pair.py, cull.py) are available at GitHub https://github.com/dedonlab/aqrnaseq.

References

1. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63 (2009).
2. Cech, T.R. & Steitz, J.A. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77-94 (2014).
3. Hafner, M. et al. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* **17**, 1697-1712 (2011).
4. Zhang, Z., Lee, J.E., Riemondy, K., Anderson, E.M. & Yi, R. High-efficiency RNA cloning enables accurate quantification of miRNA expression by deep sequencing. *Genome Biol* **14**, R109 (2013).
5. Fuchs, R.T., Sun, Z., Zhuang, F. & Robb, G.B. Bias in ligation-based small RNA sequencing library construction is determined by adaptor and RNA structure. *PLoS One* **10**, e0126049 (2015).
6. Alon, S. et al. Barcoding bias in high-throughput multiplex sequencing of miRNA. *Genome Res* **21**, 1506-1511 (2011).
7. Zhuang, F., Fuchs, R.T., Sun, Z., Zheng, Y. & Robb, G.B. Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res* **40**, e54 (2012).
8. Pang, Y.L., Abo, R., Levine, S.S. & Dedon, P.C. Diverse cell stresses induce unique patterns of tRNA up- and down-regulation: tRNA-seq for quantifying changes in tRNA copy number. *Nucleic Acids Res* **42**, e170 (2014).

9. Linsen, S.E. et al. Limitations and possibilities of small RNA digital gene expression profiling. *Nat Methods* **6**, 474-476 (2009).

10. Machnicka, M.A., Olchowik, A., Grosjean, H. & Bujnicki, J.M. Distribution and frequencies of post-transcriptional modifications in tRNAs. *RNA Biol* **11**, 1619-1629 (2014).

11. Björk, G.R. et al. Transfer RNA modification. *Ann Rev Biochem* **56**, 263-287 (1987).

12. Shigematsu, M. et al. YAMAT-seq: an efficient method for high-throughput sequencing of mature transfer RNAs. *Nucleic Acids Res* **45**, e70 (2017).

13. Gogakos, T. et al. Characterizing Expression and Processing of Precursor and Mature Human tRNAs by Hydro-tRNAseq and PAR-CLIP. *Cell Rep* **20**, 1463-1475 (2017).

14. Kim, H. et al. Bias-minimized quantification of microRNA reveals widespread alternative processing and 3' end modification. *Nucleic Acids Res* **47**, 2630-2640 (2019).

15. Dai, Q., Zheng, G., Schwartz, M.H., Clark, W.C. & Pan, T. Selective Enzymatic Demethylation of N(2) ,N(2) -Dimethylguanosine in RNA and Its Application in High-Throughput tRNA Sequencing. *Angew Chem Int Ed Engl* **56**, 5017-5020 (2017).

16. Zheng, G. et al. Efficient and quantitative high-throughput tRNA sequencing. *Nat Methods* **12**, 835-837 (2015).

17. Cozen, A.E. et al. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat Methods* **12**, 879-884 (2015).

18. Mohr, S. et al. Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* **19**, 958-970 (2013).

19. Jayaprakash, A.D., Jabado, O., Brown, B.D. & Sachidanandam, R. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res* **39**, e141 (2011).

20. Lovett, S.T. & Kolodner, R.D. Identification and purification of a single-stranded-DNA-specific exonuclease encoded by the recJ gene of Escherichia coli. *Proc Natl Acad Sci U S A* **86**, 2627-2631 (1989).

21. Shepherd, J. & Ibba, M. Bacterial transfer RNAs. *FEMS Microbiol Rev* **39**, 280-300 (2015).

22. Ardell, D.H. & Hou, Y.M. Initiator tRNA genes template the 3' CCA end at high frequencies in bacteria. *BMC Genomics* **17**, 1003 (2016).

23. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**, D68-73 (2014).

24. Hebert Comparison of RNA sequencing kits. (2019).

25. Coenen-Stass, A.M.L. et al. Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. *RNA Biol* **15**, 1133-1145 (2018).

26. Dong, H., Nilsson, L. & Kurland, C.G. Co-variation of tRNA abundance and codon usage in Escherichia coli at differeng growth rates. *J Mol Biol* **260**, 649-663 (1996).

27. Lewis, K. Persister cells. *Annu Rev Microbiol* **64**, 357-372 (2010).

28. Grant, S.S. & Hung, D.T. Persistent bacterial infections, antibiotic tolerance, and the oxidative stress response. *Virulence* **4**, 273-283 (2013).

29. Zhang, Y. Persisters, persistent infections and the Yin-Yang model. *Emerg Microbes Infect* **3**, e3 (2014).

30. Steinberg, S., Misch, A. & Sprinzl, M. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* **21**, 3011-3015 (1993).

31. Chan, P.P. & Lowe, T.M. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* **44**, D184-189 (2016).

32. Phizicky, E.M. & Hopper, A.K. tRNA biology charges to the front. *Genes Dev* **24**, 1832-1860 (2010).

33. Chernyakov, I., Whipple, J.M., Kotelawala, L., Grayhack, E.J. & Phizicky, E.M. Degradation of several hypomodified mature tRNA species in Saccharomyces cerevisiae is mediated by Met22 and the 5'-3' exonucleases Rat1 and Xrn1. *Genes Dev* **22**, 1369-1380 (2008).

34. Hopper, A.K., Pai, D.A. & Engelke, D.R. Cellular dynamics of tRNAs and their genes. *FEBS Lett* **584**, 310-317 (2010).

35. Anderson, P. & Ivanov, P. tRNA fragments in human health and disease. *FEBS Lett* **588**, 4297-4304 (2014).

36. Cruz, J.W. et al. Growth-regulating Mycobacterium tuberculosis VapC-mt4 toxin is an isoacceptor-specific tRNase. *Nat Commun* **6**, 7480 (2015).

37. Schifano, J.M. et al. tRNA is a new target for cleavage by a MazF toxin. *Nucleic Acids Res* **44**, 1256-1270 (2016).

38. Lyons, S.M., Fay, M.M. & Ivanov, P. The role of RNA modifications in the regulation of tRNA cleavage. *FEBS Lett* **592**, 2828-2844 (2018).

39. Fu, H. et al. Stress induces tRNA cleavage by angiogenin in mammalian cells. *FEBS Lett* **583**, 437-442 (2009).

40. Haiser, H.J., Karginov, F.V., Hannon, G.J. & Elliot, M.A. Developmentally regulated cleavage of tRNAs in the bacterium Streptomyces coelicolor. *Nucleic Acids Res* **36**, 732-741 (2008).

41. Boccaletto, P. et al. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res* **46**, D303-D307 (2018).

42. Kietrys, A., Velema, W. & Kool, E. Fingerprints of modified RNA bases from deep sequencing profiles. *J Am Chem Soc* **139**, 17074-17081 (2017).

43. Hauenschild, R. et al. The reverse transcription signature of N-1-methyladenosine in RNA-Seq is sequence dependent. *Nucleic Acids Res* **43**, 9950-9964 (2015).

44. Motorin, Y., Muller, S., Behm-Ansmant, I. & Branlant, C. Identification of Modified Residues in RNAs by Reverse Transcription-Based Methods. Meth Enz **425**, 21-53 (2007).

45. Levanon, E.Y. et al. Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat Biotechnol* **22**, 1001-1005 (2004).

46. Xiong, Y. et al. A Comparison of mRNA Sequencing with Random Primed and 3'-Directed Libraries. *Sci Rep* **7**, 14626 (2017).

47. Herbert, Z. et al. Multi-site Evaluation of Next-Generation Methods for Small RNA Quantification. *J Biomolec Tech*, submitted for publication (2019).

Figure Legends

Figure 1. Overview of AQRNA-seq. (**a**) Library preparation workflow. (**b**) Control/optimization experiments for steps in the workflow. Top panel: Linker 1 ligation and removal proceeds with greater than 90% efficiency at a linker to tRNA molar ratio of 50:1. Middle panel: LCMS measurements of AlkB demethylation efficiencies for various RNA modifications. Bottom panel: Linker 2 ligation proceeds with near stoichiometric efficiency at linker to tRNA molar ratios of 30:1 or above. A 50:1 ratio is used in the method. Linker 2 removal is 100% efficient under those conditions. (**c**) Computational analysis workflow. Reads are mapped against a non-redundant reference genome. cDNA libraries generated and sequenced according to the AQRNA-seq, a paired-end protocol, will yield two FASTQ files per library – one containing the forward reads, and one containing the reverse reads. After alignment, multiply-mapped reads are mined separated from uniquely-mapped reads for abundance and coverage information.

Figure 2. Validation of AQRNA-seq. (**a**) Oligo spike-in studies. Five RNA oligos between 25 and 80 nucleotides in length were prepared for sequencing at different concentrations. The sequencing response per fmol of oligo is linear with amount of material present. (**b**) An equimolar mixture consisting of 963 synthetic microRNAs was sequenced using AQRNA-seq. Sequencing reads for each microRNA were normalized to their expected values and sorted. The colored bars and dotted lines indicate the percentage of the total mixture came within 2-fold and 10-fold of their expected abundance. (**c**) A comparison of microRNA libraries prepared using the AQRNA-seq protocol and various commercially-available small RNA sequencing kits shows significantly better performance for AQRNA-seq in terms of quantitative accuracy. Data for Illumina TruSeq Small RNA Library Prep Kit, Lexogen Small RNA-seq Library Prep Kit, NEBNext Small RNA Library Prep Set for Illumina, BioO POWERFLEX, QIAseq miRNA Library Kit, and Trilink CleanTag Small RNA Library Prep Kit taken from Herbert et al.[47] (**d**) Correlation of tRNA

quantification results using AQRNA-seq versus data derived from 2D gel electrophoresis and northern blotting by Dong et al.[26]

Figure 3. Application of AQRNA-seq to quantify tRNA pool dynamics during stress-induced persistence in *M. bovis* BCG. (**a**) Growth curves of showing approximately two logs of cell killing from day 0 to day 20 of complete nutrient deprivation followed by regrowth after resuscitation. (**b**) *In vitro* model of nutrient deprivation in *M. bovis* BCG. (**c**) A breakdown of the distribution of reads after size-selection, pairing of forward and reverse reads, and mapping. Of uniquely mapped and paired reads, 75% mapped to the internal standard sequence while 25% mapped to mature tRNA sequences.

Figure 4. AQRNA-seq quantifies the complete landscape of full-length and fragmented tRNAs, with single-nucleotide resolution of cleavage sites and modifications. (a) Sequence alignment plots reveal the abundance of quantitative information provided by AQRNA-seq. This is illustrated for tRNA Gly-CCC, with a stacked horizontal bar graph of the sequence alignment showing the start and end position of reads aligning to tRNA Gly-CCC. The numbering of tRNA nucleotides follows the standard nucleotide number system with positions 1 and 76 reflecting the 5' and 3' termini respectively. The anticodon is indicated by three vertical lines and an asterisk. The complete set of alignment plots for the 45 tRNA isoacceptors in *M. bovis* BCG is shown in **Supplementary Figure S1**. (**b**) Schematic showing site of linker 1 attachment and direction of reverse transcription along the tRNA sequence annotated using the Sprinzl coordinate system.[30] (**c**) The aligned reads fall into three different categories which have different interpretations – see text. (**d**) Example alignment plot for tRNA Tyr-GTA showing polymerase blockage downstream of the anticodon resulting in a lack of type 3 (full-length) reads. (**e**) Example of an alignment plot to tRNA Glu-TTC showing polymerase blockage close to the

anticodon and an enrichment of fragments aligning inside the 3'-end of the tRNA, resulting in an increased number of type 1 reads.

Figure 5. AQRNA-seq quantifies the tRNA pool dynamics during stress-induced persistence in *M. bovis* BCG (**a**) Plots showing the normalized abundance of all expressed BCG tRNAs throughout the time course of starvation. The green line represents the mean of tRNA abundances at the S0 timepoint. (**b**) Plots showing abundances of individual isoacceptors (all reads aligning at the 3' relative to the total set of tRNAs that carry the same amino acid.

Figure 6. Application of AQRNA-seq for quantitative mapping of the tRNA epitranscriptome. Many RNA modifications block 3'-to-5' reverse transcriptase-mediated cDNA synthesis or result in mutations in AQRNA-seq (gray line and arrow in panel **b**), which can be exploited to quantitatively map the modifications. (**a**) Subsets of *E. coli* tRNAs exhibit similar reverse transcriptase blockages at positions 38 and 48. The heat map shows the percentage of sequencing reads for which reverse transcription ended at a specific location in the tRNA sequence (columns) for each tRNA isoacceptor (rows). The two positions showing the most significant accumulations of polymerase blockade are noted with a small orange or red box. The RNA sequences surrounding these positions are shown in the larger boxes that magnify the sequence location, with specific modified nucleotides noted based on existing maps of *E. coli* tRNA modifications. In the orange boxes, the 8 tRNA species showing polymerase blockade at position 38 reveals all possess i^6A at position 37. In the red boxes, the 10 tRNAs showing polymerase blockade at position 48 all possess acp^3U at position 47. (**b**) In the absence of AlkB treatment, cDNA synthesis is blocked by m^1A at position 58 in nearly half of all BCG tRNAs, which is reflected by the high proportion of aligned reads that do not extend past position 58 in the heatmap of read start positions (light blue to orange line in heat map similar to panel **a**). This is illustrated for tRNA Glu-CTC in the gray stack plot, which shows that early all the reads begin

after position 58, forming a “cliff”. (**c**) After AlkB demethylation, however, the read alignments lengthen and extend past the cliff, resulting in a more varied distribution of alignment start positions. The heat map shows a significant reduction in the number of aligned reads. (**d**) Many RNA modifications can also be mapped by polymerase-induced mutations in the resulting cDNA. This is illustrated with a striking T-to-C misreading in BCG tRNA Arg-ACG, which is consistent with the presence of inosine at position 34 of the anticodon on nearly all copies of the tRNA.

Methods

Bacterial strains, culturing conditions, growth assays, and RNA isolation

All *E. coli* strains used were taken from the Keio collection (GenoBase; http://ecoli.aist-nara.ac.jp/)[1]. The genotype of each of the bacterial strains was validated before conducting the study. Strains were cultured in 10 ml LB broth (Fisher BioReagents BP1426-2) at 37 °C with constant shaking at 180 rpm until the cultures reached a final OD_{600} of 0.6-0.7. Culture pellets were harvested by centrifugation at 13,500 rpm for 2 min and immediately used for tRNA isolation using the Purelink miRNA isolation kit (Thermofisher K15700) following the manufacturer’s protocol. Briefly, cell pellets were resuspended in Trizol Reagent (Thermofisher 15596018) for lysis, followed by treatment with chloroform to separate the aqueous layer containing bulk tRNA. The aqueous layer was then subjected to a 2-column purification process where genomic DNA, larger RNA fragments (>200 bp), and excess salts, are removed. It is important to note that 5S rRNA cannot be separated from tRNA using this method. Three biological replicates were used in *E. coli* tRNA study.

Mycobacterium bovis Bacille Calmette-Guérin (BCG) str. Pasteur 1173P2 was grown in roller bottles with 7H9 broth or PBS (with 0.05% (v/v) tyloxapol, Sigma Aldrich T8761) at 2 rpm and 37 ^{o}C. Exponentially growing cultures with an OD_{600} of 0.8-1.0 were starved by washing pellets three times with PBS-tyloxapol. Starvation cultures were inoculated into PBS-tyloxapol at a starting OD_{600} of 1.0. Samples were retrieved at 4, 10, and 20 d after starvation. At day 20, cultures were resuspended in 7H9 and resuscitated for 6 d prior to harvesting. At each time point, cultures were plated on 7H10 agar for CFU determination. Specific compositions of 7H9 media, PBS, and 7H10 agar are as follows. Middlebrook 7H9 (BD Difco 271310) was supplemented with

0.5% (w/v) albumin (Sigma Aldrich A3059), 0.2% (w/v) glucose (Sigma Aldrich G8270), 0.085% (w/v) NaCl (Sigma Aldrich S7653), 0.2% (v/v) glycerol and 0.05% (v/v) Tween 80 (Sigma Aldrich P8074) as nutrient replete media; PBS (137 mM NaCl , 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4) was supplemented with 0.05% v/v tyloxapol, a non-hydrolyzable detergent. 7H10 agar (BD Difco 262710) was supplemented with 0.5% (v/v) glycerol and 10% (v/v) oleic acid albumin dextrose catalase (OADC; BD BBL 212351).

For colony forming unit (CFU) determination, serial dilutions of BCG cultures at various nutrient starvation/resuscitation timepoints were plated on 7H10 agar and incubated at 37 ^{o}C for 3-4 weeks. Colonies of the BCG timepoints were subsequently counted and respective CFUs determined.

For RNA isolation and purification from BCG, cells collected at nutrient starvation/resuscitation timepoints were lysed in the presence of TRI reagent (Sigma Aldrich T9424) with glass beads in a FastPrep FP120 bead-beater **as previously described[2]. Small RNAs were subsequently** isolated using the Purelink miRNA isolation kit (Invitrogen K157001). **Small and large** RNA fractions were profiled using the Agilent Bioanalyzer.

Optimized AQRNA-seq protocol

Ligation of input RNA and DNA Linker 1. Small RNAs (50 ng) were mixed with an 80-mer spike-in RNA oligo internal standard. The mixture was dephosphorylated in a 5 µl reaction containing 0.5 µl of reaction buffer (NEB T4 RNA ligase buffer) and 1 U of shrimp alkaline phosphatase (rSAP, NEB) at 37 °C for 30 min. The reaction was stopped by heat inactivation at 65 °C for 5 min followed by cooling on ice. Linker 1 ligation was performed by adding the following reagents directly to the dephosphorylation product: 1 µl Linker 1 (100 pmol/ul), 3 µl ATP (10 mM, NEB), 2.5 µl T4 RNA ligase buffer (NEB), 2 µl T4 RNA ligase 1 (30U/ul, NEB), 1.5 µl water and 15 µl PEG8000 (NEB). The ligation mixture was incubated at 25 °C for 2 h and 16 °C overnight. The ligation product was purified using the Zymo Oligo Clean & Concentrator kit (Zymo Research, D4060) according to the manufacturer's instructions. The sample was eluted in 20 µl water and kept on ice prior to Bioanalyzer analysis (Agilent, small RNA kit) or used directly in the demethylation step.

Demethylation. The ligated RNA sample was demethylated by AlkB (Arraystar, rtStar tRF&tiRNA Pretreatment Kit). A 2X reaction buffer was freshly prepared before the reaction and consisted of

150 µM 2-ketoglutarate, 4 mM L-ascorbic acid, 150 uM $(NH_4)_2Fe(SO_4)_2$, 100 µg/mL BSA (NEB, molecular biology grade, 10 mg/mL), and 100 mM HEPES (pH 8.0). The demethylation reaction was performed in a 100 µl volume consisting of 50 µl 2X reaction buffer, 20 µl Linker 1-ligated tRNA sample, 2 µl AlkB demethylase, and 1 µl RNase Inhibitor (NEB, murine, 40,000U/mL). The reaction was incubated at ambient temperature. After 2 h, the reaction was stopped via the addition of 50 µl water and 100 µl phenol:chloroform:isoamyl alcohol 25:24:1, pH 5.2. The mixture was mixed by inverting several times and centrifuged at 16,000xg for 10 min. The top layer was transferred to a new Eppendorf tube. Another 100 µl chloroform was added to the original mixture to remove any remaining phenol. After centrifugation, the top layer was removed and combined with the first extraction. The extracted sample was then purified using the Zymo Oligo Clean & Concentrator kit (Zymo Research, D4060) per the manufacturer's instructions. The sample was eluted in 17 µl water before proceeding to the next step (Linker 1 removal).

Removal of excess DNA Linker 1. In this step, the DNA oligo–adenylate intermediate was de-adenylated and subsequently digested, together with unused Linker 1, by exonuclease RecJ. The de-adenylation was performed in a 20 µl reaction containing 16 µl RNA sample from demethylation step, 2 µl NEB Buffer 2 (10X), and 2 µl 5'-deadenylase (NEB, 50U/uL). After incubation at 30 °C for 1 h, 2 µl RecJ (NEB; 30U/uL) was added. The mixture was incubated at 37 °C for 30 min followed by the addition of another 2 µl RecJ and further digestion for an additional 30 min. The reaction was stopped by heating at 65 °C for 20 min. The reaction mixture was purified using a DyEx spin column (Qiagen, 63204).

Reverse transcription. The RNA sample obtained from DyEx column purification is mixed with 1 µl RT-primer (2 pmol/ µl) and 1 µl dNTPs (10 mM each). The mixture was heated at 80 °C for 2 min and cooled immediately on ice. PrimeScript Buffer (6 µl; Clontech), 1 µl RNase Inhibitor (NEB), and 1 µl PrimeScript Reverse Transcriptase (Clontech, RR014A) were added. The reverse transcription reaction mixture was then incubated at 50 °C for 2 h after which the enzyme was inactivated at 70 °C for 15 min. The RNA strand was then hydrolyzed by adding 1 µl NaOH (5 M) followed by incubation at 90 °C for 3 min. The hydrolysis product was neutralized by adding 1 µl HCl (5M) and the reaction was cleaned up using the Zymo Oligo Clean & Concentrator kit (Zymo Research, D4060). The sample was eluted with 15 µl of water before vacuum concentration to 5 ul.

cDNA ligation. The purified cDNA was ligated to Linker 2 (**Table 2-4**) in a 20 µl reaction consisting of 5 µl cDNA sample, 1 µl Linker 2 (50 pmol/µl), 2 µl T4 DNA Ligase Buffer (NEB), 1 µl ATP (10 mM, NEB), 2 µl T4 DNA ligase (400U/uL, NEB), and 9 µl PEG8000 (NEB). The mixture was mixed and then incubated at 16 °C overnight for ligation. The ligated product was purified using the Zymo Oligo Clean & Concentrator kit (Zymo Research, D4060) per the manufacturer's instructions and eluted in 16 µl of water.

Removal of excess DNA Linker 2. After cDNA ligation, excess Linker 2 was removed by a two-step purification. First, adenylated linker intermediates were de-adenylated. Then, RecJ was used to digest the de-adenylated product. The de-adenylation was performed in a 20 µl reaction containing 16 µl RNA sample from the previous, cDNA ligation step, 2 µl NEB Buffer 2 (10X), and 2 µl 5'-deadenylase (NEB, 50U/ul). After incubation at 30 °C for 1 h, 2 µl RecJ (NEB; 30U/ul) was added. The reaction was incubated at 37 °C for 30 min. Subsequently, another 2 µl RecJ was added for further digestion for an additional 30 min. The reaction was stopped through heat inactivation at 65 °C for 20 min.

PCR amplification and Illumina sequencing. Purified cDNA from the previous step was amplified by PCR in a 100 µl mixture containing 24 µl cDNA template, 50 µl seqAMP DNA polymerase buffer (2X buffer, Clontech), 2 µl each of PCR primer F and R with unique sequencing barcodes (**Table 2-4**, 1 µM each), 2 µl Clontech seqAMP DNA polymerase (Clontech) and 20 µl water. The PCR reaction was performed according to the manufacturer's instructions with an annealing temperature of 58 °C and 13 reaction cycles. The PCR product was extracted and purified from an agarose gel using a standard gel purification kit (QIAquick Gel Extraction Kit, Qiagen). The gel-extracted samples were mixed together (multiplexing) and submitted for Illumina sequencing. In the studies described, sequencing was performed on the Illumina NEXTseq sequencer (BioMicroCenter, MIT) with custom primers F and R (**Table 1-1**).

Table 1-1 DNA and RNA oligos required for the AQRNA-seq protocol. Asterisk (*) = sites of phosphorothioate modification. N = randomized position (equal mixture of all 4 bases used during oligo synthesis). Unless otherwise indicated, oligos terminate in hydroxyl groups at both 5' and 3' ends.

Name	Sequence (5'-3')
Spike-in internal RNA standard (80-mer)	ACCCACGGCAGAGUGUGUGUGGCCCACGCGAUUCGUGAAUAACAUA ACUAUGAGUAGGAUAAGGAAUGUCACCUAACAGC
Optional second spike-in internal RNA standard (40-mer)	NNNUAUCCUAAUCAUCUUAAUACUACAAUGGACCAUGNNN
Linker 1	/5Phos/-NNCAC TCG GGC ACC AAG GAddC-3'
Linker 2	/5Phos/TG AAG AGC CTA GTC GCT GTT CAN NNN NNC TGC CCA TAG AG/3SpC3/
RT-primer	G*T*C*C*T*T*GGTGCCCGAGTG
PCR primer F-1	AATGATACGGCGACCACCGAGATCTACACAGAGAGACACTCTTTCCCT ACACGACGCTCTTCCGATCTTGAACAGCGACTAGGCTCTTCA
PCR primer F-2	AATGATACGGCGACCACCGAGATCTACACGTGTGTACACTCTTTCCCT ACACGACGCTCTTCCGATCTTGAACAGCGACTAGGCTCTTCA
PCR primer F-3	AATGATACGGCGACCACCGAGATCTACACTCTCTCACACTCTTTCCCT ACACGACGCTCTTCCGATCTTGAACAGCGACTAGGCTCTTCA
PCR primer F-4	AATGATACGGCGACCACCGAGATCTACACCACACAACACTCTTTCCCT ACACGACGCTCTTCCGATCTTGAACAGCGACTAGGCTCTTCA
PCR primer F-5	AATGATACGGCGACCACCGAGATCTACACCAAGGTACACTCTTTCCCT ACACGACGCTCTTCCGATCTTGAACAGCGACTAGGCTCTTCA
PCR primer F-6	AATGATACGGCGACCACCGAGATCTACACAGGTTCACACTCTTTCCCT ACACGACGCTCTTCCGATCTTGAACAGCGACTAGGCTCTTCA
PCR primer R-1	CAAGCAGAAGACGGCATACGAGATCCTGAGCGGTCTCGGCATTCCTG CTGAACCGCTCTTCCGATCTGTCCTTGGTGCCCGAGTG
PCR primer R-2	CAAGCAGAAGACGGCATACGAGATTGCTGTCGGTCTCGGCATTCCTG CTGAACCGCTCTTCCGATCTGTCCTTGGTGCCCGAGTG
PCR primer R-3	CAAGCAGAAGACGGCATACGAGATCATCACCGGTCTCGGCATTCCTG CTGAACCGCTCTTCCGATCTGTCCTTGGTGCCCGAGTG
PCR primer R-4	CAAGCAGAAGACGGCATACGAGATTTCAGACGGTCTCGGCATTCCTG CTGAACCGCTCTTCCGATCTGTCCTTGGTGCCCGAGTG
PCR primer R-5	CAAGCAGAAGACGGCATACGAGATACGGTGCGGTCTCGGCATTCCTG CTGAACCGCTCTTCCGATCTGTCCTTGGTGCCCGAGTG
PCR primer R-6	CAAGCAGAAGACGGCATACGAGATGTAACACGGTCTCGGCATTCCTG CTGAACCGCTCTTCCGATCTGTCCTTGGTGCCCGAGTG
Custom primer F	GCTCTTCCGATCT TGAACAGCGACTAGGCTCTTCA
Custom primer R	TGAACCGCTCTTCCGATCT GTCCTTGGTGCCCGAGTG

Optimization of AlkB demethylation conditions

LC-MS/MS analysis of methylated ribonucleosides. Ribonucleosides were resolved with a Phenomenex Synergi Fusion-reverse phase column (100 x 2 mm, 2.5 µm particle size, 100 Å pore size) eluted with the following gradient of acetonitrile in 5 mM ammonium acetate (pH 5.3) at a flow rate of 0.35 ml/min and 35 °C: 0-1 min, 0%; 1-10 min, 0-10%, 10-14 min, 10-40%, 14-15 min, 40-80%. The HPLC column was coupled to an Agilent 6430 triple quadrupole LC/MS spectrometer with an electrospray ionization source where it was operated in positive ion mode with the following parameters for voltages and source gas: gas temperature, 350 °C; gas flow, 10 l/min; nebulizer, 45 psi; and capillary voltage, 3500 V. The first and third quadrupoles (Q1 and Q3) were fixed to unit resolution and the modifications were quantified by pre-determined molecular transitions. The dwell time for each ribonucleoside was 500 ms. The retention time, *m/z* of the transmitted parent ion, *m/z* of the monitored product ion, fragmentor voltage, and collision energy of each modified nucleoside are as follows: m^1A, 3.6 min, *m/z* 282 → 150, 100 V, 16 V; m^1G, 6.1 min, *m/z* 298 → 166, 90 V, 10 V; m^1I, 5.9 min, *m/z* 283 → 151, 80 V, 10 V; $m^{22}G$, 7.8 min, *m/z* 312 → 180, 100 V, 8 V. Modified ribonucleosides were identified using commercially available nucleoside standards. Three independent tRNA replicates were used to test AlkB demethylation efficiencies in this study.

LC-MS/MS data analysis. Quantitative comparisons between control and demethylase-treated samples from different buffers were made possible by correcting for variation in tRNA quantities by dividing raw peak are for each ribonucleoside by the ultraviolet absorbance peak areas for the four canonical ribonucleosides. The demethylation efficiencies were calculated by dividing the peak area of the demethylase-treated sample by the peak area of the control sample for each modification.

Optimization of linker ligation conditions and determination of linker removal

Linker 1 ligation studies. Tests of Linker 1 ligation efficiencies were performed by combining dephosphorylated tRNA with ATP, T4 RNA ligase buffer, T4 RNA ligase 1, PEG8000, and varying amounts of Linker 1. The reaction was allowed to proceed under conditions consistent with the manufacturer's recommendations and the resulting products were analyzed using the Bioanalyzer small RNA chip. Electropherogram peaks corresponding to ligated and unligated tRNA were fitted and integrated using Peakfit.m (version 9.0; Tom O'Haver, MATLAB Central File Exchange - https://terpconnect.umd.edu/~toh/spectrum/), a Matlab-based peak fitting program that uses an unconstrained non-linear optimization algorithm to decompose a complex peak signal into its fundamental underlying component parts. The fraction of ligated tRNA was calculated by dividing the summed peak area

corresponding to the ligated tRNA to the total peak area. Three independent replicates were used for linker 1 ligation efficiency testing.

Linker 2 ligation studies. To test the efficiency of the Linker 2 ligation reaction, we used single-stranded DNA oligos with phosphorothioate modification at the 5' end to simulate the reverse transcription output. The sequences for the oligos used here are listed in **Table 2-4**. To determine the optimal linker:oligo ratio, we combined the 80-nucleotide oligo with ATP, T4 DNA ligase buffer, T4 DNA ligase, PEG8000, and varying amounts of Linker 2. To determine whether oligo length altered ligation efficiencies, we used a constant linker:oligo ratio and varied the length of the oligo tested. The reactions were allowed to proceed under conditions consistent with the manufacturer's recommendations and the resulting products were analyzed using the Bioanalyzer small RNA chip. Similar to Linker 1 ligation studies, peaks were deconvoluted and integrated using Peakfit.m. The fraction of ligated oligo was used as a metric for ligation efficiency and used to determine final, optimized ligation conditions. Three independent replicates were used for cDNA ligation efficiency testing.

Table 2-4 DNA oligo sequences for Linker 2 ligation studies. Asterisk (*) = sites of phosphorothioate modification. N = randomized position (equal mixture of all 4 bases used during oligo synthesis). Unless otherwise indicated, oligos terminate in hydroxyl groups at both 5' and 3' ends.

Name	**Sequence (5'-3')**
cDNA 50-mer	G*G*T*A*G*C*ACAAAAGTATTACCATGGTCCTAGAAGTTCGGCACAGTTANNN
cDNA 60-mer	A*C*T*G*T*T*GTTTGGTAGCACAAAAGTATTACCATGGTCCTAGAAGTTCGGCACAGTTANNN
cDNA 80-mer	A*C*G*C*T*G*CCACGTGTTCATTAACTGTTGTTTGGTAGCACAAAAGTATTACCATGGTCCTAGAAGTTCGGCACAGTTANNN

Linker removal studies. The linker removal steps were performed on samples from the Linker 1 and 2 ligation studies. The extent of linker removal by deadenylase and RecJ was determined from Bioanalyzer electropherograms using peak fitting and integration. Three independent replicates were used to test linker 1 and 2 removal efficiencies in this study.

Library construction for control AQRNA-seq samples

Standard RNA oligos. Five synthetic RNA oligos of different lengths and sequences were used to study the effect of oligo concentration and length on sequencing response. The oligos, whose sequences are listed in **Table 2-5**, were diluted from stocks to varying final concentrations and mixed together into standard mix samples A-E according to the scheme

presented in **Table 2-6**. Samples A through E were prepared in triplicate and used as input RNA samples for AQRNA-sequencing.

Table 2-5 RNA oligo sequences for linearity studies. N = randomized position (equal mixture of all 4 bases used during oligo synthesis). Unless otherwise indicated, oligos terminate in hydroxyl groups at both 5' and 3' ends.

Name	**Sequence (5'-3')**
25-mer std	NNCAUUGUUAAGCGAACGGCUGCNN
40-mer std	NNNUAUCCUAAUCAUCUUAAUACUACAAUGGACCAUGNNN
60-mer std	NNNCGACCCAAGUUGUCUGAUUAAGCAUAUUAAGAGCCGAGGAUCGUAUCG CCACCCNNN
80-mer std 1	UCCUGAUAAUUGCAGAGAAGUGACUUCGUGGGUGUCACCAGUAAUAAGGCA UCGCCAGCUAACAAUCGGUGGGAACGUCA
80-mer std 2	UGUCUUUCACGGAGACUGCGGGUAGCUCUGACAAGCUCCUAUGGUGAUACC ACCACGCUCGUAAUGAGAUGGGAUAGUCG

Table 2-6 Standard mixes of RNA oligos. Target abundances (in femtomoles) of each standard within the five standard mix samples A, B, C, D, and E.

	A	B	C	D	E
25-mer std	40	80	120	160	200
40-mer std	80	120	160	200	40
60-mer std	120	160	200	40	80
80-mer std 1	160	200	40	80	120
80-mer std 2	200	40	80	120	160

MicroRNA mixture. The miRXplore Universal Reference (Miltenyi Biotec) consists of 963 synthetic unmodified, HPLC-purified RNA oligos. The sample was reconstituted according to manufacturer's instructions and aliquoted. Three aliquots were used as separate RNA inputs for AQRNA-seq library construction.

AQRNA-seq data processing

Forward and reverse reads were trimmed of their respective adapter sequences using fastxtoolkit (version 0.013). A minimum adapter alignment length of 10 bp was required, and unknown (N) nucleotides were kept. Sequences were blasted against a reference library using blast (version 2.6.0) with the parameters: blast -perc identity 90 -word_size 9 -dust no -soft_masking false. For sequencing libraries prepared from BCG tRNA, a reference library was created based on the 48 entries in the genomic tRNA database for *Myocbacterium bovis*

BCG str. Pasteur_1173P2.[3] (http://gtrnadb.ucsc.edu/GtRNAdb2/genomes/bacteria/Myco_bovi_BCG_Pasteur_1173P2_BCG_Pasteur_1173P2/Myco_bovi_BCG_Pasteur_1173P2_BCG_Pasteur_1173P2-gene-list.html). Sequences corresponding to duplicate tRNA genes (e.g., tRNA-Ala-TGC-1-2 and tRNA-Ile-GAT-1-2) and tRNA pseudogenes (tRNA-Ser-CGA-2-1) were removed to eliminate redundant entries and reduce the incidence of ambiguous or false positive matches. The terminal (3') CCA sequence was added to tRNA sequences where it is not genomically encoded. The sequence for the 80-nt RNA oligo internal standard was added to the reference library. For the control samples, the sequences of the 5 synthetic RNA oligos was used to create a reference library, along with the 80-nt RNA oligo internal standard.

For each tRNA and control sample, forward and reverse reads were merged by integrating their start and end positions to generate new start and end positions that reflect their combined coverage. Multiple alignments were reduced by ranking all the alignments for a given read by their e-value and retaining only the alignment with the lowest e-value. Forward and reverse reads were required to match the same target. Paired reads that did not match the same target were stored in a separate file and not analyzed. These manipulations were carried out with the python script cull.py. Uniquely-mapped reads were then tabulated and counted. For the microRNA samples, the set of 963 microRNA sequences contained within the miRXplore Universal Reference product was combined with the sequence for the 80-nt RNA internal standard oligo to generate the reference library. In that analysis, the number of exact sequencing reads that matched to the reference microRNA sequence in each trimmed sequencing file was determined with fgrep: numberReadsPerFile=$(fgrep $miRNA_sequence $trimmed_sequencing_file | wc -l). The read counts of the miRNAs were normalized to the summed counts for all detected miRNAs to obtain a “normalized read count”. The summed counts of all detected miRNAs were also divided by 963, the total number of detected miRNAs, to obtain the “expected read count” assuming all species were equimolar. The read ratio was calculated by dividing the normalized read count by the expected read count.

Data availability

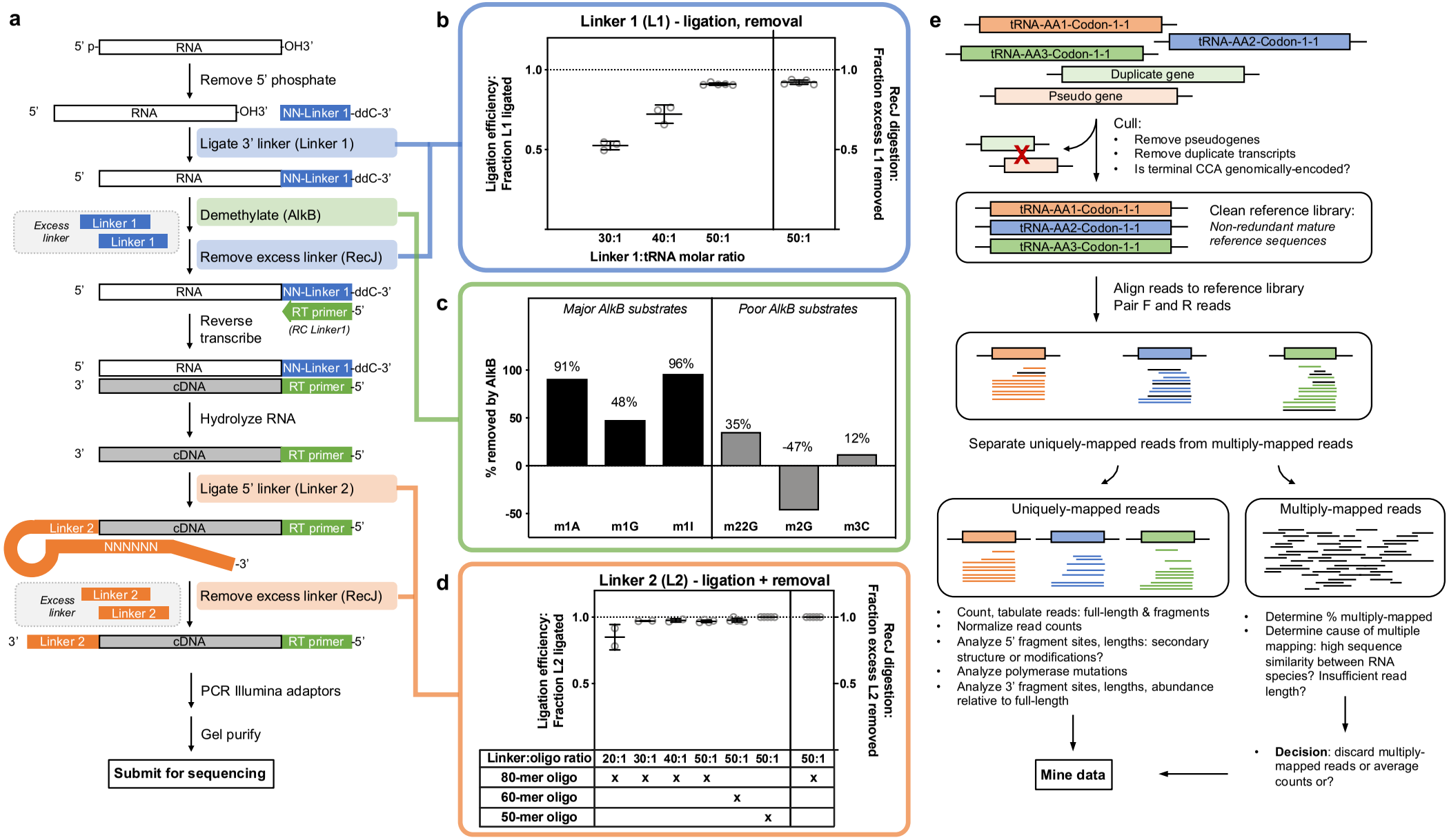
All custom scripts have been made available at https://github.com/dedonlab/aqrnaseq. Additional modified scripts can be accessed upon request. All sequencing data that support the findings of this study have been variously deposited in the National Center for Biotechnology Information Sequence Read Archive (BioProject accession number PRJNA579244; BCG starvation data files) and the Gene Expression Omnibus (GEO accession number GSE139936; control and microRNA data files). All other relevant data are available from the corresponding author upon request.

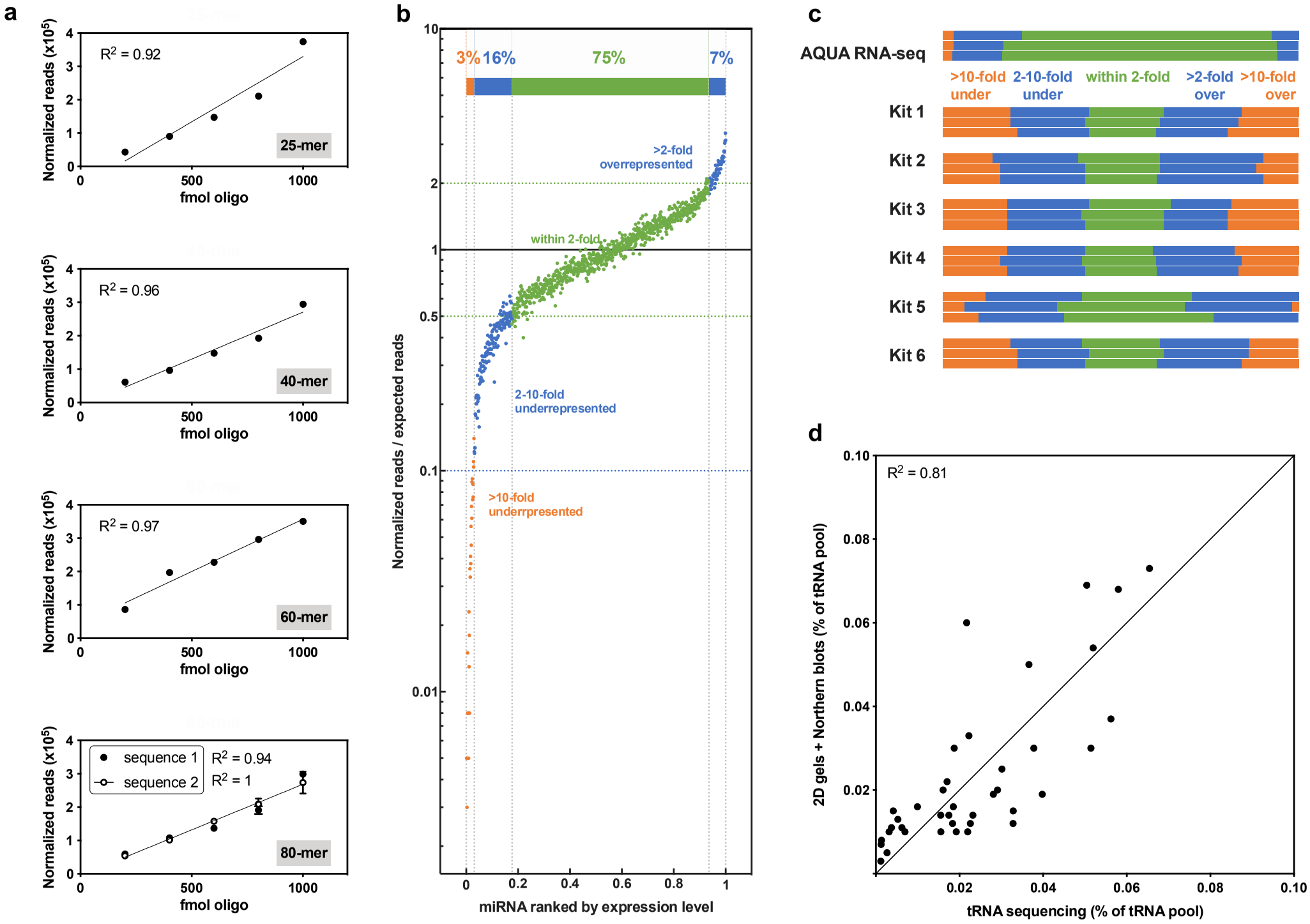
References

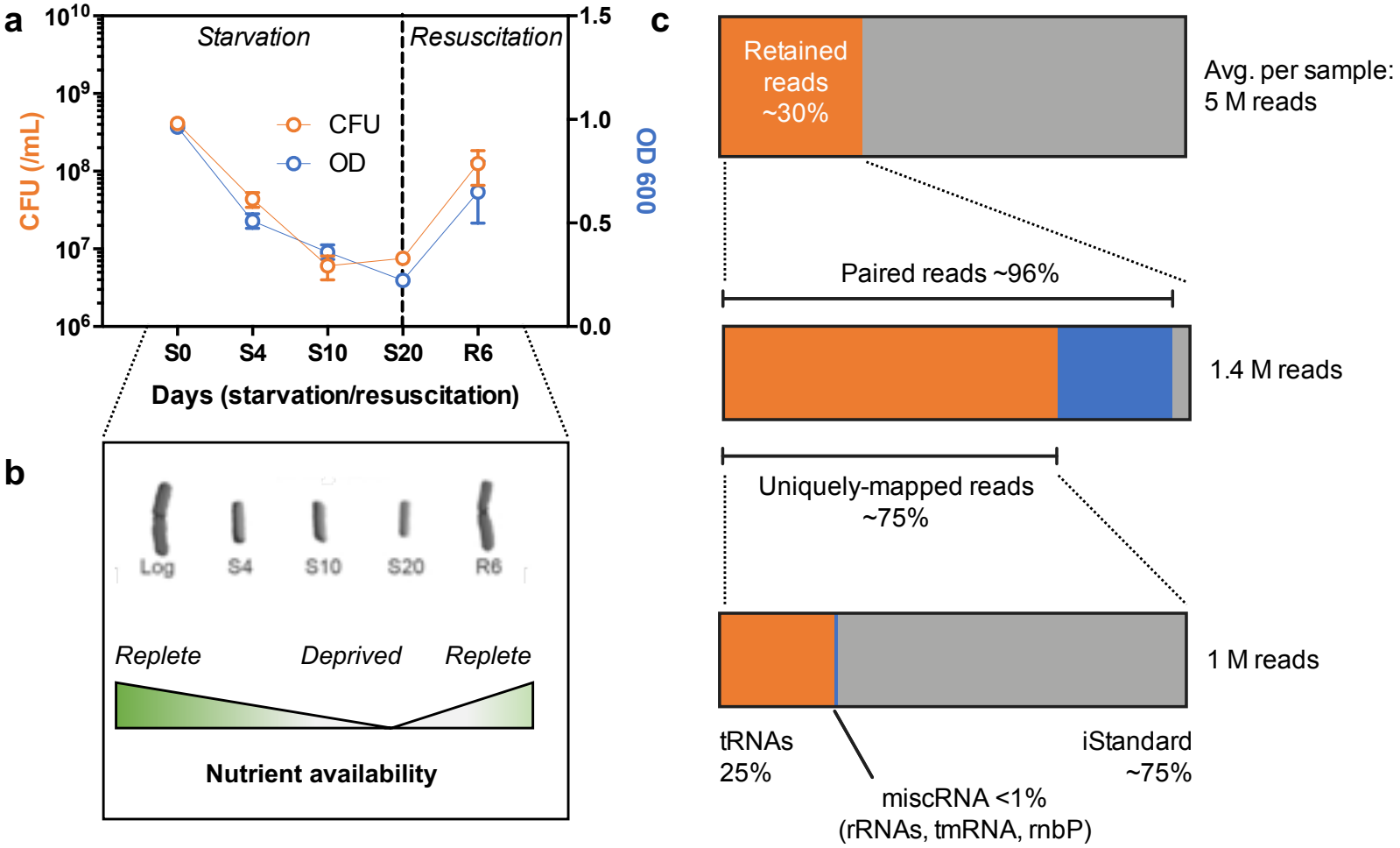
1 Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006 0008 (2006).

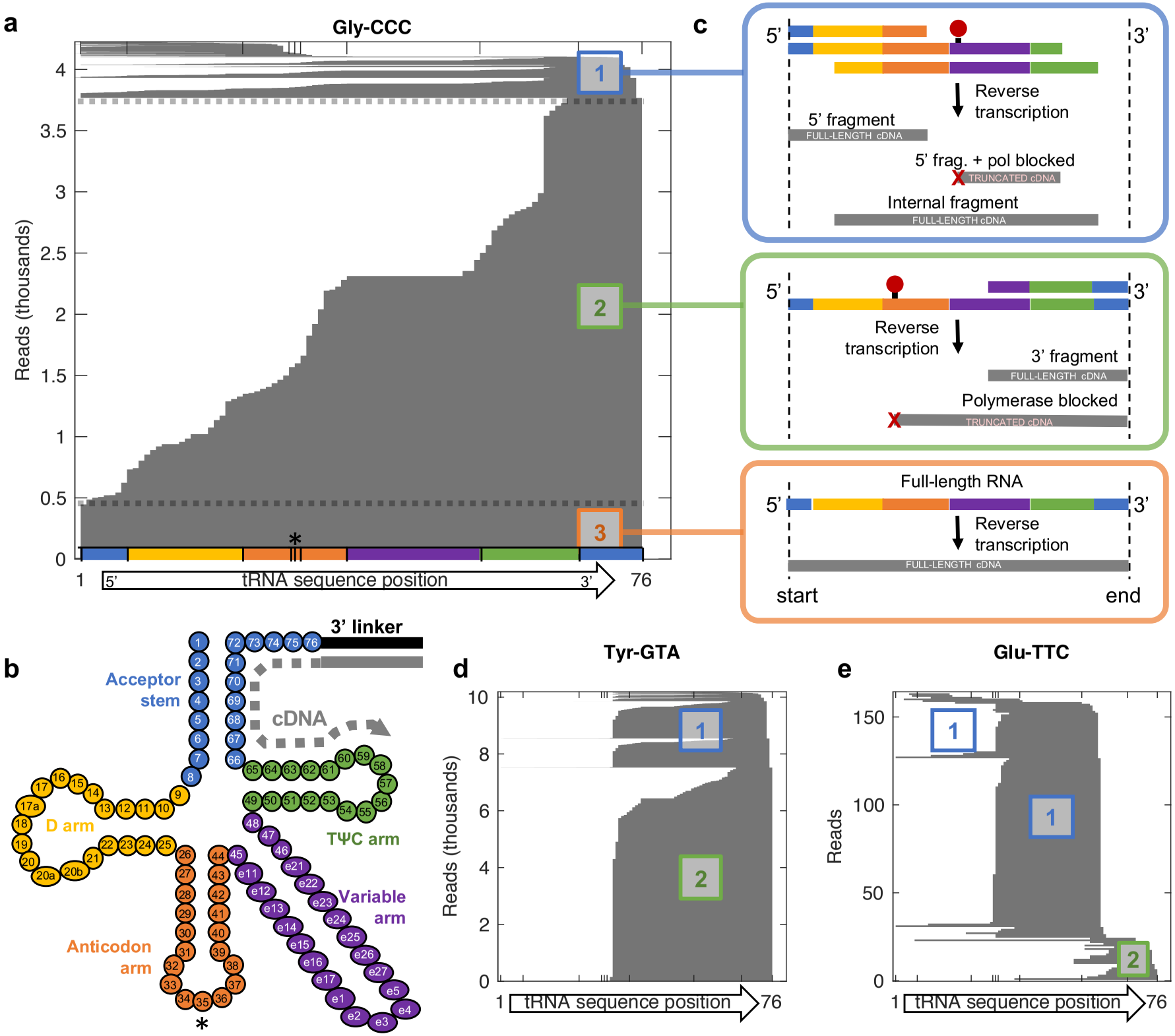
2 Hia, F. *et al.* Mycobacterial RNA isolation optimized for non-coding RNA: high fidelity isolation of 5S rRNA from Mycobacterium bovis BCG reveals novel post-transcriptional processing and a complete spectrum of modified ribonucleosides. *Nucleic Acids Res* **43**, e32 (2015).

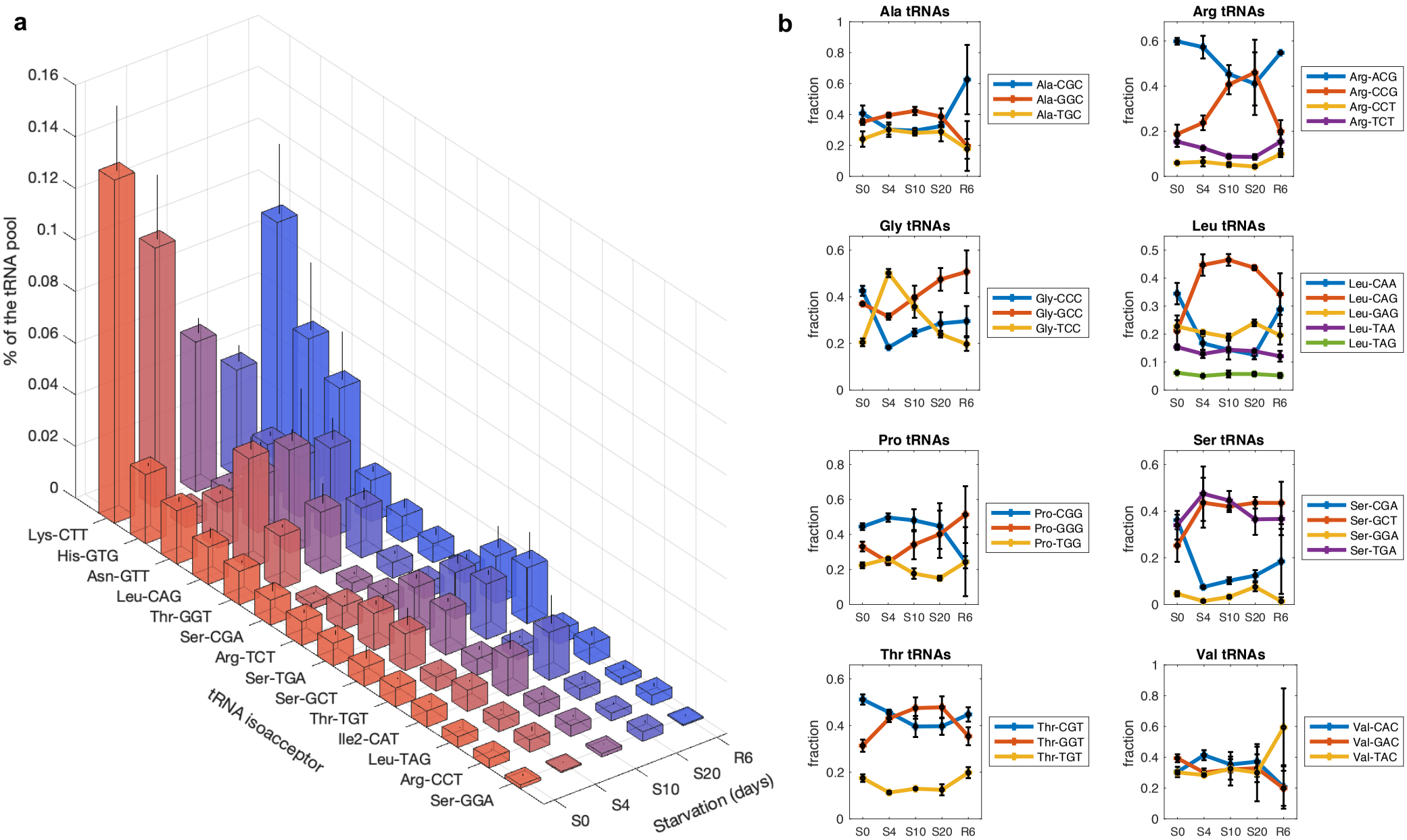
3 Chan, P. P. & Lowe, T. M. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* **44**, D184-189 (2016).

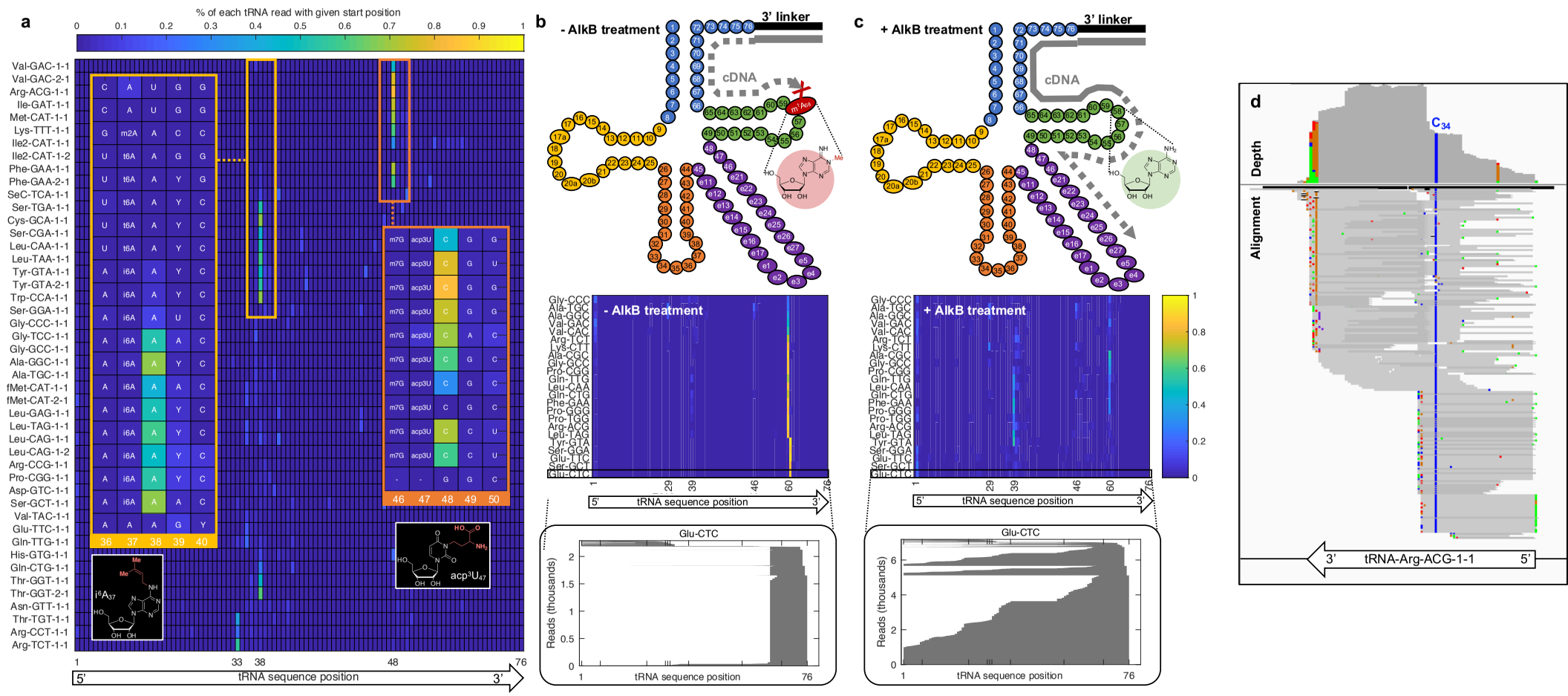

a
5' p- RNA -OH3'
Remove 5' phosphate
5' RNA -OH3' NN-Linker 1 -ddC-3'
Ligate 3' linker (Linker 1)
5' RNA NN-Linker 1 -ddC-3'
Excess linker
Linker 1
Linker 1
Demethylate (AlkB)
Remove excess linker (RecJ)
5' RNA NN-Linker 1 -ddC-3'
RT primer -5'
(RC Linker1)
Reverse transcribe
5' RNA NN-Linker 1 -ddC-3'
3' cDNA RT primer -5'
Hydrolyze RNA
3' cDNA RT primer -5'
Ligate 5' linker (Linker 2)
Linker 2 cDNA RT primer -5'
NNNNNN
-3'
Excess linker
Linker 2
Linker 2
Remove excess linker (RecJ)
3' Linker 2 cDNA RT primer -5'
PCR Illumina adaptors
Gel purify
Submit for sequencing
b
Linker 1 (L1) - ligation, removal
Ligation efficiency: Fraction L1 ligated
RecJ digestion: Fraction excess L1 removed
1.0
0.5
30:1
40:1
50:1
50:1
Linker 1:tRNA molar ratio
c
Major AlkB substrates
Poor AlkB substrates
% removed by AlkB
100
50
0
-50
91%
48%
96%
35%
-47%
12%
m1A
m1G
m1I
m22G
m2G
m3C
d
Linker 2 (L2) - ligation + removal
Ligation efficiency: Fraction L2 ligated
RecJ digestion: Fraction excess L2 removed
1.0
0.5
Linker:oligo ratio 20:1 30:1 40:1 50:1 50:1 50:1 50:1
80-mer oligo x x x x x
60-mer oligo x
50-mer oligo x
e
tRNA-AA1-Codon-1-1
tRNA-AA2-Codon-1-1
tRNA-AA3-Codon-1-1
Duplicate gene
Pseudo gene
Cull:
• Remove pseudogenes
• Remove duplicate transcripts
• Is terminal CCA genomically-encoded?
tRNA-AA1-Codon-1-1
tRNA-AA2-Codon-1-1
tRNA-AA3-Codon-1-1
Clean reference library:
Non-redundant mature reference sequences
Align reads to reference library
Pair F and R reads
Separate uniquely-mapped reads from multiply-mapped reads
Uniquely-mapped reads
Multiply-mapped reads
• Count, tabulate reads: full-length & fragments
• Normalize read counts
• Analyze 5' fragment sites, lengths: secondary structure or modifications?
• Analyze polymerase mutations
• Analyze 3' fragment sites, lengths, abundance relative to full-length
• Determine % multiply-mapped
• Determine cause of multiple mapping: high sequence similarity between RNA species? Insufficient read length?
• Decision: discard multiply-mapped reads or average counts or?
Mine data


a
Normalized reads (x10^5)
R^2 = 0.92
25-mer
fmol oligo
R^2 = 0.96
40-mer
R^2 = 0.97
60-mer
sequence 1
sequence 2
R^2 = 0.94
R^2 = 1
80-mer
b
3%
16%
75%
7%
>2-fold overrepresented
within 2-fold
2-10-fold underrepresented
>10-fold underrpresented
Normalized reads / expected reads
miRNA ranked by expression level
c
AQUA RNA-seq
>10-fold under
2-10-fold under
within 2-fold
>2-fold over
>10-fold over
Kit 1
Kit 2
Kit 3
Kit 4
Kit 5
Kit 6
d
R^2 = 0.81
2D gels + Northern blots (% of tRNA pool)
tRNA sequencing (% of tRNA pool)


a
Starvation
Resuscitation
CFU
OD
CFU (/mL)
OD 600
S0
S4
S10
S20
R6
Days (starvation/resuscitation)
b
Log
S4
S10
S20
R6
Replete
Deprived
Replete
Nutrient availability
c
Retained reads ~30%
Avg. per sample: 5 M reads
Paired reads ~96%
1.4 M reads
Uniquely-mapped reads ~75%
1 M reads
tRNAs 25%
iStandard ~75%
miscRNA <1% (rRNAs, tmRNA, rnbP)


a
Gly-CCC
Reads (thousands)
tRNA sequence position
5'
3'
1
76
1
2
3
c
5'
3'
Reverse transcription
5' fragment
FULL-LENGTH cDNA
5' frag. + pol blocked
TRUNCATED cDNA
Internal fragment
FULL-LENGTH cDNA
Polymerase blocked
3' fragment
Full-length RNA
start
end
b
3' linker
Acceptor stem
cDNA
D arm
TΨC arm
Variable arm
Anticodon arm
d
Tyr-GTA
Reads (thousands)
e
Glu-TTC
Reads


a
% of the tRNA pool
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
Lys-CTT
His-GTG
Asn-GTT
Leu-CAG
Thr-GGT
Ser-CGA
Arg-TCT
Ser-TGA
Ser-GCT
Thr-TGT
Ile2-CAT
Leu-TAG
Arg-CCT
Ser-GGA
tRNA isoacceptor
S0
S4
S10
S20
R6
Starvation (days)
b
Ala tRNAs
Ala-CGC
Ala-GGC
Ala-TGC
Arg tRNAs
Arg-ACG
Arg-CCG
Arg-CCT
Arg-TCT
Gly tRNAs
Gly-CCC
Gly-GCC
Gly-TCC
Leu tRNAs
Leu-CAA
Leu-CAG
Leu-GAG
Leu-TAA
Leu-TAG
Pro tRNAs
Pro-CGG
Pro-GGG
Pro-TGG
Ser tRNAs
Ser-CGA
Ser-GCT
Ser-GGA
Ser-TGA
Thr tRNAs
Thr-CGT
Thr-GGT
Thr-TGT
Val tRNAs
Val-CAC
Val-GAC
Val-TAC
fraction
S0
S4
S10
S20
R6


a
% of each tRNA read with given start position
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Val-GAC-1-1
Val-GAC-2-1
Arg-ACG-1-1
Ile-GAT-1-1
Met-CAT-1-1
Lys-TTT-1-1
Ile2-CAT-1-1
Ile2-CAT-1-2
Phe-GAA-1-1
Phe-GAA-2-1
SeC-TCA-1-1
Ser-TGA-1-1
Cys-GCA-1-1
Ser-CGA-1-1
Leu-CAA-1-1
Leu-TAA-1-1
Tyr-GTA-1-1
Tyr-GTA-2-1
Trp-CCA-1-1
Ser-GGA-1-1
Gly-CCC-1-1
Gly-TCC-1-1
Gly-GCC-1-1
Ala-GGC-1-1
Ala-TGC-1-1
fMet-CAT-1-1
fMet-CAT-2-1
Leu-GAG-1-1
Leu-TAG-1-1
Leu-CAG-1-1
Leu-CAG-1-2
Arg-CCG-1-1
Pro-CGG-1-1
Asp-GTC-1-1
Ser-GCT-1-1
Val-TAC-1-1
Glu-TTC-1-1
Gln-TTG-1-1
His-GTG-1-1
Gln-CTG-1-1
Thr-GGT-1-1
Thr-GGT-2-1
Asn-GTT-1-1
Thr-TGT-1-1
Arg-CCT-1-1
Arg-TCT-1-1
36 37 38 39 40
46 47 48 49 50
i6A37
acp3U47
1 33 38 48 76
5' tRNA sequence position 3'
b
- AlkB treatment
3' linker
cDNA
m1A58
- AlkB treatment
5' tRNA sequence position 3'
Glu-CTC
Reads (thousands)
tRNA sequence position
c
+ AlkB treatment
3' linker
cDNA
+ AlkB treatment
5' tRNA sequence position 3'
Glu-CTC
Reads (thousands)
tRNA sequence position
d
Depth
Alignment
C34
3' tRNA-Arg-ACG-1-1 5'