UC Riverside
UC Riverside Previously Published Works

Title

Mass Spectrometry-Based Direct Sequencing of tRNAs De Novo and Quantitative Mapping of Multiple RNA Modifications.

Permalink

https://escholarship.org/uc/item/08x970sr

Journal

Journal of the American Chemical Society, 146(37)

Authors

Yuan, Xiaohong
Su, Yue
Johnson, Benjamin
et al.

Publication Date

2024-09-18

DOI

10.1021/jacs.4c07280

Peer reviewed

Article

Mass Spectrometry-Based Direct Sequencing of tRNAs De Novo and Quantitative Mapping of Multiple RNA Modifications

Xiaohong Yuan,[⊥] Yue Su,[⊥] Benjamin Johnson, Michele Kirchner, Xudong Zhang, Sihang Xu, Sophia Jiang, Jing Wu, Shundi Shi, James J. Russo, Qi Chen, and Shenglong Zhang*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

Incomplete RNA ladder that cannot be MS sequenced

MLC fixed RNA ladder that can be MS sequenced
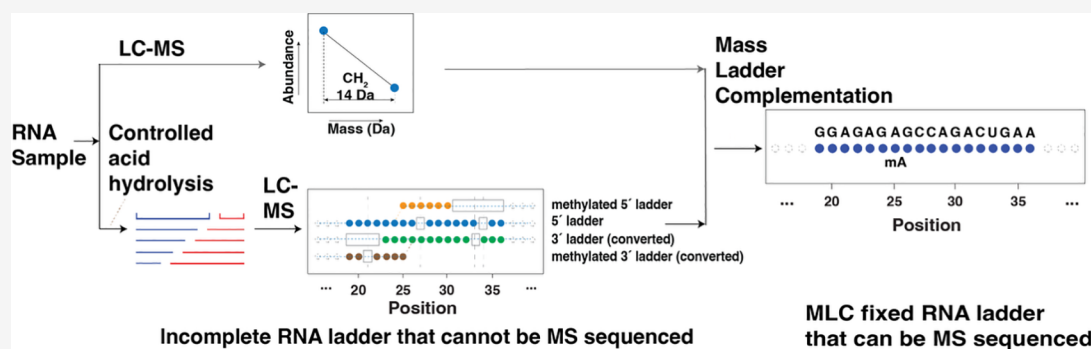
ABSTRACT: Despite the extensive use of next-generation sequencing (NGS) of RNA, simultaneous direct sequencing and quantitative mapping of multiple RNA nucleotide modifications remains challenging. Mass spectrometry (MS)-based sequencing can directly sequence all RNA modifications without being limited to specific ones, but it requires a perfect MS ladder that few tRNAs can provide. Here, we describe an MS ladder complementation sequencing approach (MLC-Seq) that circumvents the perfect ladder requirement, allowing de novo MS sequencing of full-length heterogeneous cellular tRNAs with multiple nucleotide modifications at single-nucleotide precision. Unlike NGS-based methods, which lose RNA modification information, MLC-Seq preserves RNA sequence diversity and modification information, revealing new detailed stoichiometric tRNA modification profiles and their changes upon treatment with the dealkylating enzyme AlkB. It can also be combined with reference sequences to provide quantitative analysis of diverse tRNAs and modifications in total tRNA samples. MLC-Seq enables systematic, quantitative, and site-specific mapping of RNA modifications, revealing the truly complete informational content of tRNA.

■ INTRODUCTION

Despite the widespread utilization of high-throughput next-generation sequencing (NGS), the "true" sequence of RNA, i.e., the identity and location of every nucleotide (canonical or modified) within a full-length RNA,[1] remains unknown due to the lack of a general method for direct sequencing of any nucleotide (modified or not) at single-nucleotide resolution. NGS-based RNA sequencing methods do not sequence RNA directly but rather its complementary DNA (cDNA), which contains canonical nucleotides only. To sequence modified RNA nucleotides, these NGS-based methods require additional specific procedures.[2−8] Only a small number of the over 170 known modified nucleotides can be identified by NGS sequencing, making this approach inefficient for modification-rich tRNAs. Other efforts have used direct nanopore-based sequencing for mapping modifications in long RNAs[9−13] and sequencing tRNAs[14] but suffered from high error rates.[1,5] Furthermore, RNA samples often contain coexisting isoforms, molecules that are nominally of the same RNA sequence but that have different compositions. These arise from partial

nucleotide modifications or alterations, those present in less than 100% of the molecules of a given RNA sequence. Quantifying the stoichiometries of these site-specific partial modifications remains challenging.[15]

Mass spectrometry (MS), especially tandem MS combined with liquid chromatography (LC−MS/MS), is widely used for RNA modification analysis because it does not require cDNA synthesis and is not limited to specific modifications. However, typical LC−MS/MS methods require decomposition of RNA samples to the nucleoside level prior to MS analysis,[16−18] which destroys nucleotide location information. LC−MS/MS−based mapping methods do not require full RNA degradation but must incorporate complementary methods
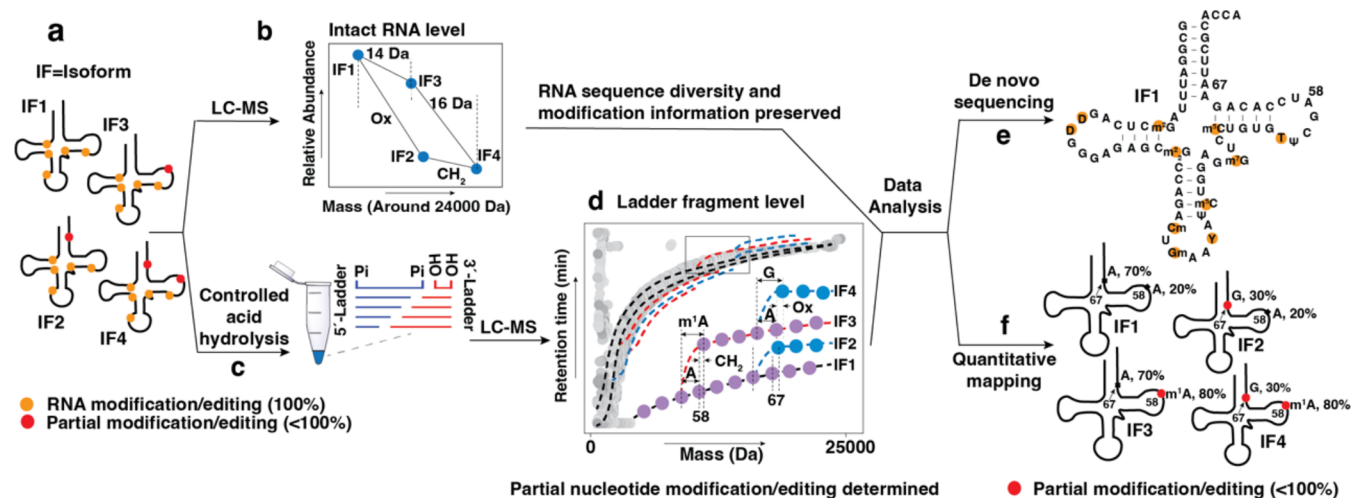
Figure 1. Schematic illustration of MLC-Seq workflow for de novo sequencing of tRNAs and site-specific quantification of RNA modification stoichiometries. (a) A tRNA sample for MLC-Seq: ∼10% is analyzed intact, while the remaining 90% is subject to controlled acid hydrolysis. (b) Direct observation of partial nucleotide modifications or editing at the intact level. For example, a 14 Da mass difference between tRNA isoforms IF1 and IF3 suggests a partial methylation (CH_2), while a 16 Da mass difference between IF3 and IF4 indicate a potential partial editing A-G, with one oxygen atom difference denoted as "Ox". (c) Controlled acid hydrolysis coverts the intact RNA into two series of ladders (5′ and 3′) composed of a series of fragments of varying lengths for MS ladder sequencing. (d) After acid degradation, the 5′ and 3′ ladders display sigmoidal curves on a t_R-mass plot, where branches in the plot indicate the position and types of partially modified or edited nucleotides. The 14 Da difference between the branched ladder fragment and original one at the same position indicates partial methylation. The branch's starting point identifies the partial modification's location at position 58 of the tRNA. (e) De novo base calling of the complete sequence of tRNA isoform IF1, using novel algorithms that identify and separate each tRNA species or isoform's MS ladders from LC−MS data. (f) Site-specific quantification of stoichiometry for partial tRNA nucleotide modifications and editing using data from both intact and ladder levels. EIC peak areas of each fragment indicates the stoichiometry ratio, e.g., m^1A vs A at position 58 of the tRNA (d). Ladder level quantification aligns with relative abundances at the intact level, confirming initially observed modifications or editing (b). More details can be found in the Supporting Information (SI) and Materials and Methods section.

(previous analyses or NGS) for location information,[19] and the complex spectra mapping data is difficult to interpret.[17] This limits its application to the analysis of relatively short (∼17 nt[1]) RNA oligonucleotides; LC−MS/MS is therefore rarely used for de novo (without prior sequence information) RNA sequencing.

LC−MS-based "ladder" sequencing methods show greater promise for de novo sequencing and site-specific quantification of nucleotide modifications.[20−23] They use controlled enzymatic or chemical degradation to randomly cut a single phosphodiester backbone bond in each RNA strand to generate a "ladder" composed of fragments[24−26] that are analyzed by LC−MS. The resulting retention time (t_R)-mass data creates a typical sigmoidal curve from which one can determine the identity, location, and stoichiometry of each nucleotide. However, this method requires a complete RNA ladder to obtain a complete sequence;[20,21,27] if the ladder is missing any fragments (a common result of, e.g., inadequate RNA sample amount, LC separation, or MS signals from low-abundance isoforms), the resulting sequence will have gaps, leading to incomplete sequencing. These rigorous sample preparation requirements, together with the need for higher sample loading (∼1000 pmol[23]) for tRNAs, short read lengths (<35 nt per run[21]), and low throughput (a highly purified single RNA sequence or only a few sequences[20,21,26]), have restricted MS-based de novo sequencing applications to lower-complexity RNA samples, making it difficult to sequence and quantify nucleotide modifications from limited amounts of heterogeneous tRNAs enriched from cells.

To systematically address the drawbacks of previous MS-based methods, we have developed a MS-based sequencing

approach, designated MLC-Seq, that addresses the shortcomings of NGS- and MS-based sequencing methods in sequencing modified nucleotides. It retains the advantages of LC−MS ladder sequencing—the ability to de novo sequence any nucleotide, modified or not, while preserving position information—while circumventing the perfect ladder requirement that might otherwise prevent complete sequencing of cellular tRNAs. As a result, MLC-Seq can perform de novo sequencing of full-length heterogeneous cellular tRNAs, including all modifications, while also identifying site-specific stoichiometry of partial modifications and the sequence of every adequately expressed tRNA in a sample. It can also be combined with reference sequence data to provide quantitative analysis of total tRNA samples containing many different tRNA sequences. As the most comprehensive method developed to date for direct sequencing and quantitative mapping of RNA modifications, MLC-Seq can decipher the true informational content of RNA and track stoichiometry changes in various cellular contexts. Such a tool will aid the RNA community in understanding how site-specific modifications control tRNA functionality,[28] contribute to disorders such as breast cancer,[29] type-2 diabetes,[30−32] and obesity,[32−35] and affect rapidly mutating viruses such as SARS-CoV-2, whose RNA contains over 40 nucleotide modifications with unknown identities or functions.[36]

RESULTS

Overview of MLC-Seq. Figure 1 outlines the MLC-Seq process, where ∼10% of a sample is analyzed intact without any treatment to preserve RNA sequence diversity and modification information (Figure 1a,b), while the remaining
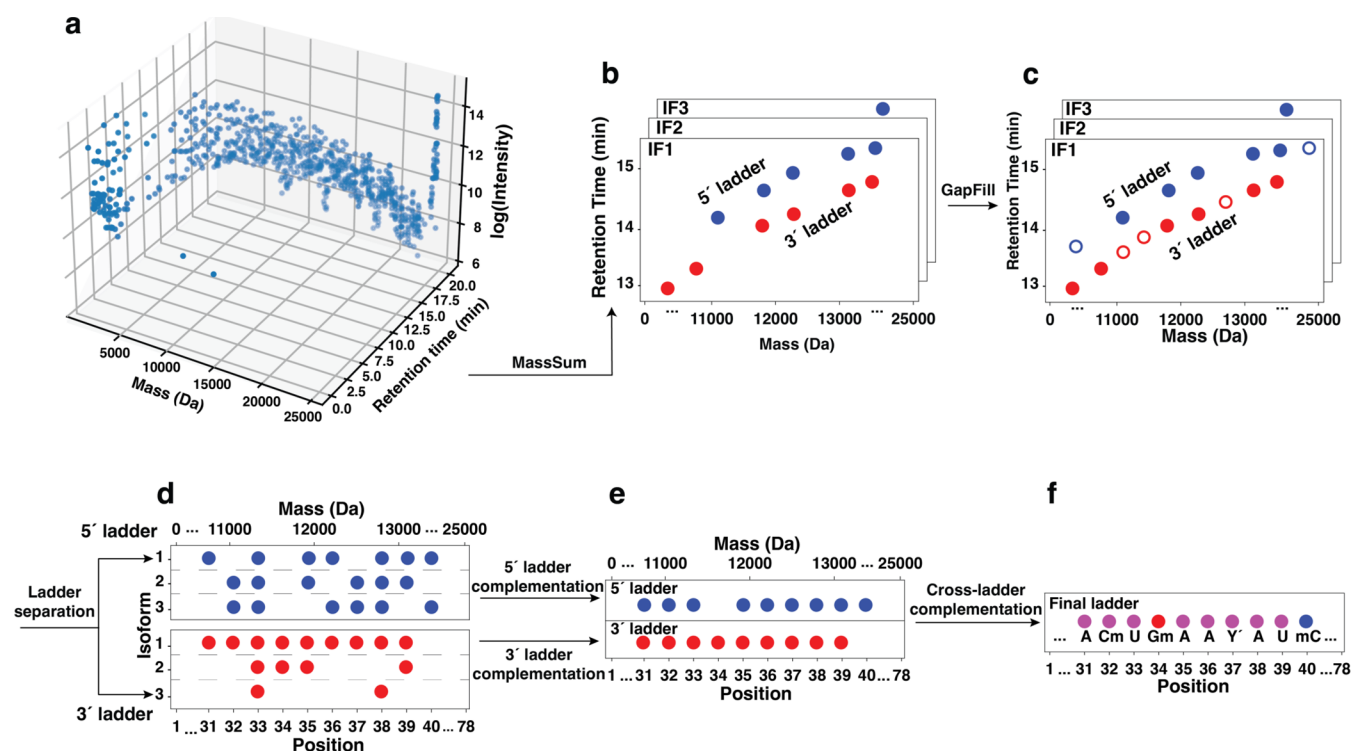
Journal of the American Chemical Society
pubs.acs.org/JACS
Article

Figure 2. MLC-Seq De novo sequencing of yeast-derived tRNA-Phe using algorithms of homology search, MassSum and ladder complementation. (a) Homology search results before acid degradation. Here, C and A indicate that the mass difference between two related tRNA isoforms, 305.0413 and 329.0525 Da, can be attributed to the nucleotides C and A, respectively. "Ox" indicates that the mass difference (15.9949 Da) between two related tRNA isoforms can be from an oxygen atom. (b) Illustration of the principle that enables MassSum to function: after a controlled single cleavage of a phosphodiester bond in the RNA strand, the combined mass of the two resulting fragments must be equal to the mass of the original RNA plus the mass of a water molecule. (c) Example of 5′ and 3′ ladders and the use of MassSum to identify pairs of linked fragments. (d) Sequencing of yeast-derived tRNA-Phe, using LC−MS deconvoluted data and the MassSum algorithm to identify ladder fragments of Phe IF1 with an intact monoisotopic mass of 24252.2692 Da. Together with other MLC-Seq algorithms, MassSum separates ladder fragments of each tRNA isoform, allowing for complete de novo sequencing. Inset shows the masses of the intact isoforms and their relative abundances. (e) The identified nucleotides for five tRNA-Phe isoforms after ladder complementation. The depth information on the bottom indicates the number of times each nucleotide was read independently from the tRNA isoform ladders. (f) The complete sequences for the five isoforms, along with a reference sequence obtained in a previous work.[21] The nucleotides highlighted in red differ from the full-length isoform Phe IF2. The sequencing result of tRNA-Phe reveals partial nucleotide modifications/editing as well as terminal truncations for five different isoforms.

90% is subjected to controlled acid hydrolysis, which converts intact RNA into ladder fragments (Figure 1c). Like our previous MS-based sequencing work,[20−23] MLC-Seq relies on a set of ladders (one each in the 5′ and 3′ directions) to

determine the identity, location, and any modifications for each nucleotide in RNA. Each ladder contains a series of fragments, each containing one more nucleotide than the previous fragment in the ladder (Figure 1c). The mass

Figure 3. Workflow for processing of MLC-Seq data. (a) Initial data after deconvolution of the LC−MS raw data file. (b) Initial partial ladders after application of the MassSum algorithm. (c) GapFill identifies additional nucleotides (empty circles) in each ladder missed by the MassSum data separation step when neither half of a 5′/ 3′ fragment pair was found. (d) Data is separated into distinct 5′ and 3′ ladders for each isoform. (e) Ladder complementation combines partial results across multiple 5′ or 3′ ladders, filling sequence gaps in one isoform using ladder fragment data from a related one. (f) Cross-ladder complementation, based on 5′ and 3′ ladders being the reverse of each other, can be applied to any remaining sequence gaps to obtain a complete sequence, a complete workflow is detailed in Figure S1.

differences between successive fragments functions as a general identifier for base calling nucleotides, including modified ones, while the ladder itself gives the order of nucleotides based on their increasing masses and t_R values.

Our previous MS-based method, however, could not directly sequence full-length tRNA-Phe. The tRNA-Phe sample contained five isoforms, labeled Phe IF1- IF5 (Figure 2), that cannot be physically separated and differ from each other by as little as a single nucleotide or modification. Furthermore, none of these isoforms yielded a perfect MS ladder (Figure 2d), which was required by previous MS sequencing. MLC-Seq addresses the "imperfect ladder" problem by combining ladder fragments from coexisting isoforms of the same specific tRNA species. Missing fragments from one isoform ladder can be filled based on other isoform data to obtain complete ladders, allowing for direct sequencing of the full length tRNA-Phe.

To obtain complete sequences, MLC-Seq incorporates a series of innovative bioinformatic tools to systematically process the MS data. First, a homology search algorithm developed in house measures the masses of all intact RNAs before and after acid degradation and groups related species together for sequencing based on known mass differences matching a nucleotide or modification. Thus, the homology search reveals not only the number of isoforms but how they differ from each other. The tRNA-Phe data in Figure 2a indicates a 3′-terminal truncation, with Phe IF2 and Phe IF4 having the longest sequences and the others losing one (Phe IF1 and Phe IF3) or two (Phe IF5) nucleotides at the 3′end. Isoform pairs with a mass difference of ~16 Da indicate a possible partial A-G transition, as their compositions differ by

one oxygen atom. These masses are also compared to the masses of intact tRNA molecules found in acid-degraded samples (see Figures S1−S3 and Table S1). The values will often be identical; any change in mass indicates the presence of acid-labile nucleotide modifications. This particular sample shows a mass decrease of ~358.14 Da for all isoforms after acid hydrolysis. This is consistent with a single wybutosine (Y) nucleotide, which converts under acidic conditions to its depurinated ribose form (Y′) (Figure S2b). This step does not reveal the location of any acid-labile nucleotide modifications, but confirming the presence (or absence) of such modifications is useful when subsequently processing MS data, identifying fragments, and constructing the ladder, which ultimately provides the location of all nucleotides, acid-labile or not.

While tRNA isoforms cannot be physically separated, the complex LC−MS data can be computationally separated into different isoforms and ladders. A novel MassSum algorithm was developed that utilizes the fact that the combined mass of any set of paired fragments generated by each single cleavage of a phosphodiester bond in the acid-mediated hydrolysis step is constant and equal to the sum of the mass of the intact RNA plus the mass of a water molecule[26] (Figure 2b). Paired 5′ and 3′ fragments originated from the same RNA sequence can therefore be isolated from the MS data because each distinct full-length RNA sequence has a unique and constant mass (Figure 2c,d).

MassSum requires the presence of two paired 5′/ 3′ fragments generated by each single cleavage of a phospho-diester bond in the acid-mediated hydrolysis step to be
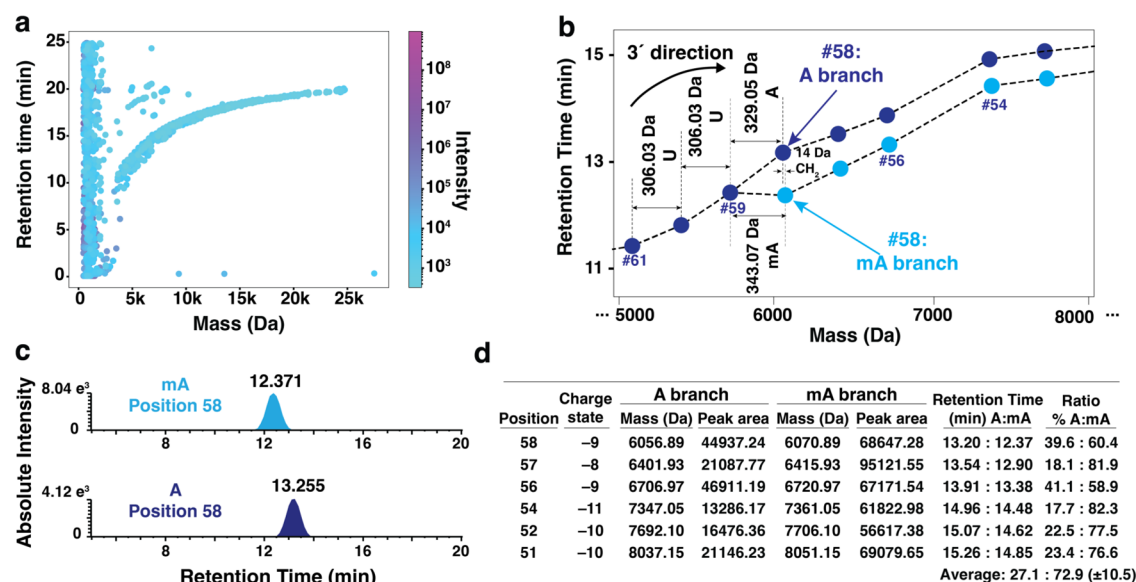
Figure 4. Quantifying stoichiometries for partial nucleotide modifications/editing. (a) t_R-mass data of tRNA-Glu output from LC−MS raw data. (b) Close-up view of t_R-mass data for tRNA-Glu showing a branch in the ladder at position 58 revealing a partial modification or alteration at this position; the mass difference of ~14 Da between the two branches specifically indicates a partial methylation. A 3′-ladder fragment is missed at position 55 because the 2′-O methyl group of m⁵Um at position 54 of tRNA-Glu blocks formic acid hydrolysis of the phosphodiester bond between these two positions. (c) EICs of 3′-ladder fragments containing mA (top, light blue, negative charge states 8−9) and canonical A (bottom, dark blue, negative charge states 9−10) at postbranch position 58. Only the peak areas of ladder fragments with single identical charge state (−9) were used to calculate the stoichiometry of canonical A and mA at this position. (d) Table of subsequent ladder fragment masses, integrated peak areas, retention times, and compositions from the partial m¹A methylation at position 58.

| | | A branch | | mA branch | | | Ratio |
Position	Charge state	Mass (Da)	Peak area	Mass (Da)	Peak area	Retention Time (min) A:mA	% A:mA
58	−9	6056.89	44937.24	6070.89	68647.28	13.20 : 12.37	39.6 : 60.4
57	−8	6401.93	21087.77	6415.93	95121.55	13.54 : 12.90	18.1 : 81.9
56	−9	6706.97	46911.19	6720.97	67171.54	13.91 : 13.38	41.1 : 58.9
54	−11	7347.05	13286.17	7361.05	61822.98	14.96 : 14.48	17.7 : 82.3
52	−10	7692.10	16476.36	7706.10	56617.38	15.07 : 14.62	22.5 : 77.5
51	−10	8037.15	21146.23	8051.15	69079.65	15.26 : 14.85	23.4 : 76.6
						Average:	27.1 : 72.9 (±10.5)

identified. A GapFill algorithm was developed to subsequently fill in fragments in the 5′ or 3′ ladders missed by the MassSum data separation step when neither half of a 5′/ 3′ fragment pair was found (Figure 3b). 5′ and 3′ ladders can be computationally separated from each other based on the sigmoidal curve that each 5′ and 3′ ladder displays in the t_R-mass plot (Figure 3c). When a perfect 5′ or 3′ ladder exists for a sequence, it can be read out via base calling of each nucleotide or modification as described in previous literature.[20−22] If ladders are still missing fragments, they can be combined with ladders from coexisting related isoforms of the same tRNA group. Fragments missing from one tRNA isoform may be complemented by counterpart fragments from a related tRNA isoform (Figure 3d,e). Gaps in a 5′ ladder can also be filled based on reversing the corresponding 3′ ladder, and vice versa (Figure 3f).

MLC-Seq was first applied to sequence the full-length tRNA-Phe (Figure 2), resolving the "imperfect ladder" problem and allowing for direct sequencing of all five tRNA isoform sequences, including three less abundant ones (Phe IF2,3,4 and 5; Figure 2a). The results are consistent with reported tRNA-Phe sequence and simultaneously reveal various modifications.[21] The reported tRNA-Phe sequence with uridine (U) in the sixth position and adenosine (A) in the 67th position was identified. However, the tRNA-Phe sequence with cytidine (C) in the sixth position and guanosine (G) in the 67th position, predicted by yeast genome,[43] was not observed in the intact and acid hydrolyzed sample, likely due to its presence being below the limit of detection of the mass spectrometer (Table S1).

Isomeric nucleotides can be distinguished by applying an extra step prior to or following MS sequencing.[21] Pseudouridine (Ψ) can be distinguished from uridine (U) with N-cyclohexyl-N′-(2-morpholinoethyl)-carbodiimide metho-p-tol-

uenesulfonate (CMC), as the CMC adducts of Ψ and U differ in mass. Methylated positions can be identified using AlkB, a demethylating enzyme whose reactivity is selective based on the methylated nucleotides. Additionally, 2′-O-methylations can often be identified based on the fact that methyl groups in this position tend to block the formic acid hydrolysis of the phosphodiester bond. Thus, these fragments do not appear in the MS data, resulting in ladders with a missing fragment. This can be observed in Figure 4b, where a missing fragment indicates the presence of a 2′-O-methylation.

Preserving RNA Sequence Diversity and Modification Information. RNA samples often contain multiple types of RNA molecules corresponding to different RNA sequences or isoforms, and a key advantage of MLC-Seq is its ability to provide detailed results regarding sample diversity. It can distinguish between sequences, even those that differ by only a single nucleotide or modification, and quantify the relative ratio of different RNA sequences and further stoichiometry of partial modifications in the RNA sample. This is aided by advanced analytical instrumentation, including a Vanquish Horizon UHPLC coupled with an Orbitrap Exploris 240 MS (Thermo Fisher Scientific) that can measure the mass of intact tRNA up to ~25 kDa without the need for a T1 digestion step. Partial T1 digestion breaks full tRNA sequences into smaller fragments (~35 nt), which complicates the data analysis but would be necessary for instruments with a lower mass resolution.[21] MLC-Seq also reduces the necessary sample loading by 3 orders of magnitude, to roughly 20 ng (1 pmol of tRNA), which allows for sequencing of heterogeneous cellular tRNA.

The ability to measure the mass of intact full-length tRNAs makes it possible to observe differences in intact RNA molecules, indicate the existence of partial RNA modifications/editing, and further identify types, locate sites and
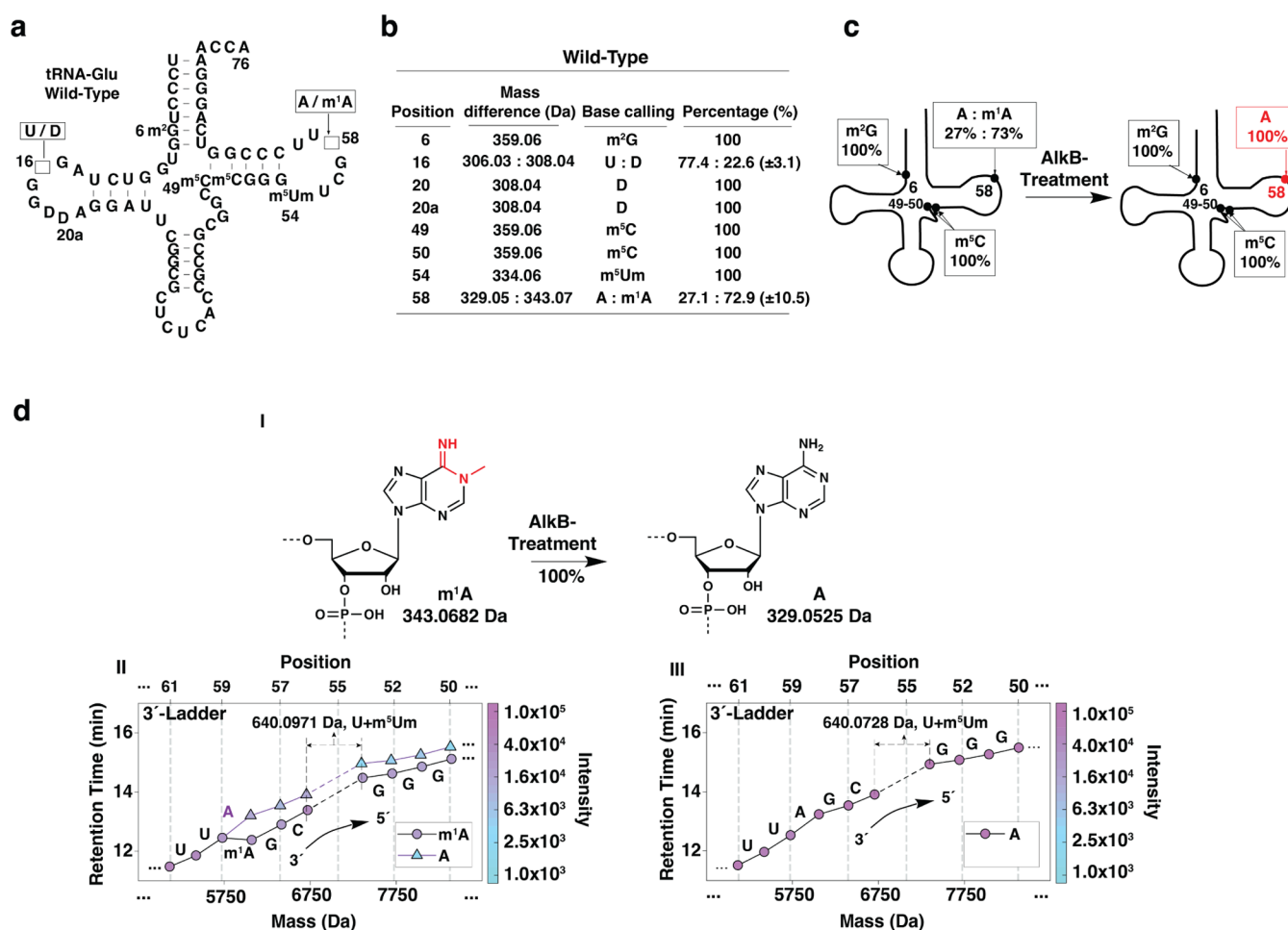
Figure 5. Using MLC-Seq to de novo sequence tRNA-Glu derived from mouse liver and track modification changes after AlkB treatment. (a) Complete sequence of tRNA-Glu. (b) Position and stoichiometry of the noncanonical nucleotides and partially modified nucleotides in tRNA-Glu. (c) Stoichiometric changes of m^{1}A at position 58 of the tRNA-Glu after AlkB-treatment. The percentage of m^{1}A is reduced from 73% to 0% at the position where it is demethylated completely and becomes 100% A. (d) I. Structural change of m^{1}A to A at position 58 caused by AlkB treatment. II–III. The disappearance of the branch of the methylated ladder in 2D t_R-mass plot caused by AlkB treatment. The 2-O′-methyl of m^{5}Um at position 54 of the tRNA-Glu blocks the acid degradation, causing a small mass gap with a ladder fragment of 640.0819 Da (the mass sum of likely m^{5}Um and U) missing between positions 54 and 55.

quantify stoichiometries of these RNA modifications/editing. This can be observed in Figure 4 using data from sequencing tRNA-Glu (which is discussed in further depth in the following section). After acid degradation, the 5′ and 3′ ladders for tRNA-Glu form sigmoidal curves on a t_R-mass plot (Figure 4a). A branch in this curve indicates a partial modification at this position, as illustrated in Figure 4b.[20,21,26] The mass difference between the branches indicates the nature of the partial modification; the 14 Da difference here indicates a partial methylation, where a canonical nucleotide in the lighter isoform is replaced in the heavier isoform by a methylated form of that nucleotide. Meanwhile, the position at which the branch begins can pinpoint the location of the partial modification in the sequence. Quantification of partial modifications can be calculated using the extracted ion chromatograms (EICs) from LC−MS raw data for each fragment[20] (Figure 4c); the ratios of the EIC peak areas correspond to the stoichiometry of m^{1}A vs A at position 58 of tRNA-Glu (Figure 4d). Previous results have shown that for sufficiently and equally long RNA (>7 nt), the addition of a single nucleotide modification does not significantly affect instrument response and that the ratio of EIC peak areas

corresponds directly to the stoichiometry of partial modifications.

Direct Sequencing of Full-Length tRNA and Quantitative Mapping of Multiple Modifications. After verification on yeast tRNA-Phe, MLC-Seq was used to sequence a tRNA enriched from mouse liver, tRNA-Glu (Figure S4). MLC-Seq was able to obtain complete sequence information for the tRNA, including the identity and position of all nucleotides (modified and canonical) and the site-specific stoichiometry of partial modifications and editing, marking the first instance of direct sequencing of full-length tRNAs with this level of detail. This is especially noteworthy as the modification-rich nature of tRNA has frustrated sequencing efforts in the past. The complete sequencing results of tRNA-Glu are shown in Figure 5.

The MLC-Seq obtained sequence for tRNA-Glu is almost identical to that reported previously,[37,38] although there are some significant differences in nucleotide modifications. It detected partial modifications at positions 16 and 58 of tRNA-Glu (Figure 5a,b). Position 16 was partially hydrogenated, containing 77.4% U and dihydrouridine (D) in the remaining 22.6% (±3.1%). MLC-Seq found m^{1}A (1- methyladenosine)

Figure 6. MLC−Seq analysis of yeast total tRNA and modifications. (a) Results of LC−MS analysis for intact tRNA molecules; orange points represent a match found between the data and a published tRNA sequence. Out of the 129 masses, only 64 match the calculated tRNA masses from the tRNAdb 2009 database and are highlighted in orange (Table S2). (b) Homology search results indicating partial 3′-end truncations; molecules with mass differences of 329.05 or 305.04 Da indicate that their structures differ by an A or C nucleotide, respectively. (c) Homology search results indicating partial methylations; molecules with mass differences of 14 Da indicate that their structures may differ by a methyl group. (d) Percentages of tRNA molecules across all subtypes possessing −CCA, −CC, and -C 3′ ends. (e) Structures and mass data for amino acid-charged tRNA molecules, indicating that MS analysis can effectively identify complete tRNA molecules charged with amino acids. (f) Partial sequences can be read out by ladder fragments computationally separated by the MassSum algorithm, utilizing the intact masses of a specific tRNA-AspGUC isoform (23904.22 Da). (g) Using BLAST to search for the partial sequences and related positions obtained from Figure 6f (shown in boxes) returns a positive result, validating the data and indicating the tRNA subtype that these fragments belong to. (h) Table of nucleotides confirmed through cross-referencing between LC−MS data and reference sequences, and a list of symbols used by BLAST and their corresponding nucleotide modifications. Shaded cells indicate nucleotides that were confirmed by 5′ and/or 3′ ladder fragments; cells with borders indicate nucleotide modifications that were confirmed by 5′ and/or 3′ ladder fragments. Colors in positions 74−76 indicate the relative amounts of each tRNA subtype with −CCA, −CC, and -C 3′ ends and in other positions indicate the stoichiometry of each modification at the site. For better visualization, the stoichiometry of partial RNA modifications and 3′ terminal truncations was represented using a six-color scale at intervals of 0, 20, 40, 60, 80, and 100. Quantification was limited to partial 3′-end truncations and partial modifications that were confirmed by at least two newly branched ladder fragments after the modification site. Additional details can be found in Figures S6−S13 and Tables S2−S4.

coexisting with A (adenosine) (72.9% m^{1}A and 27.1% A, ± 10.5%) at position 58, which was previously thought to contain only A. Additionally, it determined that position 6 contained methylated G (guanosine) and position 10 contained canonical G (these were reversed in previous studies) and that the nucleotide modification at position 20a is D rather than the previously reported Ψ. Additional information on site-specific quantification of partial modifications in the tRNA can be found in Materials and Methods section and Figure S5.

Tracking Stoichiometric Changes in RNA Modifications. A powerful application of MLC-Seq is tracking changes in RNA modifications. These changes can be caused by diseases or cellular disturbances whose progression can be tracked through monitoring the identity and stoichiometry changes in RNA nucleotides (modified or canonical). To examine this, the mouse liver RNA sample was treated with AlkB to leverage its selectivity toward specific isomeric methylated nucleotides (Figure 5c,d).[39,40] AlkB reacts with m^{1}A, m^{1}G, and m^{3}C (converting them to their respective canonical nucleotides) but is inert toward m^{6}A, m^{2}G, and m^{5}C. Measuring site-specific changes in modification stoichiometry can verify AlkB's reactivity toward specific methylated nucleotides while demonstrating the capacity of MLC-Seq to quantitatively track changes in tRNA samples site-specifically at the single-nucleotide level.

Position 58 of wild-type (i.e., pre-AlkB treatment) tRNA-Glu had a 73:27 ratio of m^{1}A to A but comprised 100% A after AlkB treatment (Figure 5b−d). Other types of methylated nucleotides in the sequence saw little to no change as a result of AlkB. This validates MLC-Seq as a way of precisely measuring site-specific RNA modification dynamics, a quality that is currently difficult to study quantitatively. To clarify, AlkB treatment itself does not quantify ratios but distinguishes m^{1}A from other modifications at position 58 of tRNA-Glu. We determined the proportions of modified m^{1}A vs canonical A at this position by integrating EIC peak areas of their corresponding ladder fragments at the identical positions (Figure 4; see Materials and Methods section), a standard method in LC−MS for relative quantification.[21,41,42]

Site-Specific Quantitative Mapping of Total tRNA Modifications. The de novo sequencing of tRNA-Glu was conducted using refined tRNA samples. However, MLC-Seq can also be used for quantitative analysis of total tRNA samples containing all tRNA subtypes. Portions of each tRNA sequence can be read out de novo using the previously described methods, although the complexity of the LC−MS data for such a diverse sample currently prevents complete sequencing. Instead, the method is combined with known reference sequences to identify each tRNA ladder fragment from LC−MS data and construct ladders for tRNA sequence and modification analysis.

To manage the LC−MS data complexity of controlled acid-hydrolyzed yeast total tRNA samples (Figure S6), the manuscript employs three functionalities of the MLC-Seq platform, including: (1) de novo sequencing (without sequence input) to identify tRNA types present in the samples, (2) site-specific mapping of tRNA modifications by cross-referencing between tRNA database sequences and the LC−MS data[43] and (3) site-specific quantification of partial tRNA modification stoichiometry. Different standards for ladder fragments are applied for each function. For de novo sequencing, a minimum of two successive ladder fragments is required for base calling any nucleotide or modifications. For

mapping, ladder fragments are selected when their observed monoisotopic masses match those calculated from the corresponding database sequence. For stoichiometric quantification of partial tRNA modifications, at least two successive branch ladder fragments are required to both confirm the modification and determine the modification's stoichiometry. Additional details about total tRNA analysis are available in the Materials and Methods section.

LC−MS analysis of intact molecules (Figure 6a and Table S2) provides information on the number of different tRNA sequences in the sample; orange dots in this figure indicate that a match was found between the LC−MS data and the mass of a known tRNA sequence; these results can also be used to identify possible partial sequence modifications/editing such as 3′-end truncations (Figure 6b) or methylations (Figure 6c). A total of 64 matches were made to known tRNAs in a database[37] (Figures 6a, S6−S13, and Tables S2), although it should be emphasized that this is a limitation of the reference data rather than the method itself.

Figure 6d summarizes the portion of molecules showing −CCA, −CC, and -C 3′ ends. This further demonstrated the capacity of MLC-Seq to precisely quantify partial modifications and truncations, although it is possible that the 3′-end truncations observed are not inherent to the tRNAs but rather are the result of sample extraction and preparation.[21] This is supported by the presence of the intact 3′ CCA in tRNA-Glu extracted from mouse liver, shown in Figure 5a. Intact molecule analysis can also indicate the presence of amino acid-charged tRNA molecules (Figure 6e).

The sample is then subject to controlled acid degradation using the previously described methods. MassSum offers a relatively straightforward and largely automated process for finding matching 5′ and 3′ fragments and facilitating the identification of which tRNA subtype they correspond to. Partial sequences can be constructed de novo from this data (Figure 6f) and used in the Basic Local Alignment Search Tool (BLAST)[44] to cross-reference between databases of known sequences and the LC−MS data (Figure 6g). Matches between the two can confirm the sequences of all tRNA subtypes in the sample (Figure 6h and Table S2), while data points not matching the known sequences may indicate partial modifications that can be further identified and pinpointed based on their intensity and mass difference relative to known sequence values. This method allows for thorough analysis and quantification of multiple RNA sequences in samples, including those of highly complex total tRNA, while also offering site-specific identification and quantification of partially modified or edited nucleotides. It was able to reveal several new partial and complete modifications for multiple tRNA subtypes. For example, we identified a partial m^{2}G methylation in tRNA-LysCUU position 9 (72 ± 16% coexisting with canonical G), a partial *Cm* modification at tRNA-ProNGG position 4 (68 ± 13%, coexisting with canonical C), and partial modified A at position 58 of both tRNA-Arg1CU (20 ± 6%) and tRNA-AsnGUU (18 ± 6%). Further details on the partial nucleotide modifications in the total yeast tRNA sample—including identity, location, stoichiometry, as well as specific tRNAs and isoforms—are available in Tables S3−4 and Figures S8−S13.

■ DISCUSSION

MLC-Seq is a significant advance over previous de novo MS sequencing methods, which require a complete ladder to

identify each nucleotide and its position in the sequence. This is particularly challenging when sequencing low-abundance tRNA species or isoforms whose ladder fragments may only be present in quantities below the instrument detection limit. MLC-Seq circumvents the need for a perfect ladder, allowing MS sequencing to be performed for a broader range of RNA samples. This method surpasses conventional MS/MS-based analytical strategies, which often require prior sequence input, and provides an unbiased approach for comprehensive de novo direct sequencing of tRNA modifications, a challenge for other sequencing methods. Contrary to traditional MS sequencing which necessitates a highly purified homogeneous short RNA sequence (<~30 nt long), MLC-Seq has the capacity to sequence each full-length tRNA type even within complex heterogeneous samples by segregating their MS ladder fragments. Algorithms in MLC-Seq can incorporate partial ladders with missing fragments, identify each RNA in a mixture, and computationally separate MS ladder fragments for sequencing of each tRNA, including those whose scarcity would prevent the acquisition of a perfect ladder.

However, for complex mixtures such as total tRNA where incomplete ladders cannot be fully fixed, incorporating known tRNA reference sequences from databases during the sample analysis may be necessary.

Additionally, MLC-Seq allows simultaneous quantitative mapping of all nucleotide modifications with single-nucleotide stoichiometric precision, potentially enabling quantitative mapping of tRNA modifications in a tissue-specific manner. The presence of a branch point in a sigmoidal fragment ladder curve on a t_R—mass plot is indicative of a partial modification/editing. MLC-Seq can monitor changes in RNA modification dynamics and map modifications that have altered stoichiometry in different cellular and disease contexts. In contrast to cDNA-based RNA sequencing, which removes information on modifications, MLC-Seq preserves information regarding tRNA sample diversity (for visualizing each tRNA) and modification (for revealing modification type, location, stoichiometry, etc.).

It is notable that tRNA samples were typically mixtures of coexisting heterogeneous cellular tRNA sequences and isoforms that cannot be physically separated. We believe that, contrary to common belief, it is not possible to obtain a single "pure" tRNA sequence, since there are always single-nucleotide sequence variations at different loci. This resonates with previously proposed ideas about the emerging complexity of the tRNA world.[28] This makes MLC-Seq convenient for tRNA samples containing multiple sequences or isoforms while remaining effective for single RNA species. It can also be used for quantitative analysis of unrefined total tRNA when used in conjunction with a sequence database, providing site-specific information on the stoichiometry of partial tRNA modifications/editing.

The ability to computationally separate LC−MS data from different sequences or isoforms in a single sample increases the throughput of de novo MS-based RNA sequencing and facilitates large-scale MS sequencing of biological samples. By incorporating advanced instrumentation and greater automation, MLC-Seq could match the capacity of classic Sanger sequencing, making it possible to directly sequence RNAs and diverse modifications at a large scale. These methods could also be used to sequence longer RNAs (e.g., rRNA, snoRNA, snRNA, Y RNA, and vault RNA) and small noncoding RNAs (e.g., miRNA, piRNA, tsRNA, rsRNA, and ysRNAs)[5] and as an orthogonal approach to verify results from high-throughput sequencing methods.

■ CONCLUSIONS

MLC-Seq addresses incomplete ladder issues in MS sequencing of cellular tRNAs. It identifies tRNA sequences and their variants, providing a powerful approach to sequence full-length tRNAs, including their modifications, even in the presence of other tRNAs. MLC-Seq effectively reveals nucleotide identities, partial modification stoichiometry of multiple tRNA modifications, and their changes after AlkB treatment. MLC-Seq can serve as a general sequencing method for revealing true RNA sequences and quantitatively studying RNA modifications. Such a tool is urgently needed in RNA research and will provide new insights and solutions for the study of COVID-19, metabolic disorders, and other diseases.

■ MATERIALS AND METHODS

Reagents and Chemicals. All chemicals were purchased from commercial sources and used without further purification. Diisopropylamine (DIPA > 99.5%) and tRNAs (specifically phenylalanine and total tRNA from brewer's yeast) was obtained from Sigma-Aldrich (St. Louis, Missouri). Formic acid (98−100%) was purchased from Merck (Darmstadt, Germany). 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP, 99%) and N,N-diisopropylethylamine (DIPEA, 99%) were purchased from Thermo Fisher Scientific (Waltham, MA). All other chemicals were obtained from Sigma-Aldrich unless indicated otherwise.

Workflow of De Novo Sequencing of tRNA Mixtures. Each tRNA sample was divided into two portions. 10% was injected directly into the LC−MS instrument, while the remaining 90% was subjected to acid degradation before LC−MS injection. After obtaining LC−MS data sets of the control and acid-degraded portions, novel algorithms (developed in house and available at https://github.com/rnamodifications/MLC-Seq) were used to identify each tRNA species or isoform and computationally separate its MS ladder fragments from the LC−MS data.

Controlled Acid Degradation of tRNA Samples. Formic acid was applied to degrade tRNA samples including yeast tRNA-Phe and mouse liver tRNA-Glu and tRNA-Gln (see "Mouse Liver tRNA Pulldown" section below) to produce MS ladders according to reported experimental protocols.[20,21,23,26] Given that a single degradation time point may not yield a complete set of ladder fragments for tRNA sequencing, our manuscript describes using (1) a combination of different time points on split samples followed by repooling, and (2) mass ladder complementation (MLC) to account for any missing ladder fragments (See "Ladder Complementation and Generation of RNA Sequences" section below). Each RNA sample solution was divided into three or more equal aliquots (each in an RNase-free thin-walled 0.2 mL PCR tube) and mixed with 50% (v/v) formic acid at 40 °C in a PCR machine. Typically, acid hydrolysis reactions were run for 2, 5, and 15 min, but additional time points (0.2, 0.5, 10, and 20 min) were included for sequencing yeast tRNA-Phe in order to obtain more complete ladders. Reaction mixtures were immediately frozen with dry ice at the specified times, followed by centrifugal vacuum concentration (Labconco Co., Kansas City, MO) to dryness. The dried samples from all time points were then usually combined and resuspended in 20 μL of nuclease-free water for the LC−MS measurement.

LC−MS Measurement and Processing of Intact and Acid-Degraded tRNA Samples. Each intact or combined acid-hydrolyzed tRNA sample was individually analyzed on an Orbitrap Exploris 240 MS (Thermo Fisher Scientific, Bremen, Germany) coupled to a Vanquish Horizon UHPLC using a DNAPac reversed-phase (RP) column (2.1 mm × 50 mm, Thermo Fisher Scientific, Sunnyvale, California). Different solvent systems were employed in the separation of the samples. One solvent system used 2% HFIP and 0.1% DIPEA as eluent A pH 9 and methanol, 0.075% HFIP, and

0.0375% DIPEA as eluent B pH 9. The other solvent system used 1.7% HPIF and 0.15% diisopropylamine (DIPA) as eluent A, and 50% methanol with 1.7% HFIP and 0.15% DIPA as eluent B. Several different gradients were also used to separate both the intact and acid hydrolyzed samples. Sorter gradients were used for intact samples; whereas a longer gradient was used for formic acid hydrolyzed samples for better sample separation. A gradient of 20 to 80% B over 6.7 min or 30 to 80% B over 22 min was used for the analysis of intact RNA samples, while a gradient from 15 to 35% B over 20 min or 20 to 40% B in 19 min was used for acid-degraded samples. The flow rate was 0.2, 0.3, or 0.4 mL/min, and the separations were performed with the column temperature maintained at 60 or 70 °C. Injection volumes were 3−25 μL, and sample amounts were 2−200 pmol of tRNA-Phe and 1 pmol (∼20 ng) of tRNA-Glu. tRNA samples were analyzed in negative-ion full MS mode from m/z 410 to 3200, m/z 550 to 2000, or m/z 600 to 2000, to obtain deconvoluted mass in the low or high region, with a scan rate of 2 spectra/s at 120k resolution and m/z = 200.

At least three samples, either intact or acid hydrolyzed, were repeated for LC−MS runs and analysis. The resulting LC−MS.raw files were deconvoluted and converted to Excel (.xlsx) files using Thermo BioPharma Finder 4.0−5.2 (ThermoFisher Scientific). The process underwent Intact Mass Analysis mode using the Xtract (isotopically resolved) deconvolution algorithm as previously described.[20,21,23,26]

Homology Search. Once LC−MS data are displayed as a two-dimensional (2D) t_R-mass plot, a homology search of intact tRNAs can be conducted in the monoisotopic mass range of >∼24k Da using an in-house algorithm in Python (see GitHub). This algorithm identifies related tRNA isoforms that may share the same ancestral precursor tRNA but are different in absolute sequence, e.g., in posttranscriptional profiles of nucleotide modifications, editing, and truncations. Mass differences between two intact tRNA isoforms are calculated and matched to the known mass of nucleotides or nucleotide modifications in the database.[26] For example, a difference of 14.0157 Da (±10 ppm)[45] can be assigned to a methylation (Me/−CH$_2$−) event, while a difference of 329.0525 Da corresponds to an additional A nucleotide. Therefore, these intact tRNAs are assigned to the same tRNA group and considered homologous isoforms of a specific tRNA for sequencing together.

The homology search is a nontarget preselection step to group possible related tRNA isoforms together for sequencing. However, only one monoisotopic mass difference is used to distinguish between two intact tRNA isoforms, which could lead to errors by including a tRNA isoform that does not belong to the specified tRNA group or omitting an isoform that does belong. These errors can be identified and corrected later when sequencing each group of tRNA isoforms, and sequencing results can further verify the interconnection between isoforms.

Detection of Acid-Labile Nucleotide Modifications. Acid-labile nucleotides are identified using another algorithm in Python (see GitHub) that analyzes the connections between the compounds (with a monoisotopic mass >24 kDa for tRNAs) measured by LC−MS before and after acid degradation. For each such compound pair, if the monoisotopic mass difference can be matched to a known mass difference corresponding to a possible structural change to a nucleotide modification during acid hydrolysis (or the sum of several such changes), the compound pair will be selected and further considered to potentially contain acid-labile nucleotide modifications. In general, if the intact mass of the RNA species does not change after acid degradation, this intact mass will be used for MassSum data separation (see below). Otherwise, the presence of acid-labile nucleotides may be indicated by matching the observed mass difference with the theoretical mass difference caused by an acid-mediated structural change in a nucleotide or a combination of several such changes (see Figure S2).

5′ and 3′ Ladder Separation. LC−MS analysis results in two different ladders, a 5′ ladder and a 3′ ladder, that can be computationally distinguished by the differences in their t_R values. This separates the data into two isolated but adjacent sigmoidal curves, one for each ladder. Due to the large number of fragment compounds, the dividing line between the 5′ and 3′ ladders is not obvious in the t_R-mass plot. Thus, we developed a computational tool (see Github) to separate the 5′ and 3′ fragments. It divides all compounds in each LC−MS data pool into two subgroup areas; compounds in the top collective curve of the t_R-mass plot are marked as 5′ fragments, while those in the bottom curve are marked as 3′ fragments. This should ideally identify as many fragments as possible while minimizing the number of fragments assigned to the incorrect ladder, although overlap between the two ladders may still occur if the t_R difference between fragments is very small. Some manual selection is also used, albeit to generate additional input fragments for the MassSum algorithm (see below) rather than as a primary means of separating 5′ and 3′ fragments.

MassSum Data Separation. MassSum is an algorithm in Python (see Github) developed based upon the controlled acid hydrolysis of RNA presented in Figure 2. MassSum takes advantage of the fact that the sum of the masses of each pair of fragments (5′ and 3′) produced from a single cut of an intact RNA is a constant value unique for each RNA isoform/species

$$\text{mass}_{3'} + \text{mass}_{5'} = \text{mass}_{\text{intact}} + \text{mass}_{\text{H}_2\text{O}} \qquad (1)$$

where mass$_{\text{intact}}$ is the intact RNA, mass$_{3'}$ and mass$_{5'}$ are the two fragment masses, and mass$_{\text{H}_2\text{O}}$ is the mass of one water molecule. This equation can be used to isolate ladder compounds corresponding to a specific isoform, which simplifies the data set by grouping MS ladder components into subsets, one for each tRNA isoform/species. MassSum operates by choosing two random compounds from the acid-degraded LC−MS data set and adding their masses; if the sum is equal to the mass of a known isoform/species, the fragments are selected into a subset corresponding to that isoform/species containing all its 3′ and 5′ ladders.

GapFill. GapFill is another Python-based algorithm (see GitHub) developed to complement MassSum, which identifies pairs of corresponding 5′/ 3′ fragments and therefore cannot separate data if, e.g., there is no 5′ fragment found in the data that pair with a given 3′ fragment. GapFill was designed to "rescue" any ladder fragments missed by MassSum separation by first identifying gaps where fragments are missing from a ladder after the MassSum algorithm was applied and the corresponding values of mass$_{\text{low}}$ and mass$_{\text{high}}$, the masses of, respectively, the heaviest known fragment below the gap range and the lightest known fragment above the gap range. The data set typically contains several fragments whose mass falls between mass$_{\text{low}}$ and mass$_{\text{high}}$, but presumably none were selected by the MassSum algorithm during data separation. GapFill iterates over each fragment LC−MS data set whose mass falls within this range and examines the mass differences between this compound and mass$_{\text{low}}$ and mass$_{\text{high}}$. If the mass difference is equal to the sum of one or more nucleotides or modifications in the RNA modification database,[26] it is noted as a connection. If the fragment in the gap has connections with both ending fragments, it is selected into a candidate pool for the subsequent sequencing process. After iteration, GapFill calculates connections of the fragments in the candidate pool and the frequency of each connection, and the fragments with the highest frequency are chosen to fill in the gap.

Ladder Complementation and Generation of RNA Sequences. After MassSum and GapFill, each tRNA isoform has its own set of separate 5′ and 3′ ladders. If any ladder is perfect (i.e., without any missing fragments), the full RNA sequence can be read, from the first to the last nucleotide in the sequence. Incomplete ladders can be completed using other related isoforms to obtain a more complete ladder for sequencing. A Python-based computational algorithm (see Github) was designed to align ladders from related isoforms based on the position of the ladder fragment in the 5′/3′ direction. For example, Figure 2e lays out the 5′ ladders for tRNA-Phe, positioned horizontally so that the nucleotide positions are aligned. Ladder complementation can be performed separately on 5′ or 3′ ladders (but not mixed ladders), resulting in one final 5′ ladder or one final 3′ ladder. Additionally, the 3′ fragments can be converted to their

corresponding 5′ fragments for each tRNA isoform based on the MassSum principle. As such, each position in an tRNA isoform could have its original 5′ ladder fragment as well as a second fragment converted from the corresponding 3′ fragment, which can be used for confirmation and/or complementation.

Stoichiometric Quantification of Partial Nucleotide Modifications/Editing. Stoichiometries of partial nucleotide modifications/alterations were quantified based on integrating EIC peaks corresponding to two or more fragments present at a single position in a sequence. EIC chromatograms were generated via BioPharma Finder 5.0–5.2 software (Thermo Scientific) using the Xtract (isotopically resolved) algorithm. In general, each EIC trace used a single m/z value corresponding to a fragment's most abundant isotope and the charge state z with the strongest MS signal; in cases where fragments at a single position had different preferred charge states, the preferred charge state for the more abundant fragment (i.e., with the greatest EIC area among all fragments of interest) was used. The ratio of EIC areas was taken as the relative abundance of their respective fragments. Each modification creates a branch in the MS ladder that is evident in all subsequent positions in the sequence, so this calculation was repeated in multiple positions for each partial modification to obtain six values that were used to calculate averages and standard deviations unless indicated otherwise.

This process needed to be modified slightly in cases where partial components at a position were close in mass to each other (mass difference of 2 Da or less). In these cases, the isotopic patterns of the fragments overlapped significantly, such that the most abundant m/z values would feature contributions from both fragments. To address this, a composite isotopic pattern was calculated and compared to the pattern obtained from MS data. This is illustrated in Figure S5 using results from position 16 of tRNA-Gln, which contains a partial U-D dehydrogenation. Theoretical single-component isotopic patterns were obtained using a Monte Carlo calculator taking 2×10^6 samples and using the following isotope probabilities: $P(^{13}C) = 0.0106$; $P(^2H) = 0.000145$; $P(^{15}N) = 0.03795$; $P(^{17}O) = 0.000385$; $P(^{18}O) = 0.002045$. In most cases, results from the first seven m/z values were sufficient to represent the entire distribution. Figure S5a shows the isotopic distributions of the obtained LC−MS data (obtained from EIC traces of each individual m/z value), a calculated distribution for U only, and an optimized composite distribution of 32.8% U and 67.2% D. Figure S5b shows the breakdown of the composite distribution and its contributions from U and D branches at each m/z value. Stoichiometries were solved using a brute-force search to determine, to the nearest tenth of a percent, the composition whose theoretical composite isotopic pattern best matched the data pattern based on minimizing the Kolmogorov−Smirnov (KS) statistic between the two isotopic distributions; this statistic was used because its value is not dependent on the test sample size, making it easier to apply to MS data where the number of "observations" is ambiguous. Figure S5c shows the cumulative probability functions for the LC−MS, U, and composite isotopic distributions, as well as the KS statistic for U only, i.e., the maximum distance between the theoretical and data-derived cumulative probability functions.

Mouse Liver tRNA Pulldown. The protocols described here were for bulk stock solution. The tRNA was enriched by an affinity pulldown assay combined with gel recovery, with modified protocols from a previous report.[46] The total RNA of mouse liver was harvested by TRIzol reagent (Invitrogen 15596026) according to the manufacturer's instructions. The concentration of the total RNA solution was adjusted to 2 mg/mL with RNase-free water. The small-RNA fraction (<200 nt) was separated in buffer containing 50% (w/v) poly(ethylene glycol) 8000 and 0.5 M NaCl solution by centrifugation at 12,000 rpm and 4 °C for 20 min. The supernatant was collected, followed by adding 1/10 volume sodium acetate (NaAc) solution (Invitrogen). One milliliter of supernatant was added to 3 mL of ethanol, and 5 μL of linear acrylamide (Invitrogen) was added to precipitate small RNAs (<200 nt) at −20 °C overnight, followed by centrifugation at 12,000 rpm at 4 °C for 20 min. The concentration of the small-RNA (<200 nt) solution was adjusted to 1 mg/mL, and 1 mL of the small-RNA solution with 6 μL of

biotinylated probe (100 μM), 26 μL of 20X saline-sodium citrate (SSC) solution (Invitrogen) and 15 μL of RNase inhibitor (NEB) was incubated at 50 °C overnight. Streptavidin sepharose (200 μL, Cytiva 17511301) was added to the hybridization solution to enrich the biotin-labeled probe captured with the targeted tRNA. After incubation at room temperature for 30 min, the streptavidin sepharose was transferred to a 1.5 mL Ultrafree-MC tube (Millipore) and washed with 0.5× SSC buffer. The washing step was repeated 5 times. Then, 500 μL of nuclease-free water was added to the MC tube and incubated at 70 °C for 15 min, followed by centrifugation at 2500g at room temperature for 1 min to elute the RNAs that are complementary to the biotinylated probe. The eluent was collected, followed by adding 1/10 volume NaAc solution. Then, 1 mL of eluent was added to 3 mL of ethanol, and 5 μL of linear acrylamide was added to precipitate RNAs at −20 °C overnight, followed by centrifugation at 12,000 rpm at 4 °C for 20 min. Nuclease-free water was added to dissolve the RNA pellets. RNA was loaded into a 7 M urea-PAGE gel for electrophoresis, and the main tRNA band was recovered from the PAGE gel as previously described[39] to obtain enriched tRNAs for MLC-Seq The DNA probe for the pull-down experiment was synthesized by Integrated DNA Technologies (IDT), and the sequence was as follows:

tRNA-Glu pulldown probe: 5′-Biotin-CTAACCACTAGACCAC-CAGGGA

tRNA-Gln pulldown probe: 5′-Biotin-TGGAGGTTCCACCGA-GATTTGA

Northern Blot. A Northern blot (Figure S4) was performed as previously described[8] to validate the captured tRNAs (as described above). RNA was separated on a 10% urea-PAGE gel stained with SYBR Gold, immediately imaged, transferred to a positively charged nylon membrane (Roche), and UV cross-linked with an energy of 0.12 J. Membranes were prehybridized with DIG Easy Hyb solution (Roche) for 1 h at 42 °C. To detect tRNAs, membranes were incubated overnight (12−16 h) at 42 °C with DIG-labeled oligonucleotide probes synthesized by IDT. The membranes were washed twice with low-stringency buffer [2× SSC with 0.1% (wt/vol) SDS] at 42 °C for 15 min each, rinsed twice with high-stringency buffer [0.1× SSC with 0.1% (wt/vol) SDS] for 5 min each, and rinsed in washing buffer (1× SSC) for 10 min. Following the washes, the membranes were transferred into 1× blocking buffer (Roche) and incubated at room temperature for 3 h, after which antidigoxigenin-AP Fab fragments (Roche) were added into the blocking buffer at a ratio of 1:10,000 and incubated for an additional 30 min at room temperature. The membranes were washed 4 times with DIG washing buffer (1× maleic acid buffer, 0.3% Tween-20) for 15 min each, incubated in DIG detection buffer (0.1 M Tris−HCl, 0.1 M NaCl; pH 9.5) for 5 min, coated with CSPD ready-to-use reagent (Roche), and incubated in the dark for 30 min at 37 °C before imaging with a ChemiDoc MP Imaging System (Bio-Rad). Digoxigenin-labeled Northern blot probes for tRNA detection were synthesized by IDT, and the sequences were as follows:

tRNA-Glu Northern blot probe: 5′-DIG-CTAACCACTAGAC-CACCA

tRNA-Gln Northern blot probe: 5′-DIG-TGGAGGTTCCACC-GAGATTT

Treatment of tRNA with AlkB. In total, 200 ng of tRNA were incubated in a 50-μL reaction mixture containing 50 mM Na-HEPES (pH 8.0; Alfa Aesar), 75 μM ferrous ammonium sulfate (pH 5.0), 1 mM α-ketoglutaric acid (Sigma-Aldrich), 2 mM sodium ascorbate, 50 μg/mL BSA (Sigma-Aldrich), 2.5 μL of RNase inhibitor (NEB), and 200 ng of AlkB enzyme at 37 °C for 30 min (the recommended mass ratio of AlkB enzyme to RNA is 1:1). The mixture was added to 500 μL of TRIzol reagent to perform the RNA isolation procedure according to the manufacturer's instructions.

Site-Specific Quantitative Analysis of Total tRNA Modifications. For each MLC-Seq experiment involving total yeast tRNA, we used approximately 10 pmol or 250 nanograms of total yeast tRNA. Of this amount, 10% was evaluated through LC−MS in its intact form without acid degradation, while the remaining 90% was degraded to produce ladder fragments for subsequent LC−MS measurement and

sequencing analysis. To quantitatively map total tRNA modifications in a site-specific manner, LC−MS data was used to create a 2D t_R-mass plot, similar to the homology section described earlier. An in-house Python algorithm (available on GitHub) conducted a homology search for monoisotopic masses above 24 kDa to identify intact tRNAs. Observed monoisotopic masses of intact tRNAs were compared to theoretical masses of tRNAs calculated using the tRNA sequence in the database. If there is a match between observed and theoretical masses, it indicates the identity of the specific tRNA type at the intact level. Additionally, the algorithm identifies differences in intact mass between observed tRNAs. A difference of 14.0157 Da (±10 ppm) indicates a possible partial methylation, while a difference of 329.0525 Da corresponds to an additional A nucleotide or a possible partial truncation event. For each observed monoisotopic mass, MassSum is performed to gather ladder fragments for the tRNA type, regardless of whether it matches the theoretical mass calculated from the tRNA database. Although some ladder fragments may be missing, the mass differences between two adjacent ladder fragments can be used for base calling and to read parts of the tRNA sequence de novo. The sequence segments, along with the location information on each nucleotide, are then used to search (via BLAST) the tRNA database and identify the specific tRNA type at the ladder fragment level. Once the tRNA type is identified, its sequences and modifications are used to find all ladder fragments that may have been missed in the MassSum data separation step, and to construct the sequences and modifications shown in Figure 6h. This also allows for verification of partial sequence modifications/editing at the fragment level in cases where this information cannot be extracted de novo from LC−MS data. While not all tRNA types can be detected at the intact level and be matched during the homology search step, it is still possible to identify the specific tRNA subtypes and their complete sequences using BLAST tRNA databases with parts of the de novo sequence results. These results are also listed in Figure 6h. Nucleotides or modifications in each position of a tRNA type are confirmed either by its 5′ end or by one or more of its 3′ ladder fragments. However, depending on the abundance and detected ladder fragments for each tRNA, discrepancies can occur when confirming the first few nucleotides at the 5′ end of different tRNA types, even if they share identical nucleotides. For example, the C at the first position of tRNA-ProUGG was not confirmed, while the C at the first position of tRNA-TyrGPA was confirmed.

Differentiating Pseudouridine and Uridine. While pseudouridine (Ψ) can be distinguished from U with N-cyclohexyl-N′-(2-morpholinoethyl)-carbodiimide metho-p-toluenesulfonate (CMC), as the CMC adducts of Ψ and U differ in mass.[20,21] This study did not incorporate CMC treatment step, and therefore could not differentiate Ψ and U.

tRNA Position Numbering. To maintain consistency with published tRNA research, we adhere to conventional numbering as recommended by the cited reference.[47] For instance, the first nucleotide of the T-loop is consistently labeled as position 54, regardless of its actual sequence position in the tRNA, unless stated otherwise.

■ ASSOCIATED CONTENT

Data Availability Statement

All MS sequence data sets used in the manuscript are publicly available through the corresponding repository on GitHub (https://github.com/rnamodifications/MLC-Seq).

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.4c07280.

> Illustration of the data analysis procedure; detection of acid-labile nucleotide modifications; PAGE and northern blot images; MLC-Seq results for specific tRNA species; MLC-Seq and LC−MS chromatographic analysis of yeast total tRNA; and pseudocodes and instructions for MLC-Seq algorithms (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Shenglong Zhang − *Department of Biological and Chemical Sciences, New York Institute of Technology, New York, New York 10023, United States; Department of Chemistry and The RNA Institute, University at Albany, State University of New York, Albany, New York 12222, United States;* ⓞ orcid.org/0000-0001-8003-7484; Email: szhang45@albany.edu

Authors

Xiaohong Yuan − *Department of Biological and Chemical Sciences, New York Institute of Technology, New York, New York 10023, United States;* ⓞ orcid.org/0000-0002-9165-5197

Yue Su − *Department of Biological and Chemical Sciences, New York Institute of Technology, New York, New York 10023, United States*

Benjamin Johnson − *Department of Biological and Chemical Sciences, New York Institute of Technology, New York, New York 10023, United States*

Michele Kirchner − *Department of Biological and Chemical Sciences, New York Institute of Technology, New York, New York 10023, United States*

Xudong Zhang − *Molecular Medicine Program, Department of Human Genetics, and Division of Urology, Department of Surgery, University of Utah School of Medicine, Salt Lake City, Utah 84132, United States*

Sihang Xu − *Department of Biological and Chemical Sciences, New York Institute of Technology, New York, New York 10023, United States*

Sophia Jiang − *Department of Biological and Chemical Sciences, New York Institute of Technology, New York, New York 10023, United States*

Jing Wu − *Department of Biological and Chemical Sciences, New York Institute of Technology, New York, New York 10023, United States*

Shundi Shi − *Department of Chemical Engineering, Columbia University, New York, New York 10027, United States*

James J. Russo − *Department of Chemical Engineering, Columbia University, New York, New York 10027, United States*

Qi Chen − *Molecular Medicine Program, Department of Human Genetics, and Division of Urology, Department of Surgery, University of Utah School of Medicine, Salt Lake City, Utah 84132, United States;* ⓞ orcid.org/0000-0001-6353-9589

Complete contact information is available at:
https://pubs.acs.org/10.1021/jacs.4c07280

Author Contributions

[⊥]X.Y. and Y.S. contributed equally to this work.

Notes

The authors declare the following competing financial interest(s): The authors have filed a patent related to the technology discussed in this manuscript.

Code Availability Pseudocodes and source codes implemented in Python of all the algorithms described in the manuscript (including homology search, identifying acid-labile nucleotides, MassSum data separation, GapFill, ladder complementation) are freely available on GitHub (https://

ACKNOWLEDGMENTS

REFERENCES

(1) Alfonzo, J. D.; Brown, J. A.; Byers, P. H.; et al. A call for direct sequencing of full-length RNAs to identify all modifications. *Nat. Genet.* **2021**, *53*, 1113−1116.

(2) Helm, M.; Motorin, Y. Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.* **2017**, *18*, 275−291.

(3) Zheng, G.; Qin, Y.; Clark, W. C.; et al. Efficient and quantitative high-throughput tRNA sequencing. *Nat. Methods* **2015**, *12*, 835−837.

(4) Behrens, A.; Rodschinka, G.; Nedialkova, D. D. High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. *Mol. Cell* **2021**, *81*, 1802−1815.e7.

(5) Shi, J.; Zhou, T.; Chen, Q. Exploring the expanding universe of small RNAs. *Nat. Cell Biol.* **2022**, *24*, 415−423.

(6) Katibah, G. E.; Qin, Y.; Sidote, D. J.; et al. Broad and adaptable RNA structure recognition by the human interferon-induced tetratricopeptide repeat protein IFIT5. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 12025−12030.

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

(8) Wiener, D.; Schwartz, S. How many tRNAs are out there? *Mol. Cell* **2021**, *81*, 1595−1597.

(9) Pratanwanich, P. N.; Yao, F.; Chen, Y.; et al. Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nat. Biotechnol.* **2021**, *39*, 1394−1402.

(10) Begik, O.; Lucas, M. C.; Pryszcz, L. P.; et al. Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat. Biotechnol.* **2021**, *39*, 1278−1291.

(11) Liu, H.; Begik, O.; Lucas, M. C.; et al. Accurate detection of m6A RNA modifications in native RNA sequences. *Nat. Commun.* **2019**, *10*, No. 4079.

(12) Parker, M. T.; Knop, K.; Sherwood, A. V.; et al. Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m6A modification. *eLife* **2020**, *9*, No. e49658.

(13) Garalde, D. R.; Snell, E. A.; Jachimowicz, D.; et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **2018**, *15*, 201−206.

(14) Thomas, N. K.; Poodari, V. C.; Jain, M.; et al. Direct Nanopore Sequencing of Individual Full Length tRNA Strands. *ACS Nano* **2021**, *15*, 16642−16653.

(15) Suzuki, T. The expanding world of tRNA modifications and their disease relevance. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 375−392.

(16) Su, D.; Chan, C. T. Y.; Gu, C.; et al. Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry. *Nat. Protoc.* **2014**, *9*, 828−841.

(17) Lauman, R.; Garcia, B. A. Unraveling the RNA modification code with mass spectrometry. *Mol. Omics* **2020**, *16*, 305−315.

(18) Wetzel, C.; Limbach, P. A. Mass spectrometry of modified RNAs: recent developments. *Analyst* **2016**, *141*, 16−23.

(19) Kimura, S.; Dedon, P. C.; Waldor, M. K. Comparative tRNA sequencing and RNA mass spectrometry for surveying tRNA modifications. *Nat. Chem. Biol.* **2020**, *16*, 964−972.

(20) Zhang, N.; Shi, S.; Jia, T. Z.; et al. A general LC-MS-based RNA sequencing method for direct analysis of multiple-base modifications in RNA mixtures. *Nucleic Acids Res.* **2019**, *47*, No. e125.

(21) Zhang, N.; Shi, S.; Wang, X.; et al. Direct Sequencing of tRNA by 2D-HELS-AA MS Seq Reveals Its Different Isoforms and Dynamic Base Modifications. *ACS Chem. Biol.* **2020**, *15*, 1464−1472.

(22) Zhang, N.; Shi, S.; Yuan, X.et al. A General LC-MS-Based Method for Direct and *De Novo* Sequencing of RNA Mixtures Containing both Canonical and Modified Nucleotides. In *RNA Modifications: Methods and Protocols*; McMahon, M., Ed.; Springer US: New York, NY, 2021; pp 261−277.

(23) Zhang, N.; Shi, S.; Yoo, B.; et al. 2D-HELS MS Seq: A General LC-MS-Based Method for Direct and *de novo* Sequencing of RNA Mixtures with Different Nucleotide Modifications. *J. Visualized Exp.* **2020**, *161*, No. e61281.

(24) Thomas, B.; Akoulitchev, A. V. Mass spectrometry of RNA. *Trends Biochem. Sci.* **2006**, *31*, 173−181.

(25) Yoluç, Y.; Ammann, G.; Barraud, P.; et al. Instrumental analysis of RNA modifications. *Crit. Rev. Biochem. Mol. Biol.* **2021**, *56*, 178−204.

(26) Björkbom, A.; Lelyveld, V. S.; Zhang, S.; et al. Bidirectional Direct Sequencing of Noncanonical RNA by Two-Dimensional Analysis of Mass Chromatograms. *J. Am. Chem. Soc.* **2015**, *137*, 14430−14438.

(27) Bahr, U.; Aygün, H.; Karas, M. Sequencing of Single and Double Stranded RNA Oligonucleotides by Acid Hydrolysis and MALDI Mass Spectrometry. *Anal. Chem.* **2009**, *81*, 3173−3179.

(28) Schimmel, P. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 45−58.

(29) Rodriguez, V.; Chen, Y.; Elkahloun, A.; et al. Chromosome 8 BAC array comparative genomic hybridization and expression analysis identify amplification and overexpression of TRMT12 in breast cancer. *Genes, Chromosomes Cancer* **2007**, *46*, 694−707.

(30) Wei, F.-Y.; Suzuki, T.; Watanabe, S.; et al. Deficit of tRNALys modification by Cdkal1 causes the development of type 2 diabetes in mice. *J. Clin. Invest.* **2011**, *121*, 3598−3608.

(31) Kirchner, S.; Ignatova, Z. Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nat. Rev. Genet.* **2015**, *16*, 98−112.

(32) Carell, T.; Brandmayr, C.; Hienzsch, A.; et al. Structure and Function of Noncanonical Nucleobases. *Angew. Chem., Int. Ed.* **2012**, *51*, 7110−7131.

(33) Lee, M.; Kim, B.; Kim, V. N. Emerging Roles of RNA Modification: m6A and U-Tail. *Cell* **2014**, *158*, 980−987.

(34) Klungland, A.; Dahl, J. A. Dynamic RNA modifications in disease. *Curr. Opin. Genet. Dev.* **2014**, *26*, 47−52.

(35) Pan, T. N6-methyl-adenosine modification in messenger and long non-coding RNA. *Trends Biochem. Sci.* **2013**, *38*, 204−209.

(36) Kim, D.; Lee, J. Y.; Yang, J. S.; et al. The Architecture of SARS-CoV-2 Transcriptome. *Cell* **2020**, *181*, 914−921.e910.

(37) Jühling, F.; Morl, M.; Hartmann, R. K.; et al. tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.* **2009**, *37*, D159−D162.

(38) Smardo, F. L.; Calvet, J. P. Sequence analysis of the glutamate tRNA family: evidence for pseudogenes. *Gene* **1987**, *57*, 213−220.

(39) Shi, J.; Zhang, Y.; Tan, D.; et al. PANDORA-seq expands the repertoire of regulatory small RNAs by overcoming RNA modifications. *Nat. Cell Biol.* **2021**, *23*, 424−436.

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

(41) Ohira, T.; Minowa, K.; Sugiyama, K.; et al. Reversible RNA phosphorylation stabilizes tRNA for cellular thermotolerance. *Nature* **2022**, *605* (7909), 372−379.

(42) Zhang, S.; Blain, J. C.; Zielinska, D.; Gryaznov, S. M.; Szostak, J. W. Fast and accurate nonenzymatic copying of an RNA-like synthetic genetic polymer. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110* (44), 17732−17737.

(43) Juhling, F.; Morl, M.; Hartmann, R. K.; et al. UCSC tRNA database Ref of tRNA sequences and tRNA genes. *Nucleic Acids Res.* **2009**, *37*, D159−162.

(44) Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezhuk, Y.; McGinnis, S.; Madden, T. L. NCBI BLAST: a better web interface. *Nucleic Acids Res.* **2008**, *36*, W5−W9.

(45) Brenton, A. G.; Godfrey, A. R. Accurate mass measurement: terminology and treatment of data. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1821−1835.

(46) Drino, A.; Oberbauer, V.; Troger, C.; et al. Production and purification of endogenously modified tRNA-derived small RNAs. *RNA Biol.* **2020**, *17*, 1104−1115.

(47) de Crécy-Lagard, V.; Boccaletto, P.; Mangleburg, C. G.; et al. Matching tRNA modifications in humans to their known and predicted enzymes. *Nucleic Acids Res.* **2019**, *47*, 2143−2159.