PhD Program in Biomolecular Sciences
Department of Cellular, Computational and Integrative Biology – CIBIO

37th Cycle

"Integrating Ribosome Profiling and tRNA Sequencing with Computational Solutions to Decode the Complexity of the Translatome"

Tutor
Prof. Dr. Alessandro Quattrone,
CIBIO, University of Trento (Italy)

Advisor
Dr. Massimiliano Clamer,
Immagina Biotechnology (Italy)

Ph.D. Thesis of
Hasan YILMAZ

Department of Cellular, Computation and Integrative Biology,
University of Trento (Italy)

Immagina Biotechnology s.r.l (Italy)

Academic Year 2023-2024

Declaration of original authorship

I, Hasan YILMAZ, confirm that this is my own work and the use of all material from other sources have been properly and fully acknowledged.

Abbreviations

Abbreviation	Definition
aa-tRNA-seq	Aminoacyl-tRNA sequencing
AGC	Automatic Gain Control
ALKBH / AlkB	AlkB homolog demethylase enzyme
ANOVA	Analysis of Variance
APPRIS	Annotation of Principal and Regulatory Isoforms
ARM-seq	AlkB-facilitated RNA Methylation Sequencing
Ars / Asn	Sodium arsenite (oxidative stress inducer)
A-site	Aminoacyl site (ribosomal site accepting incoming charged tRNA)
BAM	Binary Alignment Map
BH	Benjamini–Hochberg (multiple testing correction procedure)
CCA tail	3′-cytidine-cytidine-adenosine trinucleotide tail of tRNA, required for aminoacylation
cDNA	Complementary DNA
CDS	Coding Sequence
CCDS	Consensus Coding Sequence (database identifier)
CHX	Cycloheximide (translation elongation inhibitor)
CI	Confidence Interval
CSV	Comma-Separated Values
DE	Differential Expression
DESeq2	Differential Expression analysis using a negative binomial model (Bioconductor R package)
DM-tRNA-seq	Demethylase-tRNA sequencing
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulfoxide
DRS	Direct RNA Sequencing (Oxford Nanopore Technologies)
E-site	Exit site (ribosomal site releasing deacylated tRNA)
eIF2α	Eukaryotic Initiation Factor 2 alpha subunit
eIF4E	Eukaryotic Initiation Factor 4E
FACS	Fluorescence-Activated Cell Sorting
FBS	Fetal Bovine Serum
FDR	False Discovery Rate
GCN2	General Control Non-depressible 2 (eIF2α kinase)
GSH	Glutathione
GTF	Gene Transfer Format
HAR	Harringtonine (translation initiation inhibitor)
HCD	Higher-energy Collisional Dissociation
HEK293T	Human Embryonic Kidney 293T cell line

(Continued from previous page)

Abbreviation	Definition
HRP	Horseradish Peroxidase
HTML	HyperText Markup Language
HTSeq	High-Throughput Sequencing read counting tool
iMet-CAT	Initiator Methionine tRNA with CAT anticodon
IRES	Internal Ribosome Entry Site
ISR	Integrated Stress Response
LC-MS	Liquid Chromatography-Mass Spectrometry
LC-MS/MS	Liquid Chromatography-Tandem Mass Spectrometry
LOWESS	Locally Weighted Scatterplot Smoothing
m^1A	N1-methyladenosine (RNA modification)
m^3C	3-methylcytidine (RNA modification)
m^5C	5-methylcytidine (RNA modification)
m^5U	5-methyluridine (RNA modification)
m^7G	7-methylguanosine (RNA modification)
MapQ	Mapping Quality score
MAT2A	Methionine Adenosyltransferase 2A
MCF-7	Michigan Cancer Foundation-7 (breast cancer cell line)
mim-tRNAseq	Modification-induced misincorporation tRNA sequencing
mRNA	Messenger RNA
MTAP	Methylthioadenosine Phosphorylase
mTORC1	Mechanistic Target of Rapamycin Complex 1
nano-tRNAseq	Nanopore-based direct RNA sequencing of tRNAs
NE	Nux Enhancer
ORF	Open Reading Frame
PBS	Phosphate-Buffered Saline
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
P-site	Peptidyl site (active decoding centre of the ribosome)
PRM	Parallel Reaction Monitoring
ribo-tRNA	Ribosome-associated tRNA
RiboLace	Active ribosome profiling kit based on puromycin derivative pull-down (Immagina Biotechnology)
Ribo-seq	Ribosome profiling sequencing
RmB	RiboLace magnetic beads
RPF	Ribosome-Protected Fragment
RPKM	Reads Per Kilobase per Million mapped reads
rRNA	Ribosomal RNA
RsP	RiboLace smart probe
SAM	S-adenosylmethionine (universal methyl donor)

(Continued from previous page)

Abbreviation	**Definition**
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
SVG	Scalable Vector Graphics
TE	Translation Efficiency
TGIRT	Thermostable Group II Intron Reverse Transcriptase
TIS	Translation Initiation Site (also: TIS score)
TRAC-seq	tRNA reduction and cleavage sequencing
tRF	tRNA-derived Fragment
tRIBO-seq	Transfer RNA Ribosome Profiling by nanopore sequencing (novel method, Chapter 2)
TRMT61A	tRNA Methyltransferase 61A (deposits m^1A at Sprinzl position 58)
tRNA	Transfer RNA
TSV	Tab-Separated Values
UMI	Unique Molecular Identifier
uORF	Upstream Open Reading Frame
UTR	Untranslated Region
5′ UTR	Five-prime Untranslated Region
3′ UTR	Three-prime Untranslated Region
YAMAT-seq	Yet Another Method for tRNA sequencing
ψ (Ψ)	Pseudouridine (RNA modification)

Abstract

Translation regulation is fundamental to cellular protein synthesis and plays critical roles in development, stress responses, and disease pathogenesis. While ribosome profiling has revolutionized our understanding of translation, current computational tools remain fragmented, requiring multiple disconnected pipelines for comprehensive analysis. Moreover, studying transfer RNA dynamics, particularly from translationally active populations, remains technically challenging due to labor-intensive isolation procedures and limitations in simultaneously capturing tRNA abundance, modifications, and fragmentation patterns. This thesis addresses these challenges through the development of two complementary tools: Martian, an integrated ribosome profiling analysis pipeline, and tRIBO-seq, a novel method for profiling ribosome-associated tRNAs.

Martian is a comprehensive computational pipeline that consolidates five essential modules into a unified workflow: database creation with reference filtering and UTR correction, ribosome profiling with P-site mapping and coverage analysis, condition comparison with differential translation analysis, translation efficiency calculation integrating RNA-seq and Ribo-seq data, and nano-tRNAseq integration for tRNA quantification. Benchmarking against established tools (riboWaltz and Plastid) across 11 samples demonstrated that Martian achieves the fastest runtime (1,896 seconds versus 3,029 and 2,128 seconds), competitive P-site detection accuracy (TIS score 0.91 versus 1.00 and 0.14), and intermediate memory requirements. The pipeline successfully integrates multiple omics data types within a reproducible framework, providing researchers with an efficient platform for comprehensive translational analysis.

tRIBO-seq (transfer RNA Ribosome Profiling by nanopore sequencing) introduces a streamlined method for profiling translationally active tRNAs by combining RiboLace-based ribosome isolation with nanopore direct RNA sequencing. This approach eliminates labor-intensive sucrose gradient fractionation, requires minimal input material (3-5 µg RNA from approximately 5 million cells), and can be completed within 5 hours. Critically, tRIBO-seq enables simultaneous quantification of tRNA abundance, post-transcriptional modifications, and fragmentation patterns from a single experiment.

Application of tRIBO-seq to diverse stress conditions revealed distinct tRNAome reprogramming patterns. Under steady-state conditions, ribosome-associated and total tRNA pools showed high similarity with specific enrichment of initiator methionine tRNA in ribosomes. Amino acid deprivation (leucine or arginine) induced changes primarily in tRNA abundances with enrichment of cognate tRNA isoacceptors. Methionine starvation, in contrast, caused widespread tRNA hypomethylation linked to S-adenosylmethionine depletion, with minimal abundance changes. Oxidative stress induced by arsenite exposure triggered selective tRNA fragmentation predominantly in ribosome-embedded tR-

NAs, with cleavage occurring at anticodon regions, consistent with ribosome-localized angiogenin activation.

These findings demonstrate that ribosome-associated tRNAs respond distinctly from total tRNA pools under stress conditions, emphasizing the importance of studying translationally active tRNA populations. Together, Martian and tRIBO-seq provide the research community with accessible, robust platforms for investigating translation regulation at both ribosome and tRNA levels, enabling comprehensive analysis of protein synthesis dynamics in health and disease.

Introduction

The central dogma of molecular biology describes the fundamental flow of genetic information from DNA through RNA to proteins, with each step representing a critical control point for gene expression regulation [1]. While transcriptional control has been extensively studied, the translation of mRNA into proteins represents an important layer of biological knowledge that directly determines cellular proteomes and physiological responses [2]. In eukaryotic systems, however, correlation between mRNA abundance and protein levels is frequently weak [3, 4], underscoring the significance of post-transcriptional regulatory mechanisms that influence protein synthesis independently of mRNA abundance [5]. The ability of cells to modulate protein synthesis through diverse mechanisms, including ribosome recruitment, open reading frame (ORF) selection, elongation dynamics, and tRNA availability, enables rapid adaptations to environmental changes while maintaining precise control over cellular function [6].

The clinical significance of translation regulation is underscored by its dysregulation in human disease. Translation deregulation affects the majority of human tumors through hyperactivation of mTORC1 signaling [7], while disrupted translation control underlies neurological disorders including amyotrophic lateral sclerosis, Parkinson's disease, and Alzheimer's disease [8]. Recent developments have positioned ribosome profiling as an emerging precision medicine tool, with growing recognition of translation machinery components as therapeutic targets. Translation-targeted therapies including eIF4E inhibitors like ribavirin in acute myeloid leukemia have demonstrated clinical efficacy, with mechanism-based patient selection improving response rates [9]. Ribosome biogenesis inhibitors and translation initiation modulators represent promising therapeutic approaches, with several compounds entering clinical evaluation for cancer treatment [10]. The VICTORIA study has demonstrated ribosome biogenesis factors as predictive biomarkers for mTOR inhibitor response in endometrial cancer, highlighting the clinical utility of integrated ribosome profiling and biomarker analysis [11].

Understanding these complex regulatory networks requires sophisticated experimental approaches capable of analyzing translation at multiple levels. The translatome, the complete set of mRNAs undergoing active translation, can be interrogated through various complementary techniques including polysome profiling for assessing global translation states, ribosome footprinting for codon-resolution analysis of ribosome positioning, and emerging single-cell approaches that reveal translational heterogeneity. Among these approaches, ribosome profiling (Ribo-seq) has emerged as the gold standard for genome-wide translation analysis, offering unprecedented resolution and quantitative power for studying translational control mechanisms. The Ribo-seq technology has revolutionized our understanding of translation by providing genome-wide, codon-resolution snapshots

of ribosome positioning and translation dynamics. This technique exploits the protective effect of ribosomes on mRNA fragments during nuclease digestion, enabling sequencing of ribosome-protected fragments (RPFs) that reveal the precise locations of actively translating ribosomes across the transcriptome [12].

The success of ribosome profiling experiments critically depends on proper sample preparation to capture translational snapshots. The standard workflow begins with rapid cell lysis in the presence of translation elongation inhibitors such as cycloheximide, which blocks ribosomes in their positions along mRNAs. However, important considerations exist regarding cycloheximide treatment, as studies have revealed that it can introduce biases in ribosome positioning, particularly affecting initiation site occupancy [13, 14]. Alternative approaches using flash-freezing or rapid lysis without inhibitors have been developed to minimize such artifacts, though each method presents distinct trade-offs between preservation of ribosome positions and technical feasibility. Following lysis, ribosome-mRNA complexes are isolated through various approaches. Traditional methods employ nuclease digestion to generate ribosome-protected fragments, followed by sucrose gradient ultracentrifugation or size-selection to isolate monosomes. The optimization of nuclease treatment, including enzyme choice, concentration, and digestion time, significantly impacts the quality and interpretability of results. Recent advances have introduced ultra-rapid protocols enabling analysis from FACS-sorted cells with as little as 0.1 pmol of 30 nt RNA fragments [15], transforming ribosome profiling from a specialized technique requiring substantial material inputs to a routine genomic assay applicable to rare cell populations and clinical specimens.

A critical methodological breakthrough has been the development of active ribosome profiling approaches, which enable selective enrichment of actively translating ribosomes, thereby improving the signal-to-noise ratio and providing more accurate measurements of genuine translation events rather than ribosomal occupancy artifacts [16]. This active profiling approach addresses a fundamental limitation of traditional ribosome profiling, where inactive or stalled ribosomes could confound interpretation of translation dynamics, and enables more precise quantification of ribosome flux and elongation rates across different cellular conditions. Library construction represents a critical step that determines data quality and interpretability. Standard ribosome profiling library preparation involves several key steps: (i) RNA extraction from ribosome-protected fragments, (ii) size selection to enrich for footprints of appropriate length (typically 28-32 nucleotides), (iii) end repair and adapter ligation, (iv) reverse transcription, and (v) PCR amplification. Each step requires careful optimization to minimize biases and ensure adequate representation of the translatome. Recent methodological advances have significantly enhanced library preparation approaches, such as methods based on the circularization of RPFs for subsequent cDNA nanopore sequencing or PCR amplification [17]. Additional library preparation in-

novations include optimized adapter ligation strategies that reduce bias, improved PCR amplification protocols that maintain quantitative accuracy, and multiplexing approaches that enable cost-effective processing of multiple samples. Single-cell ribosome profiling methods have emerged that reveal cell-to-cell translational heterogeneity [18, 19], further expanding the applications of this powerful technology.

Despite the power of ribosome profiling, a major technical challenge is contamination from ribosomal RNA (rRNA) and transfer RNA (tRNA) fragments. During the nuclease digestion step, free rRNA and tRNA molecules can be fragmented to sizes similar to ribosome-protected fragments, leading to their co-purification and sequencing. This contamination can be severe, with traditional protocols often yielding only 10-30% usable reads mapping to mRNA-derived footprints, while the remaining 70-90% represent rRNA and tRNA contamination [17]. The active ribosome profiling approach exemplified by RiboLace helps address contamination issues by selectively enriching for ribosome-mRNA complexes. Building upon these insights, advanced methodological developments have emerged to directly tackle the contamination problem. The Ribo-FilterOut method represents a breakthrough in addressing rRNA contamination, employing targeted depletion strategies that increase usable reads to 83% versus the typical 10-30% obtained with standard protocols [20]. Furthermore, the Ribo-Calibration method has introduced external spike-ins of stoichiometrically defined mRNA-ribosome complexes, enabling absolute ribosome number estimation and measurement of translation initiation rates with unprecedented accuracy [20]. These calibration approaches not only improve quantification but also serve as internal controls for assessing contamination levels and technical quality throughout the experiment. The recognition of these contamination issues has driven the development of specialized computational approaches for rRNA/tRNA filtering during data analysis, as well as improved experimental protocols that minimize contamination at the source through optimized size selection, enhanced purification methods, and selective enrichment strategies.

Parallel to advances in ribosome profiling, methods for characterizing tRNA populations have evolved significantly. Transfer RNA molecules serve as molecular adaptors that translate genetic information encoded within mRNAs into corresponding amino acid sequences in proteins [21, 22]. The anticodon of each tRNA molecule pairs with the cognate mRNA codon, ensuring accurate amino acid delivery to the ribosome for incorporation into the nascent polypeptide chain. Due to wobble base pairing at the third codon position, multiple tRNA species can decode the same amino acid. tRNAs carrying the same amino acid but differing in anticodon sequence are termed *isoacceptors*, while tRNAs sharing identical anticodons but differing in body sequence are designated *isodecoders*. This tRNA diversity enables fine-tuned translational control, as different isoacceptors and

isodecoders can exhibit distinct expression patterns, modification profiles, and functional properties [23, 24].

To fulfill their essential function, tRNAs undergo extensive chemical modification—a diverse array of RNA modifications critical for stabilizing tertiary structure, maintaining codon-anticodon interactions, and preventing frameshifting during translation [25–27]. In eukaryotes, tRNA molecules bear an average of 13 modifications per molecule [28]. Multiple positions within individual tRNA molecules can undergo modification, and multiple modification types (acetylation, isomerization, methylation, thiolation) can coexist within single tRNA species [23, 29]. This chemical diversity creates extensive combinatorial potential for translational regulation, enabling cells to fine-tune translation in context- and condition-dependent manners [30–33]. Previous studies have demonstrated that specific tRNA modifications, including methylation and thiolation, can be altered under particular stress exposures [34–37]. Transfer RNAs are therefore increasingly recognized as modulators of protein synthesis that fine-tune translational output through three primary mechanisms: (i) changing relative abundances of isoacceptors, (ii) altering charging status with cognate amino acids, and (iii) modifying codon recognition through chemical modifications [28, 36, 38, 39]. The clinical significance of tRNA modifications is underscored by "modopathies", diseases affecting more than 50 tRNA modification enzymes, with 72% of tRNA-modifying proteins linked to pathological conditions ranging from microcephaly and intellectual disability to mitochondrial cardiomyopathy and cancer progression [40]. The identification of over 50 tRNA modification enzymes associated with human diseases, including NSUN2 upregulation correlating with poor survival in multiple tumor types and TRMT12/TYW2 epigenetic silencing in colon cancer, highlights the clinical potential of integrated translation and tRNA modification analysis [41].

To identify dynamically regulated tRNA modifications, highly sensitive assays such as liquid chromatography-mass spectrometry (LC-MS) can be employed to identify and quantify tRNA modification changes [37, 42]. However, these methods typically cannot identify which tRNAs and at which specific locations modifications are altered. To overcome this limitation, targeted modification-specific sequencing techniques have been developed, including hydro-tRNAseq, which uses alkaline hydrolysis to fragment tRNAs at modified sites, and ARM-seq (AlkB-facilitated RNA methylation sequencing), which employs demethylase treatment to identify methylation sites through comparative analysis [43–45]. More recent developments include DM-tRNA-seq (demethylase-tRNA-seq), which combines ALKBH treatment with high-throughput sequencing to map m^1A and m^3C modifications [46], and TRAC-seq (tRNA reduction and cleavage sequencing), which uses sodium borohydride reduction followed by aniline cleavage to detect m^7G modifications [47]. Additional methods include mim-tRNAseq for high-resolution quantitative profiling using TGIRT reverse transcriptase to overcome modification-induced stops [48],

and QuantM-tRNA seq for monitoring tissue-specific tRNA regulation [24]. These methods have provided valuable information about tRNA modification patterns but remain limited by their indirect detection mechanisms, requirement for chemical treatments to ensure proper cDNA synthesis, and inability to simultaneously assess multiple modification types [35, 48–53].

The emergence of nanopore sequencing technologies has transformed tRNA analysis by enabling direct, single-molecule detection of modifications through characteristic signal distortions. More recently, nanopore direct RNA sequencing (DRS) technologies have enabled capture of native tRNA populations (Nano-tRNAseq), allowing simultaneous assessment of tRNA abundance and modification patterns in a streamlined, high-throughput manner [54]. The nano-tRNAseq approach combines double RNA adapter ligation strategies utilizing natural $3'$ CCA overhangs with custom MinKNOW configurations that increase tRNA read recovery by 12-fold over standard nanopore settings [54]. Nano-tRNAseq relies on bidirectional extension of tRNA molecules [54–56], thereby retaining information from full-length tRNA molecules. This single-molecule approach simultaneously quantifies tRNA abundance and modification status with remarkable reproducibility ($\rho = 0.984$ between biological replicates), revealing modification interdependencies such as Ψ55 positively influencing m^5U54 and m^1A58 incorporation in *S. cerevisiae* $tRNA^{Phe}$. Nanopore-based tRNA sequencing methods circumvent the requirement for separate modification and abundance quantification assays, enabling more comprehensive analysis of tRNA dynamics under diverse conditions within a single experimental workflow [54, 57, 58]. The nano-tRNAseq platform has revealed unexpected biological insights, including limited stress-dependent modification changes contrary to previous assumptions, widespread $3'$ deadenylation under oxidative stress affecting all tRNA isoacceptors, and the identification of modification hierarchies that govern sequential installation pathways [54]. The recent development of aminoacyl-tRNA sequencing (aa-tRNA-seq) methodology represents a further improvement enabling direct nanopore sequencing of intact aminoacylated tRNAs, using chemical ligation approaches and machine learning models to distinguish amino acids based on unique nanopore signal distortions, achieving F1 scores of 0.966 for charged versus uncharged tRNA classification [59].

Computational innovations have paralleled experimental advances in both ribosome profiling and tRNA analysis. A critical step in ribosome profiling analysis is determining the precise Peptidyl site (P-site) position (where the peptidyl-tRNA resides) from sequencing reads. Advanced P-site mapping algorithms like riboWaltz achieve approximately 85% correct frame assignment compared to 65% with earlier methods by employing coherence-based optimization across read lengths [60]. Translation efficiency calculation frameworks have evolved from simple RPF/mRNA ratios to sophisticated statistical approaches including deltaTE and Mean of Typical Decoding Rates (MTDR) that address

statistical challenges in identifying translationally regulated genes while handling complex experimental designs and global translational changes [61, 62]. Quality control frameworks have matured substantially, with comprehensive pipelines like riboseq-flow providing automated diagnostic metrics, periodicity analysis, and reproducibility evaluation through Nextflow DSL2 implementation with nf-core standards [63]. The development of machine learning approaches has further enhanced analytical capabilities. The Translatomer framework achieves 72-80% accuracy for cell-type-specific translation prediction through multimodal integration of mRNA expression and gene sequence features, successfully identifying 3,041 disease-associated variants affecting translation in Alzheimer's disease, schizophrenia, and congenital heart disease [64]. Deep learning frameworks like Riboformer provide context-aware artifact correction and identify ribosome stalling motifs in aging and viral infection, demonstrating superior performance compared to traditional statistical methods [65]. Machine learning applications have also enhanced analytical capabilities for nanopore-based tRNA analysis, with computational approaches including modification detection pipelines that simultaneously identify 43+ distinct RNA modifications using integrated current-based and dwell-time signatures [66]. Advanced tRNA sequencing and quantification methods incorporating hierarchical mapping strategies and clustering approaches have been developed to manage sequence similarity among hundreds of highly similar tRNA genes [67].

Comprehensive data resources have matured significantly, with RPFdb v3.0 now containing 5,018 ribosome profiling datasets from 496 studies across 34 species, accompanied by 2,343 matched RNA-seq datasets, while the RiboSeq.Org data portal has expanded to include 14,840 preprocessed samples from 969 studies covering 96 species, providing standardized processing pipelines and browser-based analysis capabilities [68, 69]. The analytical landscape for ribosome profiling includes major integrated platforms like RiboSeq.Org [69], comprehensive analysis frameworks like riboviz 2 offering flexible workflow management for diverse experimental designs [70], and specialized tools including RiboVIEW for quality control and statistical analysis with automated diagnostic capabilities [71]. However, comprehensive benchmarking studies reveal substantial performance variations among existing pipelines [72], with ribosome profiling tools showing only 40% peak detection reproducibility between replicates [73] and only 2% agreement across major open reading frame detection tools for small ORFs [74]. These discrepancies highlight the need for standardized approaches and careful tool selection based on specific research questions.

Despite remarkable experimental and computational advances, the integration of ribosome profiling data with tRNA sequencing and other omics modalities remains a significant challenge. The analytical landscape suffers from fragmentation across numerous specialized tools that create significant analytical bottlenecks and reproducibility challenges.

Previous investigations characterizing tRNA populations and their dynamics across conditions have typically focused on total cellular tRNA populations [24, 43, 44, 46, 48, 54, 75]. However, total tRNA populations under nutrient-limiting or stressed conditions may not accurately reflect cellular translational status, as they differ from ribosome-associated tRNA populations purified from polysomal fractions [35, 76, 77]. While ribosome-associated tRNAs are expected to more accurately represent cellular translational status, their isolation remains laborious and time-consuming, requiring large input amounts. Current approaches are hampered by the inherent incompatibility of multi-omics data types, where ribosome profiling, RNA sequencing, and tRNA sequencing each produce data with distinct characteristics, measurement scales, and formats that require extensive preprocessing including data filtering, systematic normalization, batch effect removal, and quality checks, each with substantial influence on downstream analysis but lacking universal standards for implementation [78, 79].

Current fragmented approaches require researchers to navigate multiple specialized tools including riboWaltz for P-site optimization, Xtail or RiboDiff for differential translation analysis, various tRNA sequencing analysis pipelines including those for nanopore data processing, and separate statistical frameworks for multi-omics integration including pathway enrichment tools like PEANUT and ActivePathways [60, 80–83]. The manual orchestration of these diverse tools creates numerous failure points, inconsistent normalization approaches, and suboptimal resource utilization that undermines both analytical depth and reproducibility. Technical coordination challenges include synchronizing ribosome occupancy measurements with tRNA charging states, where the temporal dynamics of aminoacyl-tRNA synthetase activity operate on different time scales than ribosome translocation, creating analytical complexity for studies attempting to correlate translation efficiency with tRNA availability. Sample preparation incompatibilities between ribosome profiling (requiring cycloheximide treatment and ultracentrifugation) and tRNA sequencing (requiring specialized adapter ligation for modified RNAs) necessitate parallel sample processing approaches that may introduce systematic biases [13, 14, 38]. The complexity of multi-omics analyses requiring orchestration of numerous tools with different releases and system library dependencies creates reproducibility issues that undermine scientific rigor, while version dependency conflicts, inadequate documentation of analytical decisions, and absent provenance tracking create gaps in understanding how results were generated [84, 85].

This thesis addresses these critical challenges through two complementary approaches. First, the Martian computational pipeline provides a modular architecture that integrates database creation, ribosome profiling analysis, condition comparison, translation efficiency calculation, and tRNA sequencing analysis within a unified framework. This comprehensive approach enables standardized data processing, automated quality control, scalable

statistical analysis, and accessible interfaces while maintaining the flexibility and analytical rigor required for cutting-edge biological research. Second, through collaborative development of tRIBO-seq, a novel nanopore-based high-throughput method, this work enables simultaneous capture of tRNA abundances and modification information from ribosome-associated tRNA populations (ribo-tRNAs). An efficient and streamlined method capable of capturing actively translating ribosome-associated tRNAs is essential for understanding how the tRNAome is rapidly tuned to adapt proteomic output to environmental cues. Using tRIBO-seq, differential tRNAome reprogramming was revealed depending on cellular stress type: (i) upon arginine or leucine deprivation, altered tRNA abundances of specific tRNA isoacceptors were observed; (ii) upon methionine deprivation, global tRNA hypomethylation patterns were identified; and (iii) upon arsenite exposure, selective tRNA fragmentation was observed that occurred mainly in ribosome-embedded tRNAs, consistent with recent work suggesting endonuclease activation at the ribosome [86]. This work demonstrates that tRNA-derived fragments (tRFs) can be captured using nanopore sequencing, providing a novel framework to study both tRFs and full-length tRNAs within a single platform and experimental sample. Together, these tools advance the field of translational biology by providing accessible, robust platforms for investigating translation regulation at both ribosome and tRNA levels, enabling comprehensive analysis of protein synthesis dynamics in health and disease and facilitating clinical applications in the rapidly evolving field of translational medicine.

Aims

Translation regulation is essential for cellular protein synthesis, yet current methodologies for studying translational dynamics face significant limitations. Ribosome profiling analysis tools are fragmented and require multiple disconnected pipelines, while profiling translationally active transfer RNAs remains technically challenging and labor-intensive. This thesis aims to address these limitations through the development of two complementary approaches.

Aim 1: Develop an integrated pipeline for comprehensive ribosome profiling analysis

The first aim is to create Martian, a unified computational pipeline that consolidates the fragmented landscape of ribosome profiling tools. This pipeline will integrate database creation, P-site mapping, differential translation analysis, translation efficiency calculation, and tRNA sequencing into a single reproducible workflow. The pipeline will be benchmarked against established tools to evaluate accuracy, runtime performance, and usability.

Aim 2: Establish a streamlined method for profiling translationally active tRNAs

The second aim is to develop tRIBO-seq, a novel approach combining ribosome isolation with nanopore direct RNA sequencing to profile ribosome-associated tRNA populations. This method will enable simultaneous quantification of tRNA abundance, post-transcriptional modifications, and fragmentation patterns from minimal input material, overcoming the technical barriers of traditional approaches.

Aim 3: Characterize tRNA dynamics under diverse cellular stress conditions

The third aim is to apply tRIBO-seq to investigate how ribosome-associated tRNAs respond to different stress conditions, including amino acid deprivation and oxidative stress. This will reveal whether translationally active tRNAs exhibit distinct responses compared to total tRNA pools, providing insights into stress-specific tRNA reprogramming mechanisms.

Together, these aims will provide the research community with robust, accessible tools for comprehensive translation analysis at both ribosome and tRNA levels.

Contents

List of Figures

List of Tables

Chapter 1

Martian: A Tool-Chain for Translatome Studies

1.1 Introduction

The analytical landscape for ribosome profiling data, while rich in specialized tools, suffers from significant fragmentation that creates bottlenecks for comprehensive translational analysis. Current workflows require researchers to manually orchestrate multiple disconnected pipelines, riboWaltz for P-site mapping, separate tools for differential translation analysis, distinct frameworks for translation efficiency calculation, and independent processing streams for tRNA sequencing data. This fragmentation not only undermines analytical efficiency but also introduces reproducibility challenges, as each tool operates with different data formats, normalization strategies, and quality control metrics. The manual integration of these disparate components creates numerous failure points and prevents seamless multi-omics analysis that is increasingly essential for understanding the complex interplay between ribosome dynamics and tRNA availability.

This chapter presents Martian, an integrated computational pipeline designed to address these analytical challenges through a unified framework. Martian consolidates five essential modules, database creation with reference filtering and UTR correction, ribosome profiling analysis with P-site mapping and coverage calculation, condition comparison with differential translation statistics, translation efficiency computation integrating RNA-seq and Ribo-seq data, and nano-tRNA sequencing analysis, into a single, reproducible workflow. This integrated architecture enables standardized data processing while maintaining the flexibility required for diverse experimental designs. To evaluate Martian's performance, I conducted comprehensive benchmarking against two established tools, riboWaltz and Plastid, assessing runtime efficiency, memory requirements, P-site detection accuracy, and transcript detection patterns across multiple datasets.

The remainder of this chapter describes Martian's modular architecture and implementation details, presents benchmarking results demonstrating competitive accuracy with superior computational efficiency, and discusses the design trade-offs inherent in database curation strategies. I conclude with perspectives on future developments aimed at enhancing analytical capabilities and accessibility for the broader research community. Comprehensive documentation, example datasets, and detailed output specifications for Martian are available at `https://github.com/ImmaginaBiotechnology/Documents`.

1.2 Methods

Martian is organized into five interconnected modules: **Database Creation**, **Ribosome Profiling**, **Condition Comparison**, **Translation Efficiency**, and **Nano-tRNAseq** (Figure 1.1). The workflow begins with *Database Creation*, where transcript sequences and annotations are processed to filter transcripts, correct untranslated region (UTR) boundaries, and compile a curated transcript database.

In the **Ribosome Profiling** module, aligned ribosome-protected fragments (RPFs) [12] undergo read filtering, Peptidyl(P)-site offset calculation, codon identification, and coverage computation, producing both graphical and tabular outputs.

The **Condition Comparison** module uses curated comparison lists to perform quality control, quantify differential transcript abundances, and assess differential codon usage between experimental conditions.

The **Translation Efficiency** module integrates RNA-seq and Ribo-seq data to compute differential translation efficiency [61], incorporating directional analyses for targeted interpretation.

Finally, the **Nano-tRNAseq** module processes Nano-tRNAseq [54] alignments for read filtering, abundance estimation, modification prediction, and dynamic translation analysis. Each module generates interactive graphical reports and structured output tables, enabling flexible integration into downstream analyses.

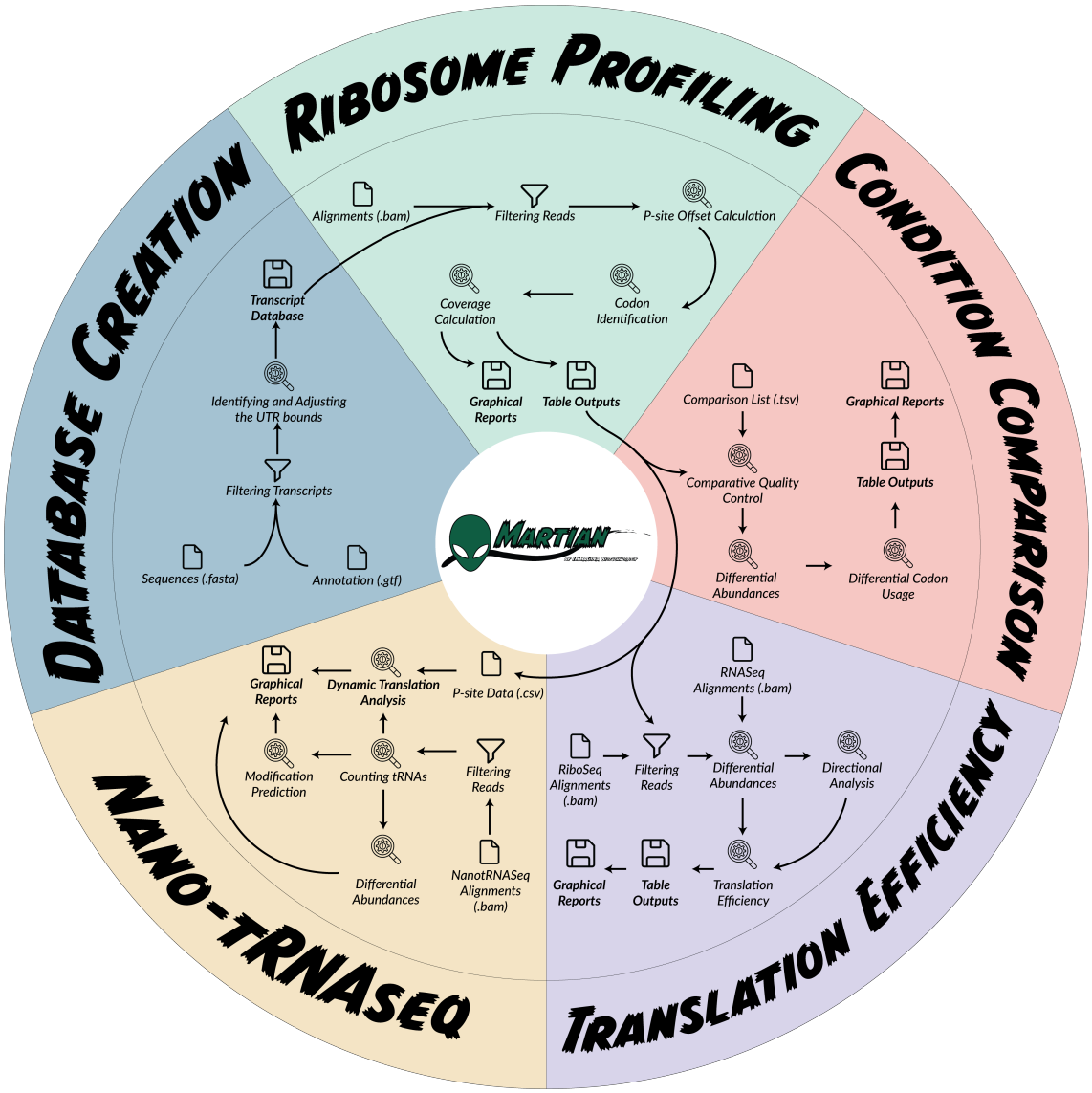

Figure 1.1: Martian's modular architecture overview. The pipeline consists of five interconnected modules: Database Creation (filtering and UTR correction), Ribosome Profiling (P-site mapping and coverage analysis), Condition Comparison (statistical testing and visualization), Translation Efficiency (integrated RNA-seq and Ribo-seq analysis), and Nano-tRNAseq (tRNA quantification and modification detection). Arrows indicate data flow within and between modules.

1.2.1 Module 1: Database Creation - Transcript Annotation Processing and Database Generation

Transcriptome annotations were parsed from a Gene Transfer Format (GTF) file to construct a reference transcript database for downstream analyses. When APPRIS [87] database file is provided by users, transcripts that do not have consensus coding sequences ID (CCDS ID) are excluded. Transcripts lacking annotated start or stop codons are excluded to ensure accurate identification of coding sequences (CDS).

For each retained transcript, the annotations are inspected for potential overlaps between the untranslated regions (UTRs) and the CDS coordinates. When such overlaps are detected, the UTR boundaries are corrected based on the CDS range to enforce mutually exclusive regions. UTRs are then annotated as 5′ or 3′ depending on their position relative to the CDS.

Following this correction, the total transcript length, CDS length, 5′ UTR length, and 3′ UTR length are computed. These features are exported to a structured CSV file and served as the reference database for subsequent modules in the analysis.

This procedure functionally mirrors the logic implemented in the Bioconductor package GenomicFeatures [88], which builds transcript databases (TxDb objects) from GTF/GFF annotations and supports region-specific feature extraction through genomic coordinate operations. The custom implementation was tailored to enable finer control over UTR-CDS boundary correction and ensure compatibility with non-standard annotations present in certain transcriptomes.

1.2.2 Module 2: Ribosome Profiling - Read Processing, P-Site Mapping, and Codon Identification

This module requires the curated transcript database generated by Module 1, specifically utilizing the corrected UTR boundaries and CDS annotations to ensure accurate P-site mapping and frame determination.

1.2.2.1 Data Analysis

This module processes ribosome sequencing (RiboSeq) data to infer P-site positions, coding frame, and region classification for aligned ribosome-protected fragments (RPFs). The inputs include a transcriptome FASTA file, a curated transcript annotation database (from Module 1), a sample metadata file with BAM file paths, and a species-specific codon table in JSON format.

The transcriptome FASTA was indexed into an in-memory dictionary of transcript sequences for use in further steps. Then, for each sample, alignments were parsed using the pysam (version 0.21.0) library, and the RPFs were grouped by read length.

The P-site represents the active decoding center, and its accurate identification is crucial for determining which codon is being translated at the time of ribosome capture. To estimate P-site offsets, a metagene-based strategy was employed. The underlying biological principle is that ribosomes exhibit consistent positioning during translation initiation, creating reproducible ”footprint” patterns. By aggregating reads aligned around the start

codon across many transcripts, it becomes possible to detect the most likely position of the ribosomal P-site for each read length.

According to precious literature [12], RPFs were grouped by read length and only reads that fully spanned the start codon, with at least 6 nucleotides upstream and downstream of the annotated CDS start, were selected. This window ensures that the start codon is centrally located within the read, improving the resolution of positional signal and reducing edge effects.

For each qualifying read, the distances from both its 5′ end and 3′ end to the CDS start were calculated. These distances represent possible offsets between the read extremities and the actual P-site, which corresponds to the active decoding center of the ribosome. The distributions of these distances across all reads of the same length were then analyzed, and the most frequent value (mode) was taken as the representative P-site offset for that read length. This mode reflects the empirical position where ribosomes most consistently locate relative to read boundaries, a common principle in Ribo-seq offset inference [12, 60].

In cases where the mode could not be reliably determined (often due to low read coverage or sparse representation of specific read lengths), a custom smoothing and imputation strategy was applied. If a read length lacked sufficient data to calculate a mode, the smallest non-missing offset in the dataset was used as a conservative estimate. The offsets were then refined by computing a local median from neighboring read lengths when available. For any remaining missing values, offsets were filled by propagating known values forward or backward to maintain continuity across the read-length spectrum. Edge cases (e.g., shortest and longest lengths) were handled by borrowing from nearby well-supported lengths, minimizing artificial discontinuities.

P-site Offset Estimation

For each read $r_i \in R$ of length L, define the spanning condition as:

$$\text{spans}(r_i, S) = (s_i \leq S - 6) \wedge (e_i \geq S + 6)$$

where S is the start codon position, ensuring the start codon lies within the read with sufficient flanking sequence for reliable offset estimation.

Then the P-site offsets are:

$$\begin{aligned} O_5(L) &= \text{mode}\left(\{S - s_i \mid \text{spans}(r_i, S)\}\right) \\ O_3(L) &= \text{mode}\left(\{e_i - S \mid \text{spans}(r_i, S)\}\right) \end{aligned} \tag{1.1}$$

Offsets from the 5′ and 3′ ends were computed independently. Each aligned read was annotated with three key features: its P-site position (calculated by applying the inferred offset to the alignment start), its reading frame (computed as the modulo-3 of the distance between the P-site and the CDS start), and its transcript region (5′ UTR, CDS, or 3′ UTR) based on the P-site coordinate.

In cases where multiple modes exist with equal frequency for length L, the offset is determined using neighboring length information:

$$O(L) = \begin{cases} O_5(L) \text{ and } O_3(L) & \text{if unique mode exists} \\ \text{mode}(M_k), k \in [L-1, L, L+1] & \text{if tied modes exist} \end{cases} \tag{1.2}$$

where M_k represents the set of modal values for length k. The neighborhood approach resolves ties by considering the most frequent offset across the local length range.

Following P-site and frame annotation, RPFs were annotated with the codon located at the inferred P-site position and the corresponding amino acid. Reads were filtered to retain only those mapping to valid transcriptome sequences. The codon was retrieved by slicing the transcript sequence at the P-site position, and the amino acid was assigned using a species-specific codon dictionary. Only in-frame reads located within CDS regions were considered to ensure codon relevant mapping. The resulting annotated dataset (transcript ID, codon, amino acid) supported codon-level ribosome occupancy analysis.

To evaluate ribosome occupancy across transcripts, read data were aggregated to compute total and in-frame read counts, CDS-normalized RPKM values, and codon-level coverage profiles. For nucleotide-resolution coverage, a per-nucleotide array was initialized for each transcript, and positions overlapping aligned reads were incremented. A separate array was constructed for CDS regions only to calculate coverage width. CDS coverage width, the proportion of CDS nucleotides covered by at least one read, was computed as a measure of translation completeness.

These computations were performed using batch-wise streaming and memory-efficient processing with Polars (version 1.31). Output was saved in compressed Parquet format for coverage vectors and CSV format for summary statistics.

Coverage Width Calculation

$$\text{Coverage Width} = \frac{|\{j \in [1, L_{\text{CDS}}] \mid \text{cov}(j) > 0\}|}{L_{\text{CDS}}} \tag{1.3}$$

where $\text{cov}(j)$ represents the number of reads covering position j within the coding sequence, and positions outside the CDS are excluded from both numerator and denominator.

After this process, the range of read lengths exhibiting strong in-frame signal (typically corresponding to monosome footprints) was determined. All analyses were then repeated using only reads within this range to generate refined, translation-specific metrics.

1.2.2.2 Visualization and Reporting

To facilitate interpretation, visualization scripts were developed using Polars (version 1.31.0), Pandas (version 2.2.3), and Bokeh (version 3.6.3). All outputs were compiled into modular HTML reports.

Read length distributions were visualized as bar plots, with the x-axis representing RPF lengths and the y-axis indicating their relative abundance as a percentage of total mapped reads. This visualization served multiple purposes: it enabled the identification of dominant read lengths (typically corresponding to monosome-protected fragments), revealed degradation patterns through the presence of shorter fragments, and helped detect disomes or other higher-order complexes through longer read length peaks. The plots provided a rapid qualitative assessment of library quality and sample-specific footprint profiles.

To assess the frame and regional distribution of RPFs, region-specific frame tile plots were generated. These tile plots visualized the fraction of reads falling into each of the three possible reading frames (0, 1, or 2), stratified by both read length and transcript region (5′ UTR, CDS, and 3′ UTR). Each cell in the tile plot represented the proportion of reads of a given length aligning in a particular frame within a specific region. This detailed resolution enabled evaluation of translation fidelity, especially the dominance of frame 0 reads within the CDS, a hallmark of high-quality ribosome profiling data. To enhance interpretability, excluding the read lengths that contributed a negligible fraction of total reads, thereby highlighting the frame distributions of the most translation-informative lengths and reducing noise from rare or artifact-prone fragments.

Global metaprofiles of ribosome occupancy were constructed by aligning P-site positions relative to annotated translation start and stop codons across all transcripts. These metagene plots captured the aggregated distribution of ribosomes around initiation and termination sites, offering insights into initiation pausing, elongation ramp-up, and termination-associated ribosome dwell times. Peaks near the start codon confirmed accurate P-site offset estimation, while smooth drops past the stop codon suggested proper termination events without extensive readthrough.

To interrogate triplet periodicity and its dependence on read length and transcript region, two complementary visualizations were generated from the annotated Ribo-seq table (psiteData.csv) (i) a *row-normalized frame map* per read length (`RowFrame.py`), and (ii) a *global frame composition* across read lengths (`totFrame.py`). Both procedures stratify reads by transcript region ($R \in \{5'\text{UTR}, \text{CDS}, 3'\text{UTR}\}$).

For region R, read length $L \in \mathbb{N}$, and reading frame $f \in \{0, 1, 2\}$ (computed as (P-site − CDS start) $\bmod 3$), let

$$F_{L,f}^{(R)} = |\{r_i : \text{region}(r_i) = R \wedge \text{length}(r_i) = L \wedge \text{frame}(r_i) = f\}|.$$

Define the within-length totals $T_L^{(R)} = \sum_{f=0}^{2} F_{L,f}^{(R)}$ and the region total $N^{(R)} = \sum_L \sum_f F_{L,f}^{(R)}$.

Calculation of contiguous length window

To remove spurious tails of scarcely represented lengths while keeping a contiguous band, first, row-normalization per length applied with a small cutoff $\tau = 10^{-3}$ (0.1%):

$$P_{L,f}^{(R)} = \frac{F_{L,f}^{(R)}}{T_L^{(R)}}, \qquad \mathcal{L}_\tau^{(R)} = \left\{L : \max_f P_{L,f}^{(R)} \geq \tau\right\}. \tag{1.4}$$

The working read-length range is then restricted to

$$L_{\min}^{(R)} = \min \mathcal{L}_\tau^{(R)}, \qquad L_{\max}^{(R)} = \max \mathcal{L}_\tau^{(R)}, \tag{1.5}$$

and all subsequent summaries use lengths $L \in \left[L_{\min}^{(R)}, L_{\max}^{(R)}\right]$.

Row-normalized frame map

For each $L \in [L_{\min}^{(R)}, L_{\max}^{(R)}]$ and frame f, the row-normalized proportion:

$$P_{L,f}^{(R)} = \frac{F_{L,f}^{(R)}}{T_L^{(R)}} \tag{1.6}$$

is visualized, producing one horizontal "row" per read length and three color-coded cells per row (frames $0, 1, 2$). The 0.1% cutoff is used solely to determine $[L_{\min}, L_{\max}]$; all counts within that window contribute to $P_{L,f}^{(R)}$. Three panels are produced side-by-side for 5′ UTR, CDS, and 3′ UTR using a shared color scale to facilitate comparison across regions.

Global frame composition map

In addition to the per-length view, a global frame composition is computed within the same contiguous length window by aggregating over L:

$$Q_f^{(R)} = \frac{\sum_{L=L_{\min}^{(R)}}^{L_{\max}^{(R)}} F_{L,f}^{(R)}}{\sum_{L=L_{\min}^{(R)}}^{L_{\max}^{(R)}} \sum_{f'=0}^{2} F_{L,f'}^{(R)}} = \frac{\sum_L F_{L,f}^{(R)}}{N_{\text{window}}^{(R)}}, \tag{1.7}$$

where $N^{(R)}_{\text{window}}$ denotes the total number of reads in region R within the retained length band. The CDS table underlying this panel is exported to CSV for downstream use.

The row-normalized map emphasizes *periodicity by length* (e.g., a dominant frame 0 band in CDS for monosome-length RPFs) and reveals off-frame enrichment in UTRs, whereas the global map summarizes *overall frame bias* within the empirically supported length band. Together, they provide complementary checks of triplet periodicity and read-length consistency.

Codon usage patterns were investigated at codon resolution by extracting the triplet codon at each in-frame P-site position within CDS regions. Using a reverse mapping from codons to amino acids, synonymous codons were grouped, and their frequencies were calculated as percentages of all codons. These values were visualized as grouped bar plots, with codons color-coded by amino acid. This allowed for the detection of codon usage biases, ribosomal pausing at specific codons, and potential elongation dynamics across synonymous codon groups.

For each codon c encoding amino acid a, let $f(c)$ be the number of in-frame P-sites mapping to c. Then:

$$\text{Codon Usage}(c) = \frac{f(c)}{\sum_{c' \in \text{Codons}} f(c')} \tag{1.8}$$

To explore transcript-level translational activity, detailed coverage profiles were generated for the 20 transcripts with the highest in-frame ribosome occupancy (i.e., frame 0 P-site counts within the CDS). Each transcript was visualized individually, showing nucleotide-level read coverage along its full length and highlighting CDS boundaries with annotated start and stop codons. The underlying nucleotide sequence was overlaid, facilitating identification of initiation peaks, elongation pauses, and potential premature termination events.

Complementing these high-resolution views, CDS coverage width was quantified for each transcript as the proportion of coding nucleotides covered by at least one read. This metric provided a coverage-normalized indicator of how completely the coding region was sampled by ribosomes. A histogram of CDS coverage widths across all transcripts was plotted to visualize the overall distribution of coverage completeness in each sample, serving as a proxy for transcriptome-wide translation consistency.

Finally, global summary plots were produced to offer an at-a-glance assessment of dataset quality. These included bar plots displaying the number of transcripts that surpassed predefined thresholds for minimal in-frame coverage or sufficient CDS coverage width. These summaries enabled rapid comparisons across samples and helped identify

technical or biological variability in ribosome occupancy. In all visualizations used output **Scalable Vector Graphics (SVG)** backend for publication-ready exports.

1.2.3 Module 3: Condition Comparison - Comparative Analysis of Ribosome Profiling Data

This module of the Martian pipeline enables comparative analysis of RiboSeq data across experimental conditions. It includes procedures for differential expression analysis, codon-level occupancy comparison, and translation frame distribution shifts, using statistical modeling, dimensionality reduction, and custom visualization scripts. All analyses were implemented in Python (version 3.12.4) using Polars (version 1.31.0), NumPy (version 1.26.4), SciPy (version 1.15.1), and Bokeh (version 3.6.3) for computation and visualization.

1.2.3.1 Data Input and Structure

The analysis requires a curated directory of annotated ribosome profiling data. Each sample is expected to have a corresponding `{sample}_psiteData.csv` file that contains per-read annotations including the P-site position, frame, region assignment (5′ UTR, CDS, or 3′ UTR), codon identity, and associated transcript metadata. Sample grouping is defined through a dictionary-like mapping to experimental conditions (e.g., `Condition A`, `Condition B`).

All input files are read using `polars` for efficient streaming and memory usage, and are joined based on shared annotation schema to enable per-sample and per-condition comparisons.

1.2.3.2 Differential Translation Analysis

Differential ribosome occupancy is assessed at the transcript level using the DESeq2 (version 1.42.0) package within rpy2 (version 3.5.11) conversion environment. First, the count matrix K is assembled such that K_{ij} is the number of in-frame (frame = 0) P-site counts mapped to transcript i in sample j.

DESeq2 models K_{ij} using a Negative Binomial distribution [89]:

$$K_{g,s,j} \sim \mathrm{NB}(\mu_{g,s,j}, \alpha_g) \tag{1.9}$$

with

$$\log(\mu_{g,s,j}) = \log(s_s) + \beta_{g,0} + \beta_{g,1} x_j \tag{1.10}$$

where s_s is the size factor for sample s, x_j is the binary condition indicator, $\beta_{g,0}$ is the baseline expression, and $\beta_{g,1}$ represents the $\log_2$ fold change.

To test for condition-dependent effects, the expected counts are modeled as:

$$\log(\mu_{ij}) = \log(s_j) + \beta_{i0} + \beta_{i1} x_j \tag{1.11}$$

where x_j is the binary condition label (e.g., control vs. treatment), β_{i0} is the baseline expression, and β_{i1} represents the log2 fold change in ribosome occupancy. A Wald test [90] is used to assess the null hypothesis $H_0 : \beta_{i1} = 0$. Multiple hypothesis testing is addressed using the Benjamini-Hochberg [91] procedure to control the false discovery rate.

Functional Enrichment Analysis of Differentially Translated Genes

To provide biological context for translation changes, significantly differentially translated genes (identified through the procedures above) underwent Gene Ontology (GO) and pathway enrichment analysis. The function accepts a sample-to-type map, a differential results container, the target organism code for g:Profiler [92] (version 1.0.0) (e.g. hsapiens), a $\log_2$(fold change) threshold, and an output directory.

For each non-control type T, the corresponding differential table was retrieved as Table[T_vs_Control]. Genes with adjusted $P-value \leq 0.05$ were considered significant, and then split into up- and down-regulated sets using a symmetric $\log_2$(fold change) cutoff:

$$\begin{aligned} G_{\uparrow}(T) &= \{\, g : \mathrm{padj}_g \leq 0.05,\ \log_2 \mathrm{FoldChange} g > t \,\}, \\ G_{\downarrow}(T) &= \{\, g : \mathrm{padj}_g \leq 0.05,\ \log_2 \mathrm{FoldChange}_g < -t \,\}, \end{aligned} \tag{1.12}$$

where t= threshold.

Each of these two gene lists (if non-empty) was submitted to g:Profiler via the gprofiler Python client using the provided species code and otherwise default settings (no custom background). g:Profiler returns a table of terms with raw p-values. Within each query, p-values were adjusted by the Benjamini–Hochberg procedure to control the false discovery rate: $padj_j = \mathrm{BH}(p_j)$, and stored in a padj column. For convenience, results were sorted by the reported source and written to tab-separated files.

Visualization was done with Bokeh. For each of the up and down sets, a jittered scatter plot was produced in which the x-axis enumerates the term sources are **"GO:MF", "GO:BP", "GO:CC", "KEGG", "REAC", "WP", "TF", "MIRNA", "HPA", "CORUM", "HP"**, where;

- GO:MF = Gene Ontology: Molecular Function,
- GO:BP = Gene Ontology: Biological Process,

- GO:CC = Gene Ontology: Cellular Component ,
- KEGG = Kyoto Encyclopedia of Genes and Genomes pathways ,
- REAC = Reactome pathways ,
- WP = WikiPathways ,
- TF = TRANSFAC transcription factor binding site predictions ,
- MIRNA = miRTarBase miRNA targets ,
- HPA = Human Protein Atlas expression data ,
- CORUM = A manually curated database of mammalian protein complexes,
- HP = Human Phenotype Ontology,

and the y-axis is $-\log_{10}(padj)$. A horizontal reference line was drawn at $-\log_{10}(0.05)$. Interactive tools and hover tooltips for GO term ID, name, and $-\log_{10}(q)$ were enabled. For each comparison T_vs_Control, the two plots ("upGO" and "downGO") were placed in tabs, and comparisons themselves were organized as top-level tabs.

1.2.3.3 Dimensionality Reduction via PCA

Principal Component Analysis (PCA) is used to visualize sample clustering and detect outliers. The input to PCA is the rlog-transformed count matrix $\widetilde{K}$, where:

$$\widetilde{K}ij = \log_2(Kij + c) \tag{1.13}$$

with a small pseudocount c to avoid undefined values.

The covariance matrix Σ is computed from the centered matrix and decomposed as:

$$\Sigma = V\Lambda V^T \tag{1.14}$$

where V contains the eigenvectors (principal components) and Λ is a diagonal matrix of eigenvalues. Samples are projected into PC space as:

$$Z = V^T \widetilde{K}^c \tag{1.15}$$

1.2.3.4 Codon-Level Occupancy Comparison

To compare ribosome occupancy at the codon level, relative codon usage ratios across conditions were calculated. Let C_{ij} be the count of in-frame P-sites mapped to codon j

encoding amino acid i in a given condition. Then, relative occupancy is:

$$R_{ij} = \frac{C_{ij}}{\sum_j C_{ij}} \tag{1.16}$$

The condition-wise codon usage is compared using a pseudo-log2 ratio:

$$\text{CodonRatio}ij = \log_2\left(\frac{Rij^{(A)} + \epsilon}{R_{ij}^{(B)} + \epsilon}\right) \tag{1.17}$$

where ϵ is a small pseudocount (10^{-10}) to avoid division by zero. These ratios are visualized using grouped bar plots.

1.2.3.5 Visualization and Reporting

All outputs, including PCA plots, length distribution barplots, metagene profile plots, differential expression volcano plots, codon usage comparisons, and frame distribution heatmaps, were generated using Bokeh and exported as HTML dashboards with SVG support. Visual summaries support the interpretation of ribosome dynamics and identification of translationally differentiating features.

1.2.4 Module 4: Translation Efficiency Analysis

This module integrates the P-site annotated data from Module 2 with matched RNA-seq data, using the in-frame read classifications to compute translation efficiency (TE) metrics. It expects: **(i)** a sample table with `SampleID`, `Type` $\in$ {RNASeq, RiboSeq}, `Condition`, and file paths of genomic alignment files of RNASeq and RiboSeq; **(ii)** a comparison table specifying the `Control` and `Treatment`; **(iii)** a reference GTF; and **(iv)** Ribo-seq tables produced earlier in the ribosome profiling module. Outputs include gene-level count files, DESeq2 result tables per contrast and assay, and interactive HTML reports (volcano, PCA, and directional plots).

1.2.4.1 Preprocessing and Gene-level quantification (HTSeq)

For each RiboSeq samples, **read IDs** gathered from the reads that are in-frame (frame = 0) to ensure those RPFs counted over CDSs are translation-informative. The original Binary Alignment Map (BAM) file is indexed and filtered for those read IDs. This yields BAM files containing only in-frame footprints.

For each sample, a BAM file is counted with `htseq-count` (`--format bam`, `--order pos`). Feature class and strandedness are provided by the module parameters and differ for RNA-seq and Ribo-seq:

- RNA-seq: features typically set to `exon`; strandedness set by user.
- Ribo-seq: features typically set to `CDS`; strandedness set by user.

The counting mode (`--mode`) is also parameterized (e.g. `intersection-strict`). Raw gene-by-sample tables are written as `{SampleID}_{Type}.txt`.

Count tables are merged by gene ID. Genes with very low information content were removed prior to modeling to stabilize dispersion estimates. Specifically,

$$\sum_{i=1}^{k} sample_i(count_{gene}) \leq 10.$$

The design matrix contains `SampleID`, `Condition` (control vs. treatment), and `Type` (RNASeq vs. RiboSeq).

1.2.4.2 Differential modeling with DESeq2

Differential abundance analysis performed with count tables created in pre-processing. Differential abudance analysis modeled as followed;

Let K_{gjs} denote the raw count for gene g, assay type $j \in \{\text{RNA}, \text{Ribo}\}$, and sample s with size factor s_s. DESeq2 models

$$K_{gjs} \sim \text{NB}(\mu_{gjs},\, \alpha_g) \tag{1.18}$$

$$\log \mu_{g,s,j,t} \;=\; \log(s_{s,t}) \;+\; \beta_{g,0} \;+\; \beta_{g,\text{Condition}}\, x_s \;+\; \beta_{g,\text{Type}}\, t_j \;+\; \beta_{g,\text{Interaction}}\, x_s t_j \tag{1.19}$$

where $x_s \in \{0, 1\}$ encodes Control/Treatment and $t_j \in \{0, 1\}$ encodes Type with RNA as the baseline ($t_{\text{RNA}} = 0$, $t_{\text{Ribo}} = 1$). Three result tables are produced per contrast:

1. **RNA-seq model:** design $\sim$ Condition on RNA-seq counts only; test

$$H_0 : \beta_{g,\text{Condition}} = 0$$

2. **Ribo-seq model:** design $\sim$ Condition on Ribo-seq counts only; test

$$H_0 : \beta_{g,\text{Condition}} = 0$$

3. **TE model (interaction):** design $\sim \mathrm{Condition} * \mathrm{Type}$ on the combined matrix; test

$$H_0 : \beta_{g,\mathrm{int}} = 0,$$

which captures the condition effect in Ribo-seq *relative* to RNA-seq. Wald statistics are used; P-values are adjusted by Benjamini–Hochberg to control FDR, yielding padj.

Interpretation of the TE coefficient.

Let RNA and Ribo denote the two assays, and let C (control) and T (treatment) denote the conditions. From (1.19), the difference-in-differences that defines the TE change is

$$\underbrace{\left[\log \mu_{g,\mathrm{Ribo},T} - \log \mu_{g,\mathrm{RNA},T}\right]}_{\text{log TE in Treatment}} - \underbrace{\left[\log \mu_{g,\mathrm{Ribo},C} - \log \mu_{g,\mathrm{RNA},C}\right]}_{\text{log TE in Control}} = \beta_{g,\mathrm{int}}. \tag{1.20}$$

Thus, $\beta_{g,\mathrm{int}}$ (reported as a $\log_2$ fold change in DESeq2 output) estimates the condition-induced change in $\log \mathrm{TE}$.

1.2.4.3 Principal component analysis (PCA)

Normalized counts from each fitted DESeq2 object are used to compute PCA (PC1 vs. PC2). Samples are colored by condition and annotated by `Type`. Axis labels display the percent variance explained.

1.2.4.4 Volcano plots and thresholds

For each result table (RNA-seq, Ribo-seq, TE), volcano plots display $\log_2 \mathrm{FC}$ vs. $-\log_{10}(\mathrm{padj})$. Significance is defined by $\mathrm{padj} \leq 0.05$ and a user-specified effect size threshold $t > 0$ on $|\log_2 \mathrm{FC}|$. A horizontal reference line indicates $-\log_{10}(0.05)$.

1.2.4.5 Directional categorization of regulation

For each differential expression analysis performed with RiboSeq and RNASeq, the following criteria were applied to show the driven force (transcription or translation) for each individual gene.

Let Δ_g^{RNA} and Δ_g^{Ribo} be the gene-level $\log_2$ fold changes from the RNA-seq-only and Ribo-seq-only models, respectively. With threshold $t > 0$, genes are classified as:

$$\text{Homo-direction: } \left(\Delta_g^{\mathrm{RNA}} \geq t \ \textit{and}\ \Delta_g^{\mathrm{Ribo}} \geq t\right) \ \textit{or} \ \left(\Delta_g^{\mathrm{RNA}} \leq -t \ \textit{and}\ \Delta_g^{\mathrm{Ribo}} \leq -t\right), \tag{1.21}$$

$$\text{Cross-direction: } \left(\Delta_g^{\mathrm{RNA}} \geq t \ \textit{and}\ \Delta_g^{\mathrm{Ribo}} \leq -t\right) \ \textit{or} \ \left(\Delta_g^{\mathrm{RNA}} \leq -t \ \textit{and}\ \Delta_g^{\mathrm{Ribo}} \geq t\right), \tag{1.22}$$

$$\text{Translation-dominant: } \left(|\Delta_g^{\mathrm{RNA}}| < t \ \textit{and}\ \Delta_g^{\mathrm{Ribo}} \geq t\right) \ \textit{or} \ \left(|\Delta_g^{\mathrm{RNA}}| < t \ \textit{and}\ \Delta_g^{\mathrm{Ribo}} \leq -t\right), \tag{1.23}$$

$$\text{Transcription-dominant: } \left(|\Delta_g^{\mathrm{Ribo}}| < t \ \textit{and}\ \Delta_g^{\mathrm{RNA}} \geq t\right) \ \textit{or} \ \left(|\Delta_g^{\mathrm{Ribo}}| < t \ \textit{and}\ \Delta_g^{\mathrm{RNA}} \leq -t\right), \tag{1.24}$$

$$\text{Stable: } \left(|\Delta_g^{\mathrm{RNA}}| < t \ \textit{and}\ |\Delta_g^{\mathrm{Ribo}}| < t\right). \tag{1.25}$$

The directional plots show $\log_2 \mathrm{FC}_{\mathrm{Ribo}}$ (x-axis) vs. $\log_2 \mathrm{FC}_{\mathrm{RNA}}$ (y-axis) with vertical and horizontal reference bands at $\pm t$.

1.2.4.6 Multiple testing and outputs

All p-values are adjusted with the Benjamini–Hochberg procedure within each contrast/model to control the false discovery rate. The module writes:

- HTSeq tables: `Counts/{SampleID}_{Type}.txt`.
- TE results: tab-delimited files with `gene_id`, `gene_name`, `baseMean`, `log2FoldChange`, `lfcSE`, `stat`, `pvalue`, `padj`.
- Reports: interactive HTML pages (volcano, PCA, directional).

1.2.5 Module 5: nano-tRNA-Seq Quantification and Position-Specific Modification Analysis

1.2.5.1 Notation and Definitions

Throughout this section, the following notations was used:

- r_i = individual sequencing read i
- L_{ref} = reference sequence length
- α_5, α_3 = 5' and 3' adapter lengths respectively

- δ = margin tolerance for alignment boundaries
- $\mu_{g,s}$ = expected count for gene g in sample s
- $q_{\min}$ = minimum mapping quality threshold
- $C_{g,s}$ = observed count for gene g in sample s

The module processes Direct RNA Seqeuncing (DRS) alignments with a curated tRNA reference to **(i)** quantify tRNA abundances, **(ii)** test the differential abundance between conditions, and **(iii)** derive single-nucleotide basecall profiles to test for position-specific differential modification signals.

It takes as input;

- a tab-delimited sample sheet (`Sample Path`, `Condition`, optional `Experiment`, `Barcode`),
- a tRNA reference FASTA for counting,
- an aligned FASTA (converted from Stockholm) embedding Sprinzl numbers [93] indices in `reference_annotation`,
- and a contrast table specifying `Treatment` vs. `Control`.

with key parameters; $q_{\min}$(default: 10), a_5(default: 24nt), a_3(default: 30nt), δ(default: 0nt), t(default: 0.5), optional read-class filters:

- `--complete` (require full-length),
- `--fragment` (exclude full-length),
- when none of them selected it returns all reads (full-length & fragment)

and library orientation `--firststrand` (antisense cDNA vs. sense DRS).

1.2.5.2 nanoCount: Read Processing and Quantification

The **nanoCount** module serves as the primary data preporcessing engine for nanotRNA-Seq analysis. It performs length filtering for reads, quality filtering, nucleotide level quantification and base-calling error quantification of nanotRNA-Seq data.

Read filtering and Quality Control

Sequenced reads undergo multi-stage filtering to ensure data quality, sequence integrity and, remove artifacts that could cause noise on downstream analysis.

Strand specificity check: For directional sequencing protocols:

$$\text{antisense}(r_i) = \begin{cases} \text{True} & \text{if (!firststrand and reverse}(r_i)\text{) or (firststrand and !reverse}(r_i)\text{)} \\ \text{False} & \text{otherwise} \end{cases} \tag{1.26}$$

Where firststrand indicates first-strand synthesis protocols (cDNA sequencing) versus direct RNA sequencing (DRS).

Three prime adapter integrity check: In nano-tRNASeq method motor protein binds to $3'$ adapter sequence, all sequenced reads that not contain complete $3'$ adapter are considered sequencing artifacts:

$$\text{valid}(r_i) = \begin{cases} \text{True} & \text{end}(r_i) \geq L_{\text{ref}} - \delta \\ \text{False} & \text{otherwise} \end{cases} \tag{1.27}$$

where $L_{\text{ref}} = a_5 + Length_{tRNA} + a_3$.

Completeness assessment: Full-length tRNA detection is crucial for accurate quantification, as partial reads may introduce systematic biases in abundance estimates. A read was classified as complete based on its alignment boundaries relative to adapter sequences:

$$\text{complete}(r_i) = \begin{cases} \text{True} & \text{if start}(r_i) \leq a_5 \text{ and end}(r_i) \geq L_{\text{ref}} - a_3 + \delta \\ \text{False} & \text{otherwise} \end{cases} \tag{1.28}$$

Histidine tRNAs contain an additional guanine nucleotide (G_1) at position -1 of the acceptor stem, which is added post-transcriptionally by $tRNA^{His}$ guanylyltransferase [94, 95]. This additional nucleotide creates a longer sequence that reduces efficient $5'$ adapter hybridization. This structural feature necessitates extended adapter tolerance ($\alpha_5 = 44$ nt instead of 24 nt) to accommodate the reduced adapter binding efficiency while maintaining specificity for full-length tRNA detection.

Alignment Coverage Analysis and Quality Metrics

Alignment coverage provides information on read integrity and potential systematic biases in library preparation or sequencing.

Coverage proportion calculation: For each successfully aligned read, the alignment proportion quantifies the fraction of the reference sequence covered:

$$p_{align}(r_i) = \frac{L_{aligned}(r_i)}{L_{\text{ref}}(r_i)} \tag{1.29}$$

where $L_{aligned}(r_i)$ represents the aligned length, excluding soft-clipped regions. Coverage proportions approaching 1.0 indicate full-length reads, while systematically lower values may reveal fragmentation.

Alignment Statistics Generation: The module generates comprehensive alignment statistics including:

- Total demultiplexed reads per sample,
- Percentage of reads passing quality filters,
- Antisense read proportions (indicating protocol specificity),
- Adapter detection rates (5' and 3' adapters separately),
- Unique mapping percentages versus multi-mapping reads.

Single-Nucleotide Resolution Counting

Nucleotide-level quantification enables detection of RNA modifications through analysis of systematic base-calling deviations from reference sequences. The module employs *pysam mpileup* with optimized parameters to extract base composition at each genomic position:

```
pysam.mpileup(--ff 3840 --no-output-ends --no-output-del \\
--no-output-ins -q {mapq} -Q 0 -f {fasta} {bam} -r {reference})
```

where:

- `--ff` 3840: Excludes secondary alignments and PCR duplicates,
- -q {mapq}: Applies mapping quality filter,
- -Q 0: Includes all base qualities (Nanopore-specific).

Pileup strings cleaned from the elements of "+N" and "-N", where N indicates the sequence of insertions or deletions to count insertion and deletion events rather than their lengths to focus on modification-induced artifacts. At each genomic position j in reference sequence k, the base composition vector was defined as:

$$b_{j,k,s} = [A_{j,k,s}, T_{j,k,s}, G_{j,k,s}, C_{j,k,s}, I_{j,k,s}, D_{j,k,s}]^T \tag{1.30}$$

where:

- $A_{j,k,s}, T_{j,k,s}, G_{j,k,s}, C_{j,k,s}$ = counts of respective nucleotides at position j in reference k for sample s,

- $I_{j,k,s}$ = insertion count (each element fits to the regex pattern: "\+\d+")
- $D_{j,k,s}$ = deletion count (asterix symbols "*" in pileup string)

Reference Position Mapping: Each position is annotated with its corresponding Sprinzl number, enabling standardized comparison across different tRNA sequences:

$$\text{Position}(j,k) = \begin{cases} \text{S1...S24} & \text{if } j \leq 24 \text{ (5}' \text{ adapter region)} \\ \text{Sprinzl numbering} & \text{if } 24 < j \leq L_{\text{ref}} - 30 \\ \text{S25...S54} & \text{if } j > L_{\text{ref}} - 30 \text{ (3}' \text{ adapter region)} \end{cases} \tag{1.31}$$

Error Rate Calculation and Quality Assessment

Systematic error analysis done by sum of errors methodology which identifies position with elevated mismatch rated that may indicate RNA modifications or sequencing artifacts. The sum of errors ratio at each position is computed as:

$$e_{j,k} = \frac{\sum_{b \neq ref_{j,k}} b_{j,k}}{\sum_{b \in \{A,T,G,C,I,D\}} b_{j,k}} \tag{1.32}$$

where $\text{ref}_{j,k}$ is the reference base at position j in sequence k.

Count Matrix Generation and Data Structure

The module generates multiple output matrices optimized for different downstream analyses: **Gene-Level Count Matrix:** Primary quantification aggregates reads mapping to each tRNA reference:

$$C_{g,s} = \sum_{r_i \in R_{g,s}} w(r_i) \times filter(r_i) \tag{1.33}$$

where:

- $R_{g,s}$ represents all reads mapping to gene g in sample s,
- $w(r_i) = 1$ for uniquely mapped reads,
- $filter(r_i)$ is 1 if read passes all quality filters, 0 otherwise

Single-Nucleotide Count Matrix: Position-specific quantification preserves base-resolution information:

$$M_{j,k} = [b_{j,k,1}, b_{j,k,2}, ..., b_{j,k,n}] \tag{1.34}$$

where each column represents a sample's base composition vector at position (j,k).

Each count matrix includes comprehensive metadata with sprinzl position annotations, reference base identities and quality metrics and filtering statistics.

1.2.5.3 nanoDiff: Differential tRNA Expression Analysis

The **nanoDiff** module implements sophisticated statistical frameworks for identifying condition-dependent changes in tRNA abundance. This analysis addresses unique challenges in tRNA quantification, including sequence similarity between family members, variable expression levels across amino acid types, and the need for robust statistical inference with typically modest sample sizes.

DESeq2 Statistical Framework

Differential expression analysis employs the DESeq2 framework (via rpy2 library on Python), which models count data using negative binomial distributions to account for overdispersion commonly observed in RNA-seq data. It uses all assumptions of DESeq2's previously explained **Generalized Linear Model, Negative Binomial Distribution and Wald Test**.

The module implements a robust three-stage approach for estimating dispersion parameters:

- **Parametric Fit:** Initial estimation assumes a parametric relationship between mean and dispersion $\alpha_g = \frac{a}{\mu_g} + b$,
- **Local Regression:** If parametric fitting fails, local regression (LOWESS) models the mean-dispersion relationship non-parametrically.
- **Mean Dispersion:** As a final fallback, all genes are assigned the mean dispersion across the dataset $\alpha_g = \frac{1}{G} \sum_{h=1}^{G} \hat{\alpha}_h$

where α_g represents dispersion parameter, μ_g represents mean. This three-stage approach enables continuous sample processing without sacrificing statistical power.

1.2.5.4 modFinder: RNA Modification Detection and Analysis

RNA modifications alter the chemical structure of nucleotides, affecting their electrical resistance and sequencing speed during nanopore sequencing. This results in systematic base-calling deviations that are reproducible and position-specific, unlike random sequencing errors, which occur stochastically. The **modFinder** algorithm exploits this signal-to-noise difference to identify genuine modification events by analyzing the underlying mechanism of these systematic errors.

Standard base-calling tools for nanopore sequencing rely on machine learning (ML) algorithms that are trained to recognize the four canonical bases (A, T, G, C). When these algorithms encounter modified bases, they lack the training data to classify them correctly and instead force a prediction to one of the four canonical bases. This creates the systematic misclassification patterns [96] that modFinder detects by identifying positions where base-calls consistently deviate from the expected sequence in a non-random manner, modified positions can be accurately identified and quantified.

Modification Signal Preprocessing and Count Categorization

Modification-Sensitive Count Classification: For each Sprinzl-numbered position j in sample s, nucleotide observations are classified based on their consistency with the reference sequence:

$$\text{mod count}_{j,s} = \sum_{b \neq \text{ref}_j} b_{j,s} = A_{j,s} + T_{j,s} + G_{j,s} + C_{j,s} + I_{j,s} + D_{j,s} - b_{\text{ref},j,s} \tag{1.35}$$

$$\text{unmod count}_{j,s} = b_{\text{ref},j,s} \tag{1.36}$$

where ref_j is the canonical reference base at Sprinzl position j. This binary classification assumes that:

- Unmodified positions generate reads matching the reference sequence
- Modified positions generate systematic deviations (mismatches, indels)
- Technical errors are randomly distributed and canceled out by statistical analysis

Statistical Model for Differential Modification Analysis

The core innovation of modFinder is the use of an interaction model that simultaneously analyzes modification levels and their condition-dependent changes. By separating the counts of modified and unmodified reads per position, statistical testing is enabled between conditions as follows:

$$\log_2(\mu_{j,s,t}) = \log_2(s_{s,t}) + \beta_{0,j} + \beta_{1,j} \cdot \text{Condition}_s + \beta_{2,j} \cdot \text{Type}_t + \beta_{3,j} \cdot (\text{Condition}_s \times \text{Type}_t) \tag{1.37}$$

where:

- j indexes Sprinzl positions
- s indexes samples
- $t \in \{\text{mod}, \text{unmod}\}$ indexes read type

- $Condition_s \in \{\text{control, treatment}\}$ (reference level = control)
- $Type_t \in \{\text{mod, unmod}\}$ (reference level = unmod)
- $\beta_{0,j}$: Baseline unmodified read count in control samples
- $\beta_{1,j}$: Condition effect on unmodified reads (expression change)
- $\beta_{2,j}$: Modification effect in control samples (baseline modification level)
- $\beta_{3,j}$: **Interaction term**, condition-specific change in modification level

The interaction coefficient $\beta_{3,j}$ directly measures differential modification:

$$\Delta_{mod,j} = \frac{mod\,ratio_{treatment,j}}{mod\,ratio_{control,j}} = e^{\beta_{3,j} ln(2)} \tag{1.38}$$

To ensure all Sprinzl positions are analyzed across all tRNAs, the reference annotation is systematically expanded:

$$\mathrm{R}_{expanded} = \bigoplus_{t \in T} \mathrm{R}_{base} \tag{1.39}$$

where $\bigoplus$ denotes the direct sum (vertical concatenation) and T is the set of all tRNA genes. This creates a complete position × gene matrix:

$$dim(\mathrm{R}_{expanded}) = |T| \times |\text{Sprinzl positions}| \tag{1.40}$$

Missing combinations (tRNA-position pairs without reads) are filled with zero counts, ensuring statistical analysis covers the complete modification landscape.

Modification count data often exhibit greater overdispersion than expression data due to position-specific technical noise and biological heterogeneity in modification efficiency. The negative binomial framework accommodates this through position-specific dispersion estimates ($\alpha_j = f(\mu_j, \sigma_j^2)$).

The interaction design naturally controls for batch effects that affect both modification and expression where

$$\text{Batch effect} = \beta_1 * \text{Batch} + \beta_2 * Type \tag{1.41}$$

the interaction term β_3 remains unbiased as long as batch effects are not differential with respect to modification status.

1.2.5.5 Visualization and Reporting Framework

The visualization module provides comprehensive graphical representations of analysis results, implementing both standard bioinformatics plots and specialized visualizations

for tRNA modification analysis. The module produces scatter plots, volcano plots, PCA plots, coverage plots per tRNA and cloverleaf secondary structure plots and heatmaps with differential modification analysis.

Modification-Specific Visualizations

Modification heatmaps display log2 fold changes across Sprinzl positions and tRNA families and color mapping accommodates both positive and negative fold changes. With this modification patterns become visually distinctive between loss of hypo and hyper-modification with respect to control samples. Statistically significant positions are highlighted with border annotations:

$$\text{border}(i,j) = \begin{cases} \text{thick black} & \text{if } q_{i,j} \leq 0.05 \\ \text{none} & \text{otherwise} \end{cases} \tag{1.42}$$

The canonical tRNA secondary structure produced as dot-plots with fixed points on xy coordinate system and it is represented using a standardized coordinate system as "$Coordinates : \mathbb{R}^2 \rightarrow$ Sprinzl positions" and each structural domain has a distinct background color:

$$\text{Color}(\text{region}) = \begin{cases} \text{Light green} & \text{Acceptor stem} \\ \text{Light blue} & \text{D-arm} \\ \text{Orange} & \text{Anticodon arm} \\ \text{Red} & \text{Anticodon triplet} \\ \text{Yellow} & \text{T}\psi\text{C arm} \\ \text{Purple} & \text{Variable loop} \\ \text{Dark green} & \text{Discriminator base} \end{cases} \tag{1.43}$$

while significantly modified positions are highlighted with red indicators superimposed on the structural diagram, enabling immediate identification of functionally relevant changes.

Interactive plots enable real-time exploration of high-dimensional datasets for:

1. **Sample selection dropdowns:** Toggle between different experimental comparisons,
2. **Gene family filtering:** Focus on specific tRNA types (e.g., only Leu tRNAs),
3. **Hover tooltips:** Display detailed statistics for individual data points.

This comprehensive methodological framework provides a complete foundation for nanotRNA-Seq analysis, addressing both the computational and statistical challenges inherent in small RNA modification detection and quantification.

1.3 Results

1.3.1 Benchmark Dataset and Experimental Design

Martian was benchmarked against two established ribosome profiling tools: riboWaltz and Plastid. The benchmark dataset consisted of 11 publicly available ribosome profiling samples from human cells with an average of 15 million aligned reads per sample [97]. All three tools were executed on identical hardware configurations (128 GB RAM, 32-core processor) with consistent input parameters to ensure direct comparison across runtime, memory usage, and analytical accuracy metrics.

1.3.2 Computational Performance and Resource Utilization

1.3.2.1 Runtime Efficiency

The average processing time varied among tools (Figure 1.2). Martian completed processing in 1,895.93 seconds (approximately 32 minutes), Plastid in 2,127.82 seconds (35 minutes), and riboWaltz in 3,028.95 seconds (51 minutes). Statistical analysis revealed significant differences in runtime between all pairwise comparisons (Martian vs. riboWaltz: $p_{FDR-adj}$ = 1.0510^{-4}; Martian vs. Plastid: $p_{FDR-adj}$ = 1.3110^{-5}; Plastid vs. riboWaltz: $p_{FDR-adj}$ = 8.2210^{-6}), with a moderate effect size (ω_p^2 = 0.45, 95% CI [0.14, 1.00]).

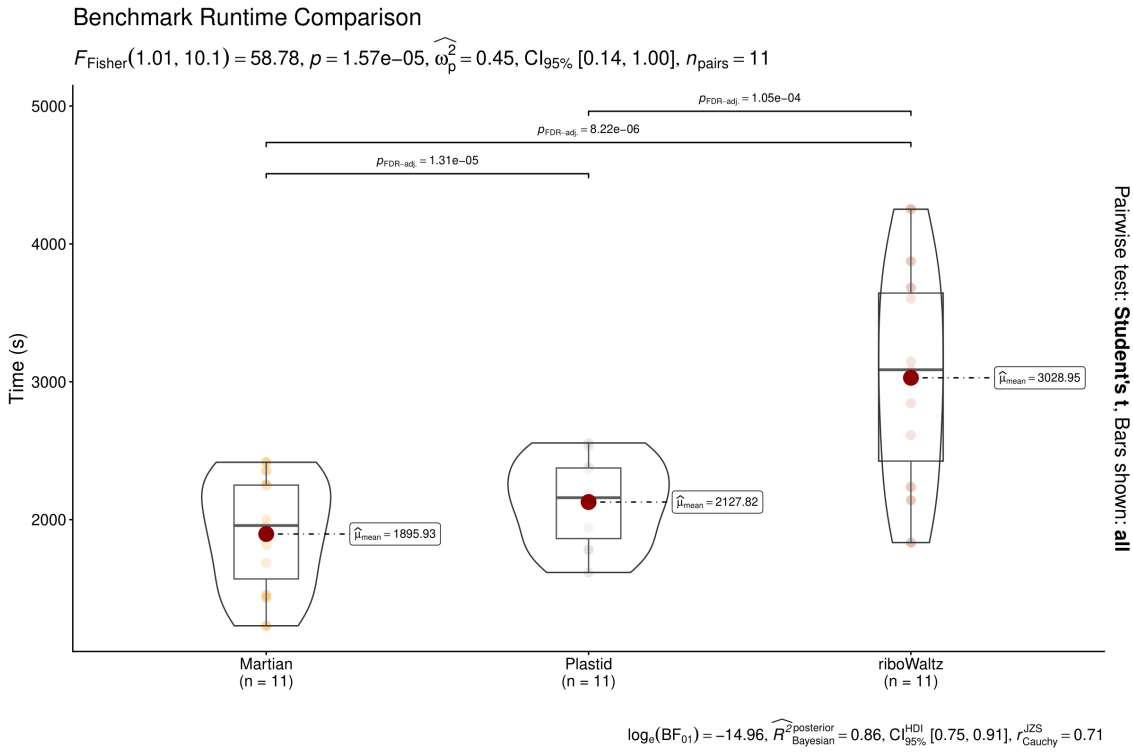

Figure 1.2: Runtime comparison across ribosome profiling tools. Violin plots show the distribution of processing times for 11 samples (n = 11 per tool) with mean values indicated. Martian achieved the fastest runtime (1,896 seconds), followed by Plastid (2,128 seconds) and riboWaltz (3,029 seconds). Statistical significance was assessed using one-way ANOVA with Student's t-test for pairwise comparisons (FDR-adjusted p-values shown).

1.3.2.2 Memory Usage

Memory consumption analysis revealed substantial differences in resource requirements among the three tools (Figure 1.3). Plastid exhibited the lowest memory usage with a mean of 547.64 MB (0.5 GB). Martian showed intermediate consumption at 20,987.36 MB (21 GB), representing a 38-fold increase over Plastid. riboWaltz displayed the highest memory requirements at 31,828.25 MB (32 GB), approximately 58-fold higher than Plastid and 1.5-fold higher than Martian. All pairwise comparisons showed highly significant differences ($p_{FDR-adj} < 10^{-6}$), with a large effect size ($\omega_p^2 = 0.87$, 95% CI [0.77, 1.00]).

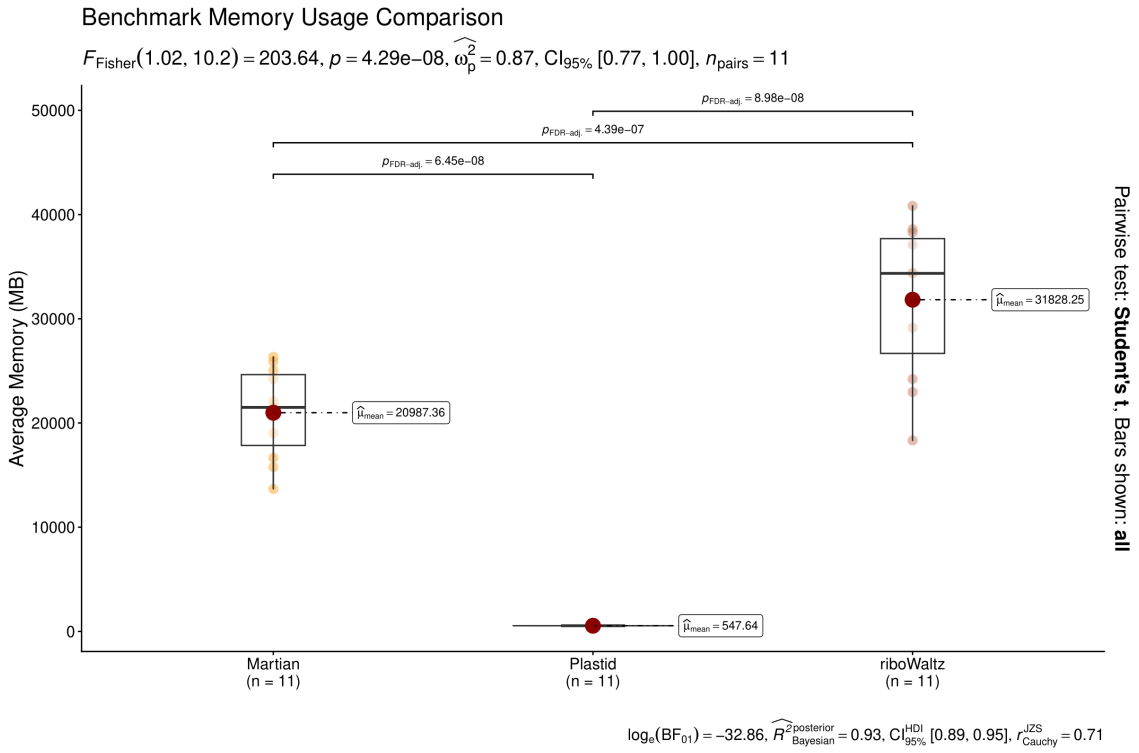

Figure 1.3: Memory consumption comparison across ribosome profiling tools. Box plots display memory usage distributions for 11 samples (n = 11 per tool). Plastid exhibited minimal memory requirements (547.64 MB), Martian showed intermediate usage (20,987.36 MB), and riboWaltz required the highest memory (31,828.25 MB). Statistical comparisons performed using one-way ANOVA with Student's t-test for pairwise comparisons (FDR-adjusted p-values shown).

1.3.3 Database Composition and Pre-processing Differences

The three tools operate on different transcript databases due to distinct preprocessing strategies. Martian's Database Creation module (Module 1) applies filtering criteria wherein transcripts lacking annotated start or stop codons are excluded from the reference database, and UTR boundaries overlapping with CDS coordinates are corrected to enforce mutually exclusive regions. This preprocessing resulted in a curated database containing approximately 100,000 transcripts with complete, well-defined coding sequences.

In contrast, riboWaltz and Plastid process the unfiltered reference transcriptome provided in the input GTF file, including transcripts with incomplete or ambiguous annotations. These tools operate on databases containing approximately 412,000 transcripts, approximately 4-fold larger than Martian's curated database.

This fundamental difference in database composition affects all downstream metrics, particularly detection percentages, which use different denominators across tools.

1.3.4 Analytical Accuracy and Quality Metrics

1.3.4.1 Translation Initiation Site (TIS) Detection

The Translation Initiation Site (TIS) score quantifies P-site offset estimation accuracy by measuring ribosome occupancy enrichment at annotated start codons. For each transcript, the TIS score were calculated as:

$$\text{TIS Score} = \frac{\sum_{i=0}^{15} N_i \cdot \mathbb{1}_{[i \bmod 3=0]}}{\sum_{j=-15}^{15} N_j \cdot \mathbb{1}_{[j \bmod 3=0]}} \tag{1.44}$$

where N_i represents the number of reads with P-sites mapping to position i relative to the annotated start codon (position 0), and $\mathbb{1}_{[i \bmod 3=0]}$ is an indicator function ensuring only in-frame reads are counted. The numerator captures reads positioned from the start codon to 15 nucleotides downstream (the expected initiation region), while the denominator includes reads in a 30-nucleotide window centered on the start codon. A TIS score approaching 1.0 indicates strong enrichment of ribosome occupancy immediately downstream of start codons in the correct reading frame, reflecting accurate P-site offset estimation. The sample-level TIS scores were calculated by aggregating counts across all transcripts in each tool's respective database (Figure 1.4). riboWaltz achieved a TIS score of 1.00, Martian demonstrated a score of 0.91, and Plastid showed a score of 0.14. Statistical comparisons confirmed highly significant differences (F_{Fisher}(1.13, 11.31) = 3706.24, p = 8.4810^{-16}, ω_p^2 = 1.00), with all pairwise tests yielding $p_{FDR-adj} < 10^{-9}$.

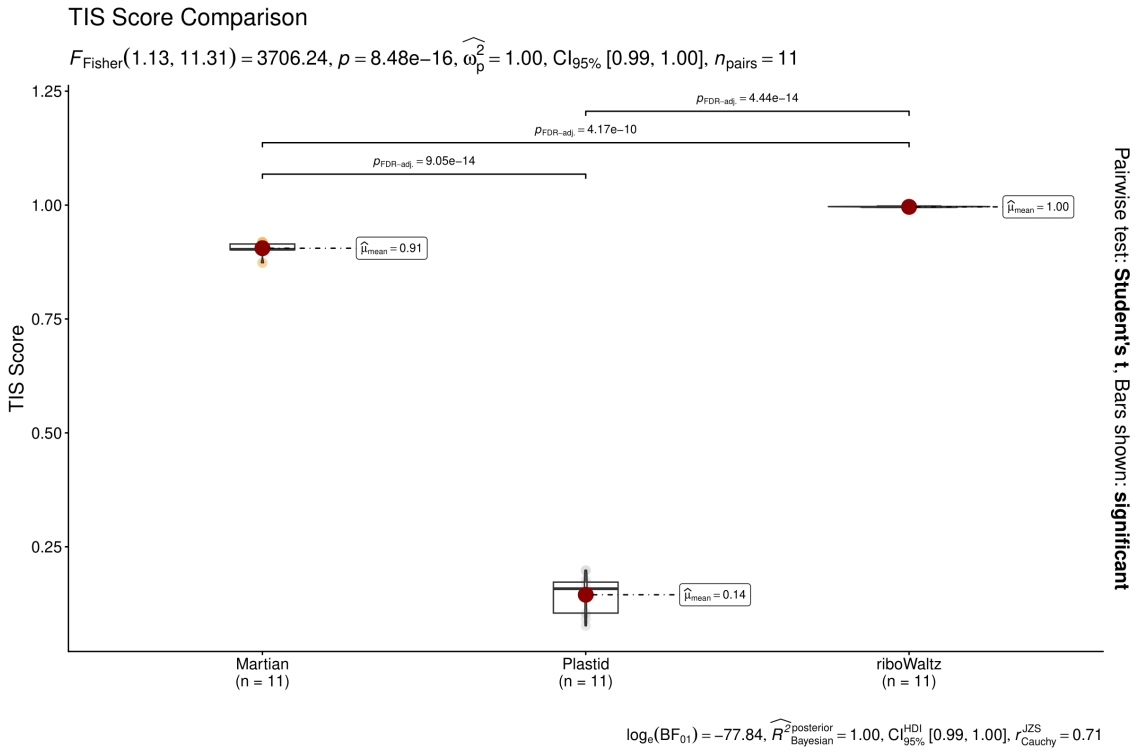

Figure 1.4: Translation Initiation Site (TIS) score comparison. Violin plots show TIS score distributions across 11 samples (n = 11 per tool), measuring P-site offset accuracy through ribosome occupancy enrichment at annotated start codons. riboWaltz achieved perfect scores (1.00), Martian demonstrated near-optimal performance (0.91), and Plastid showed substantially reduced accuracy (0.14). Statistical significance assessed using one-way ANOVA with Student's t-test (FDR-adjusted p-values shown).

1.3.4.2 Reading Frame Assignment

Analysis of reading frame distributions across the coding sequence (CDS) revealed differences in frame assignment among tools (Figure 1.5). For Frame 0 (the correct reading frame), Plastid showed the highest proportion at 63.1%, followed by Martian at 55.39% and riboWaltz at 53.92%. Kruskal-Wallis tests indicated significant differences in Frame 0 distribution ($\chi^2 KW$ = 14.82, p = 6.0510^{-4}, ε^2 = 0.43), with Plastid exhibiting significantly higher Frame 0 enrichment compared to both Martian ($p_{FDR-adj}$ = 0.003) and riboWaltz ($p_{FDR-adj}$ = 0.001). Frame 1 distributions showed no significant differences among tools ($\chi^2 KW$ = 3.39, p = 0.183, ε^2 = 0.05), with proportions of 21.53% (Martian), 17.96% (Plastid), and 22.24% (riboWaltz). Frame 2 distributions demonstrated significant variation ($\chi^2 KW$ = 15.15, p = 5.1410^{-4}, $p_{FDR-adj}$ = 0.44), with Plastid showing the lowest proportion at 18.94%, compared to Martian at 23.08% ($p_{FDR-adj}$ = 0.005) and riboWaltz at 23.84% ($p_{FDR-adj}$ = 6.9410^{-4}).

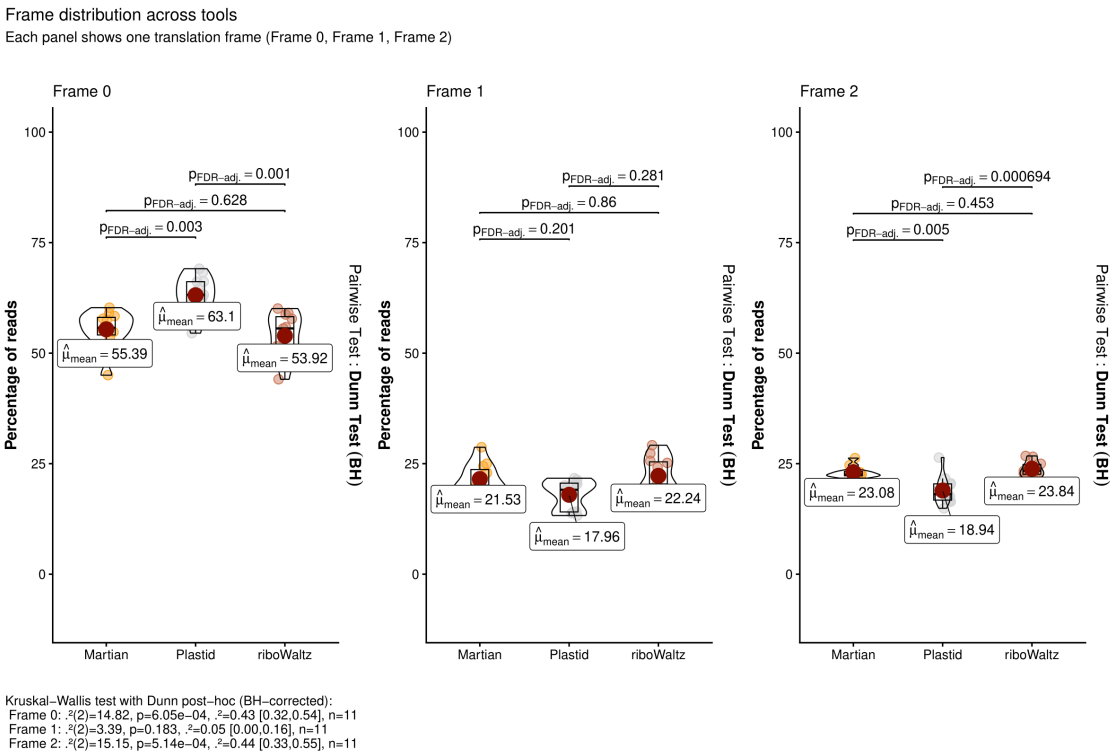

Figure 1.5: Reading frame distribution comparison across tools. Violin plots display the percentage of reads assigned to each reading frame (Frame 0, 1, and 2) within coding sequences across 11 samples (n = 11 per tool). Frame 0 represents the correct reading frame. Plastid showed highest Frame 0 enrichment (63.1%), followed by Martian (55.39%) and riboWaltz (53.92%). Statistical analysis performed using Kruskal-Wallis test with Dunn's post-hoc test for pairwise comparisons (FDR-adjusted p-values shown).

1.3.4.3 Region Assignment Patterns

Regional read distribution analysis revealed differences in transcript region classification among tools (Figure 1.6). For CDS assignment, Plastid classified 95.09% of reads as CDS-mapping, compared to 85.01% for riboWaltz and 84.68% for Martian. All pairwise comparisons for CDS assignment were highly significant ($\chi^2 KW$ = 25.84, p = 2.4510^{-6}, ε^2 = 0.79; all $p_{FDR-adj}$ < 0.05). For 5′UTR assignment, Plastid showed 4.66%, Martian 3.43%, and riboWaltz 3.70%, with significant differences detected ($\chi^2 KW$ = 26.04, p = 2.2210^{-6}, ε^2 = 0.80; all pairwise $p_{FDR-adj}$ < 0.05). The 3′UTR assignment showed the largest variation: Plastid identified 0.25% of reads as 3′UTR-mapping, while Martian identified 11.9% and riboWaltz identified 11.28% ($\chi^2 KW$ = 28.52, p = 6.4210^{-7}, ε^2 = 0.88). This represents an approximately 47-fold difference between Plastid and the other two tools. All pairwise comparisons for 3′UTR assignment were highly significant (all $p_{FDR-adj}$ < 0.01).

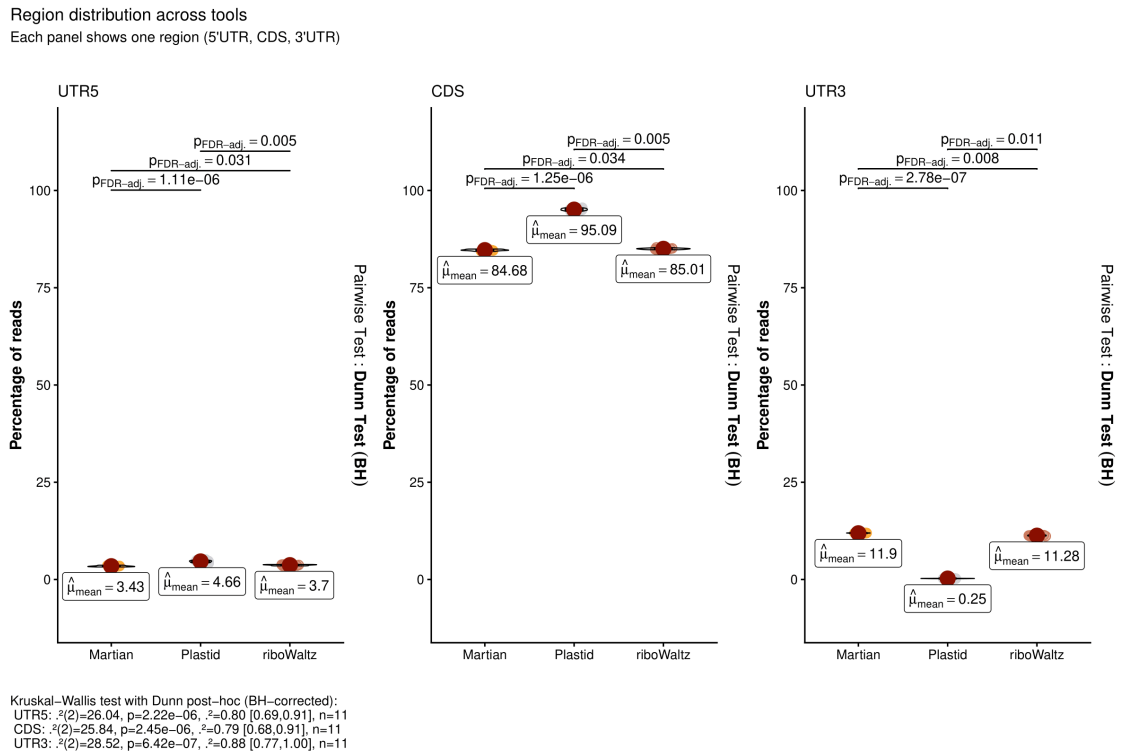

Figure 1.6: Regional read assignment patterns across tools. Violin plots show the distribution of reads assigned to 5'UTR, CDS, and 3'UTR regions across 11 samples (n = 11 per tool). Notable differences emerged in 3'UTR detection: Martian (11.9%), riboWaltz (11.28%), and Plastid (0.25%), representing a 47-fold difference. Statistical analysis performed using Kruskal-Wallis test with Dunn's post-hoc test (FDR-adjusted p-values shown).

1.3.4.4 Transcript Detection: Absolute Counts and Percentages

Analysis of transcript overlap revealed that the three tools detected different numbers of transcripts in absolute terms (Figure 1.7). Based on the union of all detected transcripts (109,421 total):

Absolute transcript counts:

- Martian: 72,789 transcripts (151 unique + 89 shared with Plastid only + 8,713 shared with riboWaltz only + 63,836 shared by all three)
- Plastid: 94,031 transcripts (1,412 unique + 89 shared with Martian only + 28,694 shared with riboWaltz only + 63,836 shared by all three)
- riboWaltz: 107,769 transcripts (6,526 unique + 8,713 shared with Martian only + 28,694 shared with Plastid only + 63,836 shared by all three)

In absolute terms, Martian detected 34,980 fewer transcripts than riboWaltz (32% fewer) and 21,242 fewer transcripts than Plastid (23% fewer). Detection percentages relative to each tool's database: When expressed as percentages of each tool's respective reference

database, the detection rates were: Martian 60%, Plastid 20%, and riboWaltz 23% (Figure 1.8). These percentage differences were highly significant ($F_{Fisher}(1.08, 10.8) = 51812.32$, p = 2.4410^{-21}, $\omega_p^2 = 1.00$; all pairwise $p_{FDR-adj} < 10^{-15}$). However, these percentages reflect different denominators due to database size differences. Martian's percentage (60%) is calculated from its curated database of 100,000 transcripts, while Plastid (20%) and riboWaltz (23%) percentages are calculated from databases of 412,000 transcripts. Therefore, direct comparison of these percentages requires consideration of the underlying database compositions. Overlap analysis: Of the 109,421 transcripts detected by at least one tool, 63,836 (58%) were detected by all three tools, representing a core set of commonly detected transcripts. Tool-unique detections were limited: Martian uniquely identified 151 transcripts (0.14% of detected transcripts), Plastid identified 1,412 unique transcripts (1.3%), and riboWaltz identified 6,526 unique transcripts (6.0%). The largest pairwise-only overlap occurred between Plastid and riboWaltz (28,694 transcripts, 26% of all detected), followed by Martian and riboWaltz (8,713 transcripts, 8.0%), with minimal overlap between Martian and Plastid only (89 transcripts, 0.08%).

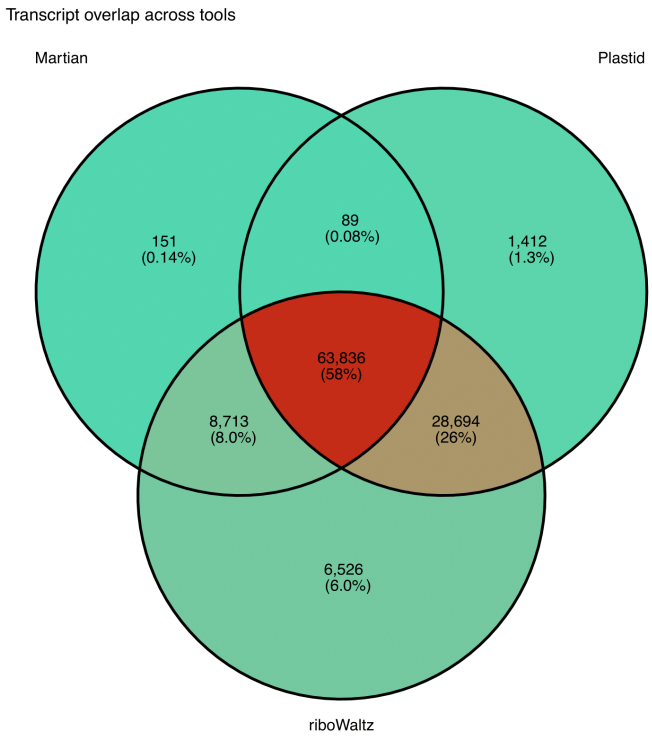

Figure 1.7: Transcript overlap analysis across tools. Venn diagram displays absolute transcript counts detected by each tool and their overlaps. A core set of 63,836 transcripts (58% of total detected) was identified by all three tools. Martian detected 72,789 transcripts, Plastid detected 94,031 transcripts, and riboWaltz detected 107,769 transcripts. Tool-unique detections were limited: Martian (151), Plastid (1,412), and riboWaltz (6,526).

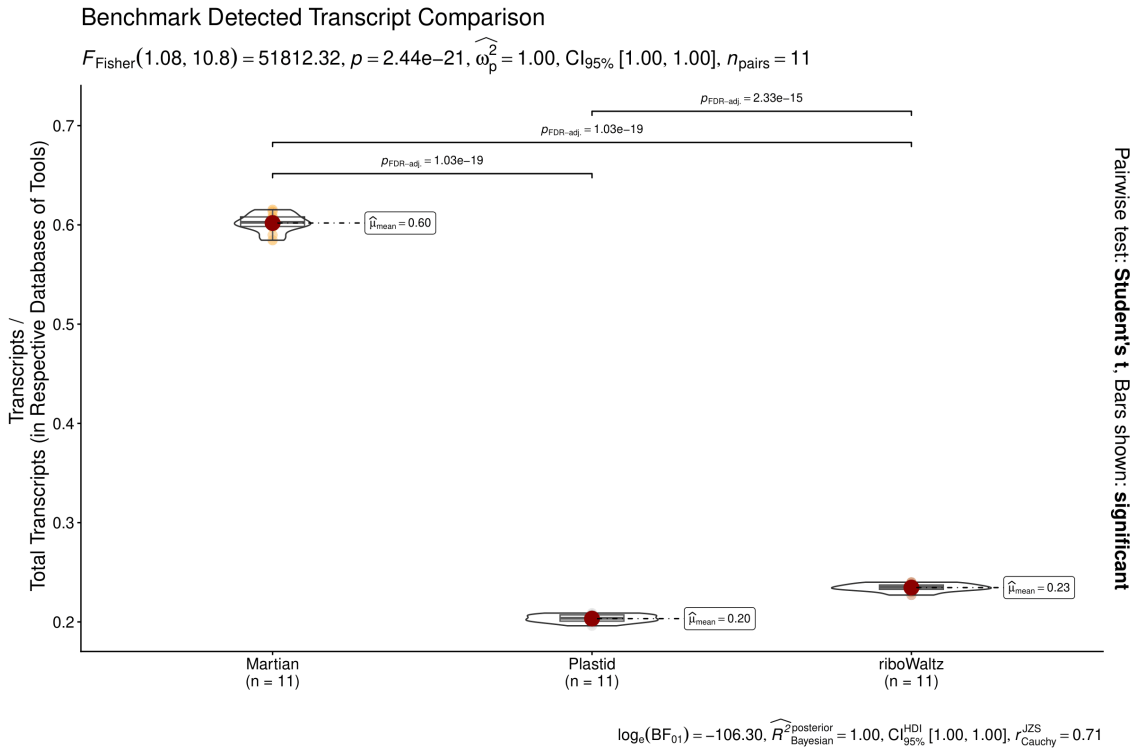

Figure 1.8: Transcript detection rates relative to reference databases. Violin plots show the percentage of transcripts detected with adequate coverage from each tool's respective reference database across 11 samples (n = 11 per tool). Detection rates were: Martian 60%, riboWaltz 23%, and Plastid 20%. These percentages reflect different denominators: Martian operates on a curated database (100,000 transcripts) while riboWaltz and Plastid use the unfiltered reference (412,000 transcripts). Statistical analysis performed using one-way ANOVA (FDR-adjusted p-values shown).

1.3.5 Functional Capabilities Comparison

1.3.5.1 Module Architecture and Analysis Features

Comparative analysis revealed differences in functional scope among tools (Table 1.1). Martian implements five modules: Database Creation, Ribosome Profiling, Condition Comparison, Translation Efficiency analysis, and nano-tRNAseq integration. riboWaltz provides P-site offset optimization and read quantification with codon-level resolution. Plastid offers P-site mapping with partial offset optimization capabilities. Key distinguishing features include Martian's implementation of annotation filtering and UTR boundary correction (Module 1), integrated statistical comparison between experimental conditions with DESeq2 and functional enrichment analysis (Module 3), translation efficiency quantification using interaction modeling to distinguish transcriptional from translational effects (Module 4), and integration of tRNA abundance and modification analysis from nanopore sequencing data (Module 5). All three tools implement P-site offset optimization, though with different algorithmic approaches.

1.3.5.2 Visualization and Output Formats

Martian generates interactive HTML reports using Bokeh-based visualizations with dynamic filtering and hover tooltips. Graphical outputs are exportable in SVG format. Tabular outputs are in CSV and TSV formats. riboWaltz produces static visualizations in PDF and PNG formats. Tabular outputs are in TSV and CSV formats. Generation of visualizations requires custom R scripting. Plastid generates static PNG graphics. Tabular outputs are in TXT format. Visualization generation requires command-line programming.

1.3.5.3 Accessibility and Usability

Installation complexity: Martian and Plastid show comparable installation procedures. riboWaltz requires more complex dependency management (Table 1.1).
Scripting requirements: Martian provides a complete workflow without requiring custom scripting for standard analyses. riboWaltz requires R programming for utilization, visualization and downstream analysis. Plastid requires command-line scripting for complete analytical workflows.
Built-in reporting: Martian generates automated HTML reports. riboWaltz and Plastid do not provide integrated reporting functionality.

1.3.6 Summary of Benchmarking Results

Table 1.1: Comparison of ribosome profiling analysis tools

Feature/Metric	Martian	riboWaltz	Plastid
Accessibility and Usability			
Need of scripting	×	✓	✓
Installation ease	Easy	Moderate	Easy
Built in reports	✓	×	×
Functional Capabilities			
Annotation correction	✓	×	×
Translation efficiency analysis	✓	×	×
Sample-wise comparison	✓	~	×
Condition-wise comparison	✓	×	×
Compatible analyses	✓(nanotRNASeq)	×	×
Ribosome Profiling			
P-site offset optimization	✓	✓	~
Coverage calculation	✓	✓	×
Coverage width calculation	✓	×	×
Codon usage calculation	✓	✓	×
Monosome length detection	✓	×	×
Visualization Capabilities			
Interactive plots	✓	×	×
Graphical output format	html/svg	pdf/png	png
Output and Analysis			
Post-process ready table outputs	✓	~	~
Table output format	csv/tsv	tsv/csv	txt
Technical Specifications			
Processing time (min)*	36	51	38
Memory usage (GB)	20	31	0.5
Adaptive memory usage**	✓	×	–

* Processing time for average 15M reads on standard hardware
** Ability to adapt memory usage with respect to hardware
✓= Available/Supported, ×= Not available,
–= No information found, ~= Partially Supported

Across 11 ribosome profiling datasets, the following performance characteristics were observed:

Computational metrics:

Runtime: Martian 1,896 seconds (32 min), Plastid 2,128 seconds (35 min), riboWaltz 3,029 seconds (51 min)

Memory usage (mean): Martian 21 GB, Plastid 0.5 GB, riboWaltz 32 GB

Accuracy metrics:

TIS scores: riboWaltz 1.00, Martian 0.91, Plastid 0.14
Frame 0 percentage: Plastid 63.1%, Martian 55.39%, riboWaltz 53.92%
3'UTR detection: Martian 11.9%, riboWaltz 11.28%, Plastid 0.25%
Detection metrics:
Absolute transcript counts: riboWaltz 107,769, Plastid 94,031, Martian 72,789
Detection percentage (relative to each database): Martian 60%, riboWaltz 23%, Plastid 20%
Common transcripts (detected by all three): 63,836 (58% of union)
Functional capabilities:
Translation efficiency analysis: Martian only
tRNA integration: Martian only
Annotation correction: Martian only
Interactive visualization: Martian only
Scripting requirement: None (Martian), Required (riboWaltz, Plastid)

These results indicate that tool selection involves trade-offs between computational resource requirements, analytical scope, detection breadth, and user accessibility.

1.4 Discussion

The benchmarking of Martian against riboWaltz and Plastid revealed that these tools represent fundamentally different design philosophies, each optimizing for different priorities. Martian achieved the fastest runtime (1,896 seconds) while maintaining near-optimal P-site accuracy (TIS score 0.91) but required substantial memory (21 GB) and detected fewer transcripts in absolute terms (72,789) due to stringent annotation filtering. riboWaltz demonstrated optimal P-site inference (TIS score 1.00) and broadest transcript detection (107,769) but required the most memory (32 GB) and longest runtime (3,029 seconds). Plastid minimized memory requirements (0.5 GB) but showed substantially reduced P-site accuracy (TIS score 0.14) and detected the fewest $3'$UTR reads (0.25%). These differences are not simply technical details; they reflect fundamental trade-offs in how ribosome profiling analysis tools balance analytical accuracy, computational efficiency, data quality, and detection breadth. Understanding these trade-offs is essential for selecting appropriate tools and interpreting published ribosome profiling studies.

The substantial differences in TIS scores—riboWaltz (1.00), Martian (0.91), Plastid (0.14)—reflect different algorithmic strategies for P-site offset calculation. As described in the Introduction, accurate P-site positioning is critical for identifying translation initiation sites, quantifying ribosome occupancy at specific codons, and analyzing triplet periodicity. The riboWaltz algorithm achieves approximately 85% correct frame assignment

through coherence-based optimization, compared to 65% with earlier methods. Martian's implementation employs a mode-based approach with smoothing and imputation for read lengths with sparse coverage. This design choice prioritizes consistent performance across varying data quality over maximal optimization for individual datasets. The 9% TIS score difference between Martian and riboWaltz represents this trade-off: accepting slightly reduced optimization to maintain stability when read length distributions are irregular or sequencing depth is limited.

The rationale for this design reflects practical realities in ribosome profiling experiments. While high-quality datasets with deep coverage and narrow read length distributions are ideal, many experiments—particularly those involving clinical samples, rare cell populations, or exploratory conditions—yield more variable data. Recent methodological advances have enabled ribosome profiling from FACS-sorted cells with as little as 0.1 pmol of RNA fragments, but such low-input protocols may produce irregular read distributions where aggressive optimization becomes unstable. Plastid's substantially lower TIS score (0.14) indicates reduced enrichment at translation initiation sites compared to the other tools. The cause of this difference was not investigated in this study, but the practical consequence is clear: Plastid's P-site assignments show weaker correspondence with annotated start codons, which may affect analyses dependent on accurate initiation site mapping.

The TIS score differences have direct implications for research questions focused on translation initiation. Studies investigating ribosome recruitment mechanisms, alternative start codon usage, or upstream open reading frame (uORF) regulation require accurate identification of initiation events. As noted in the Introduction, translation deregulation through mechanisms including altered ribosome recruitment affects the majority of human tumors and underlies multiple neurological disorders. Tools with higher TIS scores provide more reliable data for such investigations. However, for studies focused on elongation dynamics, codon-specific pausing, or overall transcript-level ribosome occupancy, the TIS score may be less critical than other performance characteristics such as accurate reading frame assignment or computational efficiency.

Beyond P-site accuracy, the most striking difference in regional assignment was the 47-fold variation in 3′UTR detection: Martian (11.9%) and riboWaltz (11.28%) versus Plastid (0.25%). This difference has important implications given the growing recognition of 3′UTR translation regulation in cellular processes. Martian's implementation of UTR boundary correction in Module 1 addresses a common problem in reference transcriptomes: overlapping or ambiguous coordinate annotations between UTR and CDS regions. When GTF files contain such inconsistencies, reads mapping near UTR-CDS boundaries may be misclassified depending on how tools handle overlapping annotations. By system-

atically correcting these boundaries to enforce mutually exclusive regions, Martian ensures that reads assigned to 3′UTR truly map outside the CDS.

Both riboWaltz and Plastid use the GenomicFeatures package or direct GTF parsing without additional coordinate correction steps, processing annotations as provided in the input file. Whether the 47-fold difference between Plastid and the other tools stems solely from this annotation handling difference, from P-site offset variations, or from other algorithmic factors was not determined in this study. However, the practical consequence is clear: studies using Plastid will detect substantially fewer 3′UTR-mapping reads, which could affect interpretation of results related to 3′UTR translation events. The biological significance of 3′UTR translation detection extends beyond simple regional quantification. Researches has demonstrated roles for 3'UTR translation in stop codon readthrough, regulatory uORFs positioned in 3′UTRs, and IRES-mediated internal initiation [44, 98]. While the magnitude of 3′UTR translation in normal physiology remains debated, tools that systematically underdetect such events will miss or underestimate their contribution when they do occur.

The differences in regional assignment connect directly to perhaps the most consequential design decision in Martian's architecture: the stringent transcript filtering in Module 1. Transcripts lacking annotated start or stop codons are excluded, resulting in a curated database of ~ 100,000 transcripts compared to ~ 412,000 in unfiltered references. This decision reflects prioritization of analytical robustness over comprehensive coverage. The rationale for this approach is straightforward: transcripts lacking complete CDS annotations create multiple downstream problems. Codon usage calculation requires defined start-to-stop reading frames; without clear boundaries, codon identification becomes ambiguous. Translation efficiency quantification depends on CDS-normalized metrics; incomplete annotations prevent reliable normalization. Regional assignment to 5′UTR, CDS, and 3′UTR requires definitive coordinate boundaries; overlapping or missing annotations cause misclassification. By filtering at the database creation stage, all downstream analyses operate on well-defined features with unambiguous interpretations. This approach eliminates a class of potential errors before they can propagate through the analysis pipeline.

The direct consequence of stringent filtering is reduced transcript detection in absolute terms. Martian detected 34,980 fewer transcripts than riboWaltz (32% fewer) and 21,242 fewer transcripts than Plastid (23% fewer). This represents a substantial reduction in detected RNA complexity that has to be taken into account. The ~312,000 transcripts excluded from Martian's curated database include multiple categories. Some represent genuine protein-coding genes with incomplete or ambiguous annotations in the reference GTF. Others represent non-coding transcripts, pseudogenes, or assembly artifacts with nominal "CDS" annotations. Still others represent alternative isoforms with uncertain structures.

Distinguishing legitimate biology from annotation artifacts within this excluded set was beyond the scope of the current work.

The 63,836 transcripts detected by all three tools represent 88% of Martian's total detections but only 59% of riboWaltz's detections. This indicates that Martian's curated approach captures the core set of well-annotated, confidently detected transcripts, but riboWaltz detects an additional 43,933 transcripts beyond this core set. Whether these additional transcripts represent genuine translation events that Martian misses due to overly conservative filtering, or whether they include substantial numbers of false positives from poorly annotated regions, would require case-by-case validation beyond the scope of this benchmarking.

This trade-off between quality and comprehensiveness occurs within a broader context of reproducibility challenges in ribosome profiling analysis. As noted in the Introduction, comprehensive benchmarking studies reveal that ribosome profiling tools show only 40% peak detection reproducibility between replicates, and major open reading frame detection tools show only 2% agreement for small ORFs. These discrepancies highlight fundamental challenges in the field. Martian's design choice to prioritize annotation quality represents one approach to improving reproducibility: reducing false positives from ambiguous annotations at the cost of potentially missing some true positives from poorly annotated transcripts. Alternative approaches—such as riboWaltz and Plastid's inclusion of all reference transcripts—prioritize comprehensive coverage at the risk of including more annotation-driven artifacts. Neither approach is categorically superior; they represent different prioritizations within an inherent trade-off.

The analytical trade-offs discussed above must also be considered alongside computational resource requirements, which have direct practical implications for tool accessibility. The memory consumption differencesk, Plastid (0.5 GB), Martian (21 GB), riboWaltz (32 GB), determine which computational environments can support each tool. Plastid's minimal memory footprint enables deployment on standard laptop computers, while Martian requires workstations with ≥32 GB RAM and riboWaltz requires ≥64 GB RAM. However, these runtime memory requirements do not capture the complete computational picture. During this benchmarking study, Plastid's database creation step required workstation-level resources and could not complete with 16 GB RAM, despite Plastid's minimal memory requirements during the analysis phase itself. In contrast, both Martian and riboWaltz database creation completed successfully on the same 16 GB system. These database creation requirements represent an often-overlooked practical consideration that affects initial tool setup even when runtime requirements appear favorable.

The memory allocation in Martian's implementation supports several resource-intensive capabilities: in-memory transcript sequence indexing for rapid codon identification, complete coverage arrays for nucleotide-resolution analysis, and simultaneous processing of

multiple read length cohorts during P-site optimization. While adaptive memory management partially mitigates these requirements by adjusting batch sizes based on available system resources, the fundamental architectural choices that enable Martian's analytical capabilities require substantial memory allocation.

Martian's fastest runtime (1,896 seconds) represents a 12% improvement over Plastid and a 37% improvement over riboWaltz. For single-sample analyses, these differences (4-19 minutes) may have limited practical significance. However, for large-scale studies—increasingly common as ribosome profiling databases like RPFdb v3.0 now contain 5,018 datasets and RiboSeq.Org includes 14,840 preprocessed samples—these per-sample differences scale substantially. More significantly, runtime efficiency affects iterative analyses during methods development, parameter optimization, and quality control troubleshooting. Researchers performing multiple analysis rounds to refine approaches or diagnose technical issues benefit disproportionately from reduced processing times. The combination of fastest runtime and intermediate memory positions Martian favorably for production analyses on standard computational infrastructure.

Beyond computational performance and analytical accuracy, Martian's integrated capabilities address a broader challenge in the field. As described in the Introduction, current ribosome profiling analysis suffers from fragmentation across numerous specialized tools, creating analytical bottlenecks and reproducibility challenges. Researchers must manually orchestrate riboWaltz for P-site optimization, Xtail or RiboDiff for differential translation, various tRNA sequencing pipelines, and separate statistical frameworks for multi-omics integration. This fragmentation creates numerous failure points, inconsistent normalization approaches, and suboptimal resource utilization. Martian's implementation of five integrated modules directly addresses this fragmentation. Module 3 provides native condition-wise comparison with integrated DESeq2, eliminating manual data export and custom statistical scripting. Module 4 implements translation efficiency analysis through interaction modeling, a capability absent in both riboWaltz and Plastid that would otherwise require substantial custom development. Module 5 provides the first integrated framework for joint ribosome profiling and tRNA sequencing analysis, enabling investigation of coordinated regulation across these molecular layers.

The practical value of this integration extends beyond time convenience. Integrated workflows ensure consistent data handling, reduce opportunities for errors during manual tool orchestration, and enable analyses that would be technically challenging to implement through tool combinations.

The elimination of scripting requirements in Martian represents a conscious design choice to broaden accessibility. Both riboWaltz and Plastid require programming skills, R for riboWaltz, command-line scripting for Plastid, for complete analyses. While this

provides flexibility for computational specialists, it creates barriers for experimental biologists, and researchers without programming training. This accessibility consideration has practical implications for technology translation. Broader adoption of ribosome profiling in translational research and pharmaceutical contexts requires tools accessible to domain experts who may not possess extensive computational training. Martian's automated workflows with parameter files rather than custom scripts represent one approach to reducing these barriers.

The interpretation of these benchmarking results must be tempered by acknowledgment of the study's scope and limitations. This benchmarking evaluated performance across 11 samples from a single study using human cell lines. Performance characteristics on samples from different organisms, different library preparation protocols, sequencing platforms, or biological conditions were not evaluated. Conclusions about relative tool performance are limited to samples. Several potentially relevant performance metrics were not quantified: reproducibility across biological replicates, sensitivity and specificity for detecting differential translation, performance across varying sequencing depths, accuracy of translation efficiency quantification, tRNA modification detection accuracy, and computational scalability with increasing sample numbers. Future benchmarking addressing these dimensions would provide more comprehensive tool characterization.

Additionally, the benchmarking documented performance differences but did not systematically investigate their underlying causes. For example, why Plastid's TIS score is substantially lower, what specific algorithmic choices drive memory differences, and whether tool-unique transcript detections represent signal or noise was not investigated in this work. Several ribosome profiling analysis tools described in the Introduction were not included in this comparison, such as RiboSeq.Org, riboviz 2, and RiboVIEW. The comparison focused on riboWaltz and Plastid as representative established tools, but other tools may offer different performance trade-offs.

Despite these limitations, the benchmarking results provide a guidance for tool selection which should align measured performance characteristics with specific research requirements. Researchers requiring minimal memory usage must choose Plastid (0.5 GB), as it is the only tool deployable on standard laptop computers, though they should be aware of reduced 3'UTR detection (0.25%) and lower TIS scores (0.14) relative to alternatives. Researchers prioritizing P-site accuracy for translation initiation studies should consider riboWaltz (TIS 1.00), accepting higher memory requirements (32 GB) and longer runtime (3,029 seconds), though riboWaltz lacks integrated translation efficiency and tRNA analysis capabilities. Researchers requiring translation efficiency analysis or tRNA integration should go with Martian, as these capabilities are not implemented in other evaluated tools, but at the same time they should ensure ≥32 GB RAM availability and understand that absolute transcript detection will be reduced due to annotation filtering. Researchers pri-

oritizing comprehensive transcript coverage including poorly annotated transcripts should use riboWaltz (107,769 detected) or Plastid (94,031 detected), accepting potential inclusion of annotation artifacts alongside genuine biology. Researchers without programming expertise should consider Martian for its automated workflows, while those with strong computational skills may prefer the flexibility of riboWaltz (R-based) or Plastid (command-line) approaches.

This benchmarking provides the quantitative foundation for evidence-based tool selection, enabling researchers to make informed decisions that balance analytical requirements against practical constraints.

Moreover, this work highlights broader challenges facing the ribosome profiling field. The 47-fold variation in 3′UTR detection, 40% peak detection reproducibility between replicates, and 2% agreement among ORF detection tools documented here and in previous studies reveal fundamental inconsistencies that undermine cross-study comparisons and meta-analyses. These discrepancies stem not from technical failures but from legitimate algorithmic differences reflecting different biological assumptions and analytical priorities. Addressing these challenges will require community-wide efforts to establish standardized benchmarking datasets, define minimal quality metrics, and develop guidelines for transparent reporting of analytical choices.

The development of Martian represents a contribution toward addressing the fragmentation, accessibility barriers, and reproducibility challenges currently limiting translation research. By integrating database curation, P-site inference, statistical comparison, translation efficiency quantification, and tRNA analysis within a unified framework accessible to users without programming expertise, Martian reduces analytical bottlenecks while maintaining scientific rigor. Continued tool development, comprehensive benchmarking across diverse experimental conditions, and community coordination on analytical standards will be essential to fully realize ribosome profiling's potential as both a research tool and a clinical diagnostic platform.

Chapter 2

Profiling Translationally Active Transfer RNAs and Stress-Induced Dynamics

2.1 Introduction

Transfer RNA molecules serve as critical adaptors that decode mRNA sequences into amino acid chains, with their abundance, modification status, and charging dynamics collectively determining translational efficiency and fidelity. While recent methodological advances have enabled comprehensive profiling of total cellular tRNA populations, a fundamental question remains: do total tRNA pools accurately reflect the tRNA landscape engaged in active translation? Previous studies have revealed discrepancies between total tRNA populations and ribosome-associated tRNA pools purified from polysomal fractions, particularly under stress conditions where selective tRNA recruitment to ribosomes may drive adaptive translation responses. However, the isolation of ribosome-associated tRNAs has traditionally required labor-intensive sucrose gradient fractionation with substantial input material, limiting the application of this approach across diverse experimental conditions and preventing systematic comparisons between total and ribosome-engaged tRNA populations.

Short-read tRNA sequencing methods have provided much of our current knowledge of tRNA biology, but they carry intrinsic limitations that motivate the nanopore-based approach used in tRIBO-seq. Methods such as DM-tRNA-seq, mim-tRNAseq, ARM-seq, and YAMAT-seq rely on reverse transcription followed by PCR amplification and short-read sequencing [43, 46, 48, 75]. Because most tRNA modifications interfere with processive reverse transcription, these methods require chemical pre-treatments (e.g., AlkB

demethylation, periodate oxidation) or specialised reverse transcriptases (e.g., TGIRT) to read through modified positions.

Consequently, the sequencing signal reflects modification-induced misincorporation rates rather than direct physical evidence of the modification, making it difficult to simultaneously assess multiple modification types or to distinguish primary from secondary effects of chemical treatments. Furthermore, PCR amplification introduces sequence-dependent biases that disproportionately affect low-abundance tRNAs and tRNA families with high sequence similarity. Critically, none of these methods selectively capture ribosome-associated tRNAs; profiling the translationally active tRNAome requires a prior sucrose gradient fractionation step, which is laborious, requires substantial input material (typically tens to hundreds of micrograms of total RNA), and introduces batch effects when paired with independent tRNA sequencing experiments.

This chapter describes my contributions to the collaborative development of tRIBO-seq, a novel nanopore-based high-throughput method that simultaneously captures tRNA abundance, modification patterns, and fragmentation status from ribosome-associated populations. This work was conducted in collaboration with the laboratory of Dr. Eva Maria Novoa and IMMAGINA Biotechnology s.r.l., and is currently under revision at *Nature Methods*. My specific contributions included the majority of bioinformatics analyses, development of computational pipelines for tRNA modification detection and quantification, and complete processing and analysis of arsenite-treated samples including both ribosome profiling and nano-tRNA sequencing data. The tRIBO-seq methodology combines RiboLace-based active ribosome isolation with nanopore direct RNA sequencing, eliminating the need for sucrose gradient ultracentrifugation while requiring minimal input material and enabling completion within a single working day.

tRIBO-seq applied to characterize tRNAome dynamics under three distint stress paradigms: amino acid deprivation (arginine and leucine starvation), methionine limitation, and oxidative stress induced by arsenite exposure. These conditions were selected to probe different aspects of tRNA-mediated translational control—nutrient sensing through tRNA abundance changes, metabolic coupling through modification dynamics, and stress-induced tRNA fragmentation. The analyses revealed condition-specific tRNAome reprogramming patterns, with amino acid deprivation primarily affecting tRNA isoacceptor abundances, methionine starvation inducing widespread hypomethylation, and arsenite exposure triggering selective fragmentation of ribosome-embedded tRNAs. Critically, comparison between ribosome-associated and total tRNA populations demonstrated that these pools respond differently to stress, with ribosome-associated tRNAs exhibiting distinct enrichment patterns and modification profiles that more accurately reflect translational status.

The remainder of this chapter details the tRIBO-seq methodology, presents comprehensive analyses of tRNA dynamics across the three stress conditions, and discusses the biological implications of differential tRNA recruitment to ribosomes. I conclude by highlighting how the integration of ribosome profiling with ribosome-associated tRNA sequencing provides unprecedented insight into the coordination between codon usage, tRNA availability, and translation dynamics under cellular stress.

2.2 Materials and Methods

2.2.1 Cell Culture and Chemical Treatments

HEK293T (ATCC) and MCF-7 (Michigan Cancer Foundation-7, ATCC) cells were maintained at 37°C in a humidified incubator with 5% CO_2. Unless specified otherwise, HEK293T cells were maintained in Dulbecco's Modified Eagle Medium (DMEM, Sigma-Aldrich, cat. no: D6429) supplemented with 10% FBS (Thermo Fisher, cat. no: 10082147). MCF-7 cells were cultured in DMEM (Thermo Fisher, cat. no: 10566016) supplemented with 10% heat-inactivated FBS (Thermo Fisher, cat. no: 10082147), 1% (v/v) penicillin-streptomycin (Thermo Fisher, cat. no: 15140122), and 1 mM L-glutamine (Thermo Fisher, cat. no: 25030081). For routine passaging, cells were washed with 1X phosphate-buffered saline (PBS), detached using 0.25% trypsin-EDTA, and centrifuged at 300×g for 5 minutes. Cells were seeded at appropriate densities depending on experimental requirements.

For chemical treatments, MCF-7 cells (7×10 cells per 10-cm dish) were treated with 1 mM sodium arsenite (Asn) for 50 minutes or 2 µg/mL Harringtonine (HAR) for 3 minutes, followed by 10 µg/mL cycloheximide (CHX) for 5 minutes. In all cases, DMSO was used as a vehicle control at an equivalent volume. All cell lines were routinely screened for mycoplasma contamination using Eurofins Scientific services, and only mycoplasma-negative cells were used for experiments.

2.2.2 Arginine and Leucine Starvation

HEK293T cells were initially maintained in standard DMEM (Sigma-Aldrich, cat. no: D6429) supplemented with 10% fetal bovine serum (FBS) and incubated at 37°C in a humidified atmosphere containing 5% CO_2. For deprivation, cells were washed with pre-warmed PBS and transferred to custom arginine-free or leucine-free media for 3 hours. Deprivation media were prepared by supplementing amino acid-free DMEM (Genaxxon Bioscience, cat. no: C4150) with 3.5 g/L glucose and the following amino acids (mg/L): glycine (30), cysteine (24), glutamine (580), histidine (31), isoleucine (105), lysine (131),

methionine (30), phenylalanine (66), serine (42), threonine (95), tryptophan (16), tyrosine (71), and valine (94). Arginine-free medium was supplemented with leucine (105 mg/L), and leucine-free medium was supplemented with arginine (69 mg/L). Control medium was prepared using the same base formulation with either L-arginine (69 mg/L) or leucine (105 mg/L). All media were pH-adjusted to 7.2–7.4, sterile-filtered, and supplemented with 10% dialyzed FBS (Gibco, cat. no: A3382001) and 1% (v/v) penicillin–streptomycin (Thermo Fisher, cat. no: 15140122). Media preparation and handling were performed under sterile conditions.

2.2.3 Methionine Starvation

For methionine deprivation, MCF-7 cells at approximately 70% confluency were washed with pre-warmed PBS and incubated in methionine-free medium (MFM) for 6–16 hours. MFM was prepared using methionine-free DMEM (Thermo Fisher, cat. no: 21013024), supplemented with 10% dialyzed FBS (Gibco, cat. no: A3382001) and 1% (v/v) penicillin–streptomycin (Thermo Fisher, cat. no: 15140122). Control medium was generated by supplementing MFM with L-methionine to a final concentration of 0.2 mM (Thermo Fisher, cat. no: J6190418). All experimental and control conditions were prepared in parallel and maintained under identical incubation parameters.

2.2.4 tRIBO-seq: Ribo-Embedded tRNA Isolation

Ribosomes were isolated using the RiboLace kit (Immagina Biotechnology S.r.l., cat. no: GF001-12), following manufacturer's recommendations. Briefly, approximately 7×10 MCF-7 cells seeded in 10-cm dish plates were allowed to reach 70-80% confluency before treatment (see 'Cell Culture and Treatment' for dosage details). Prior to harvesting, cells were incubated with 10 µg/mL cycloheximide (CHX, SIGMA cat. no. 01810) for 5 minutes at 37°C. Cells were then washed twice with ice-cold PBS containing 20 µg/mL CHX and resuspended in 300 µL Lysis Buffer (LB; Immagina #1BT0031) supplemented with sodium deoxycholate 10% (v/v), DNase I (1 U/µL, Thermo Scientific catalog no. 89836) and RiboLock RNase Inhibitor (40 U/µL, Thermo Scientific catalog no. EO0381).

Cells were pre-treated with cycloheximide (CHX) to preserve ribosome-mRNA-tRNA interactions during isolation, which stabilizes translating ribosomes by binding the E-site [99]. Absorbance of the cell lysate was measured at 260 nm using a Nanodrop ND-1000 UV-VIS Spectrophotometer, and 0.9 AU was used for ribosome pull-down, adjusted to a final volume of 450 µL with Wash Buffer (WB; Immagina #IBT0071). To each sample, 0.9 µL of Nux Enhancer (NE; Immagina #IBT0081) and 4.5 µL diluted Nuclease solution

(equivalent to 6.75U, dNux; Immagina #IBT0081) were added, and samples were incubated at 25°C for 45 minutes to allow digestion. Digestion was terminated by adding 1.5 µL SUPERaseIn (Invitrogen, catalog no. AM2696) and incubating on ice for 10 minutes.

While digestion occurred, ribosome pull-down beads were prepared by functionalizing RiboLace magnetic beads (RmB; Immagina #IBT0042) with the RiboLace smart probe (RsP; Immagina #IBT0011). The probe was allowed to bind to the beads for 1 hour at room temperature with shaking at 1400 RPM. After incubation, mPEG (Immagina #IBT0061) was added to block non-specific binding, and beads were washed twice with WB. After stopping the digestion the digested cell lysate was added to the functionalized beads and incubated for 70 minutes on a rotating wheel (3-10 RPM) at 4°C. Following incubation, beads were washed twice with 1 mL WB. RNA was then eluted from ribosomes using the RNA Clean & Concentrator™-5 kit (Zymo catalog. No. R1015 or R1016) according to manufacturer instructions, with a modification: 200 µL of Zymo RNA Binding Buffer was added directly to the beads, followed by thorough mixing. Recovered RNA was quantified at 260 nm using a Nanodrop ND-1000 UV-VIS Spectrophotometer, and RNA integrity and size were assessed via 15% TBE-UREA gels and/or Agilent 4200 TapeStation RNA HS ScreenTape Assay (cat. no: 5067-5579).

2.2.5 tRIBO-seq: Library Preparation

For building ribo-embedded nano-tRNA-seq libraries, the protocol of Lucas et al. [54] was followed, with minor adaptations specific to ribo-embedded tRNA sequencing. Briefly, since CHX treatment retains deacylated tRNAs at the peptidyl (P) site, as described by Schneider-Poetsch et al. [99], the deacylation step was omitted for these samples. However, as harringtonine (HAR) blocks translation before peptide bond formation between P-site and aminoacyl (A) site tRNAs [99], HAR-treated libraries were deacylated to enable adaptor ligation for sequencing of iMet tRNAs. Nano-tRNAseq libraries from total RNA were prepared following the protocol by Lucas et al. [54]. All libraries were sequenced using the Oxford Nanopore Direct RNA Sequencing (DRS) protocol with the DRS kit (ONT, SQK-RNA002).

2.2.6 tRIBO-seq: Data Analysis

2.2.6.1 Basecalling and Mapping

Basecalling was performed using Guppy (v 6.5.7), as described by Lucas et al. [54]. Basecalled reads were aligned to the reference genome using minimap2 (v2.24) [100] with the parameters: -ax splice -k 7 -w 3 -n 1 -m 13 -s 30 -A 2 -B 1 -O 1,32 -E 1,0. Demultiplexing

was conducted using SeqTagger [101]. Reads were subsequently separated based on their alignment positions to annotated reference tRNAs.

2.2.6.2 Filtering

For downstream analyses, reads lacking the 3′ splint adapter were discarded, as those most likely correspond to mismapped reads. Full-length reads were defined by their start positions relative to the reference tRNAs. Reads containing full or partial 5′ splint adapters were kept for downstream analyses and were labeled as ”full-length tRNAs” and ”fragmented tRNAs,” respectively. All reads were further filtered based on mapping quality (MapQ ≥ 10). All filtering steps were carried out using pysam (v0.21.0) within custom Python scripts.

2.2.6.3 Differential tRNA Abundance Analysis

Differential abundance analysis was conducted using DESeq2 (v1.44.0) within a custom R script. tRNAs with fewer than 10 counts across all samples were excluded. Size-factor normalization from the DESeq2 [102] package was applied prior to differential testing. Comparisons between groups (Control vs. Treatment) within each experimental condition (Ribo-embedded vs. Total) were modeled using a one-factor design matrix: ∼ Group. P-values were adjusted for multiple testing using the Benjamini-Hochberg procedure [91]. This analysis was performed separately for each of the three datasets: ”All,” ”Full-Length,” and ”Fragment.”

2.2.6.4 Differential Modification Analysis

Differential modification analysis was based on the detection of basecalling errors, as previously described by Begik et al. [96]. Only the ”Full-Length” subsample was used in this analysis to minimize false positives due to positional coverage variability. Positions with fewer than 10 total counts across all samples were excluded. For each reference tRNA, positions matching the expected nucleotide at each position were labeled as ”not-modified,” while mismatches, insertions, or deletions were classified as ”modified” counts. This was implemented via a custom R script. Size-factor normalization from the DESeq2 [102] package was applied prior to differential modification testing. Differential modification analysis was performed using DESeq2 (v1.44.0), applying a two-factor design: ∼ Condition * Type, where ’Condition’ refers to control versus treated and ’Type’ refers to modified versus not-modified. P-values were corrected using the Benjamini-Hochberg method. Log_2 fold changes were visualized using ComplexHeatmap [103] (v2.22.0). Final heatmaps were generated filtering out positions with adjusted P-values > 0.05.

2.2.7 Ribo-seq Library Preparation

Library preparation of eluted ribosome-protected fragments (RPFs) was performed using the RiboLace Gel Free kit (Immagina Biotechnology S.r.l, cat. no: #GF001-12). Briefly, 1.5 µg of RNA was 5' phosphorylated by incubation with Buffer L1 (Immagina #IBT0151), ATP (Immagina #IBT0171), and enzyme L1 (Immagina #IBT0161) in a final volume of 50 µL for 1 hour at 37°C. The reaction was cleaned to retain small RNAs using the RNA Clean & Concentrator-5 kit (Zymo, cat. no. R1015 or R1016). Adaptors were ligated to 5′ phosphorylated RNA, followed by incubation with Buffer L2 (Immagina #IBT0181), GTP (Immagina #IBT0201), MnCl (Immagina #IBT0211), L2 enzyme (Immagina #IBT0191), and Linker MC+ (Immagina #IBT0222) for 1 hour at 37°C, then purified with the RNA Clean & Concentrator-5 kit.

Circularization was achieved by mixing adaptor-ligated RNA with Buffer L3 (Immagina #IBT0231), ATP (Immagina #IBT0172), PEG8000 (Immagina #IBT0251), and enzyme L3 (Immagina #IBT0241), then incubating for 2 hours at 25°C, followed by purification to retain small RNAs. Reverse transcription was performed by incubating circular RNA with dNTPs (Immagina #IBT0301) and Primer L4 (Immagina #IBT0262) at 70°C for 5 minutes, cooling on ice, and then adding Buffer L4 (Immagina #IBT0271), DTT (Immagina #IBT0291), and enzyme L4 (Immagina #IBT0281). The reaction was incubated at 50°C for 40 minutes, heat-inactivated at 70°C for 10 minutes, then treated with AR enzyme (Immagina #IBT0311) at 37°C for 1 hour, followed by 80°C for 20 minutes.

PCR amplification was conducted on 20 µL cDNA with L5 enzyme (Immagina #IBT0321), primers F1 (Immagina #IBT0331) and R1 (Immagina #IBT0341), and nuclease-free water. The program included an initial denaturation at 98°C for 1 minute, followed by 8 cycles of 98°C for 30 seconds, 61°C for 30 seconds, and 72°C for 10 seconds, with a final hold at 4°C. The product was purified with Agencourt AMPure XP beads (Beckman Coulter, cat. no. A63881). A second PCR for library multiplexing incorporated unique dual indices (UDIs) using previous PCR product, L5 enzyme (Immagina #IBT0321), and LACEseq UDIs (10 µM). The program included an initial denaturation at 98°C for 1 minute, followed by 6 cycles of 98°C for 30 seconds, 60°C for 30 seconds, and 72°C for 10 seconds, with a final hold at 4°C. The product was purified using the NucleoSpin Gel and PCR Clean-Up kit. The final library was evaluated with an Agilent 2100 Bioanalyzer using the High Sensitivity DNA kit. Sequencing was performed using the NovaSeq platform, achieving coverage of 150 million reads per sample.

2.2.8 Ribo-seq Data Analysis

Sequenced reads were pre-processed to trim adapter sequences using Cutadapt (version 4.9). Unique Molecular Identifiers (UMIs) were then extracted using UMI-tools (version 1.1.6), and contaminant reads corresponding to rRNAs, tRNAs, and ncRNAs were filtered out by pre-mapping to an artificial 'contaminant' genome using Bowtie2 (version 2.5.4). The pre-processed reads were then aligned to the human transcriptome (ensembl 108) using STAR (version 2.7.10b), and the resulting BAM files were sorted and indexed using Samtools (version 1.21).

Ribosome profiling analysis was performed using the Martian pipeline (Chapter 1, version 1.5.0), using the GENCODE human database release 42. The lengths of ribosome-protected fragments (RPFs) corresponding to monosomes were identified by determining those aligning to the correct reading frame. The primary monosome length was defined as the most frequent RPF length aligning to the correct frame, and a 3-nt frame interval was selected as the monosome length range. Subsequent analyses were restricted to reads within this defined range.

2.2.9 Codon-Anticodon Analysis

Codon-anticodon analysis was performed using the Percudani rules [104], which assume that adenosine at the first (5′) position of the tRNA anticodon undergoes A-to-I (inosine) editing, enabling wobble pairing. In cases where adenosine is absent at this position, guanosine is assumed to wobble instead.

2.2.10 Polysome Fractionation

Cells were treated as described in 'Cell Culture and Treatment', and lysates were prepared as outlined in 'RiboLace-based Ribosome Isolation and RNA Elution'. Lysates were loaded onto a linear 15–50% sucrose gradient and subjected to ultracentrifugation using a Beckman SW41Ti rotor at 40,000 rpm for 1 hour and 40 minutes at 4°C in a Beckman Optima LE-80K Ultracentrifuge. Following ultracentrifugation, gradients were fractionated into 1 mL volumes while continuously monitoring absorbance at 254 nm using an ISCO UA-6 UV detector.

2.2.11 Puromycin Incorporation Assays

Puromycin incorporation was performed using the Surface Sensing of Translation (SunSET) method [105] to assess global protein synthesis. HEK293T cells were seeded in

6-well plates and cultured to approximately 70% confluency. Cells were washed with pre-warmed PBS and treated with puromycin (Merck, cat. no: P8833-25MG) at a final concentration of 0.5 µg/mL for 10 minutes at 37°C. Immediately following treatment, cells were washed twice with ice-cold PBS and lysed in RIPA buffer [50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1% Triton X-100, 0.1% SDS, 0.5% sodium deoxycholate] supplemented with protease inhibitor (Roche cOmplete™ Protease Inhibitor Cocktail used at 1× final concentration, cat. No: #11873580001).

Lysates were additionally treated with Turbo DNase (Thermo Fisher, cat. no: AM2239) at 2 U per sample for 1 hour on ice. Following lysis, samples were centrifuged at maximum speed (≥14,000 × g) for 5 minutes at 4°C to remove debris. Protein concentrations were determined using the Pierce™ BCA Protein Assay Kit (Thermo Fisher, cat. no: 23225) according to manufacturer's instructions. Equal amounts of protein were resolved by SDS–PAGE and analyzed by Western blot using an anti-puromycin primary antibody (DSHB, cat. no: PMY-2A4) and an HRP-conjugated anti-mouse secondary antibody (Sigma-Aldrich, cat. no: #31430).

2.2.12 Data Analysis and Statistics

2.2.12.1 Principal Component Analysis (PCA)

PCA was performed on tRNA-Seq data using custom R scripts. For Ribo-seq data, PCA was conducted on a DESeq2 data frame, which utilized read counts per transcript for all samples (Control and Treatment groups n=3) modeled with a one-factor model 'y ~ Treatment'.

2.2.12.2 Amino-Acid Percentage Calculation

Codon counts were aggregated based on their corresponding amino acids. Amino-acid percentages were calculated for each sample as the ratio of amino acid count to total codon counts.

2.2.12.3 Correlation Analysis

Associations between variables were assessed using either Spearman's rank correlation coefficient (rho) or Pearson's correlation coefficient (r). These were used as estimates of effect size, quantifying the strength and direction of linear (Pearson) and non-linear (Spearman) relationships, respectively. Correlation coefficients and corresponding P-values were calculated using the `cor.test()` function in R (method = 'pearson' or 'spearman').

2.2.13 Per-Site Fragmentation Analysis

After counting reads mapping to full-length tRNAs and tRNA fragments for each reference tRNA, one-factor differential expression analysis was performed (y ~ Type, where Type = {Full-length, Fragment}) separately within each sample group (control, treatment, ribo-tRNA and total-tRNA) using the edgeR package [106]. Resulting p-values were corrected using the Benjamini-Hochberg method, and volcano plots were generated using the p-adjusted values and corresponding log fold changes (Fragment vs Full-length).

To identify precise cleavage sites within tRNAs, a custom python script was implemented. Briefly, the analysis focused on the 5′ termini of aligned reads to detect cleavage hotspots. The 5′ termini of aligned reads from BAM files were extracted and aggregated at start positions by reference tRNA. Positions with sufficient read coverage were tested for significant enrichment of 5′ termini between conditions (Control vs. Treatment, Ribo-tRNA vs. total-tRNA) using edgeR's negative binomial framework. P-values were corrected with the Benjamini–Hochberg method, and sites with significant differences in cleavage frequency were annotated as putative fragmentation loci relative to the tRNA reference coordinates.

2.2.14 LC-MS/MS Analysis of RNA Modifications

100 ng of small RNA (<200 nt) fraction were digested with 1 µL of the Nucleoside Digestion Mix (New England BioLabs, #M0649S) and the mixture was incubated at 37°C for 1 h. Samples were analyzed using an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) coupled to an EASY-nLC 1200 (Thermo Fisher Scientific (Proxeon), Odense, Denmark). Ribonucleosides were loaded directly onto the analytical column and were separated by reversed-phase chromatography using a 50-cm column with an inner diameter of 75 µm, packed with 2 µm C18 particles (Thermo Fisher Scientific, cat # ES903).

Chromatographic gradients started at 93% buffer A and 3% buffer B with a flow rate of 250 nl/min for 5 minutes and gradually increased to 30% buffer B and 70% buffer A in 20 min. After each analysis, the column was washed for 10 min with 0% buffer A and 100% buffer B. Buffer A: 0.1% formic acid in water. Buffer B: 0.1% formic acid in 80% acetonitrile. The mass spectrometer was operated in positive ionization mode with nanospray voltage set at 2.4 kV and source temperature at 275°C.

For the Parallel Reaction Monitoring (PRM) method, the quadrupole isolation window was set to 1.4 m/z, and MS/MS scans were collected over a mass range of m/z 50-450, with detection in the Orbitrap at resolution of 60,000. MSMS fragmentation of defined masses with schedule retention time was performed using HCD at NCE 20, the auto gain control

(AGC) was set to ”Standard” and a maximum injection time of 118 ms was used. In each PRM cycle, one full MS scan at resolution of 120,000 was acquired over a mass range of m/z 220-700 with detection in the Orbitrap mass analyzer. Auto gain control (AGC) was set to 1×10 and the maximum injection time was set to 50 ms.

Serial dilutions were prepared using commercial pure ribonucleosides (0.005-150 pg, Carbosynth, Toronto Research Chemicals) to establish the linear range of quantification and the limit of detection of each compound. A mix of commercial ribonucleosides was injected before and after each batch of samples to assess instrument stability and to be used as an external standard to calibrate the retention time of each ribonucleoside. Acquired data were analyzed with the Skyline-daily software (v24.1.1.254) and extracted precursor areas of the ribonucleosides were used for quantification. The metabolomics data have been deposited to MetaboLights [107] repository with the study identifier MTBLS12806.

2.3 Results

2.3.1 Sequencing Translationally-Active tRNAs

tRIBO-seq was developed to provide simultaneous quantitative measurements of both tRNA abundance and modification patterns from tRNAs actively engaged in translation and total tRNA pools. To specifically sequence 'active' tRNAs, ribosomes were isolated using a puromycin derivative (RsP) [16] followed by an antibody-free and tag-free pull-down procedure (see Methods), thus obtaining the ribosome-bound fraction in a single step (Figure 2.1A). The small RNA fraction was then isolated from each pool and used as input for ligation with splint adaptors [54], required for nanopore tRNA library preparation and sequencing. Notably, mRNA populations from ribo-embedded (Ribo-mRNAs) and total RNA fractions (Total-mRNA) can also be isolated and sequenced from the same samples (Figure 2.1B).

To benchmark the method, tRIBO-seq was applied to HEK293T cell cultures maintained under nutrient-rich conditions (see Methods). Principal component analysis (PCA) of tRNA abundances revealed distinct clustering of ribo-tRNAs and total-tRNAs along the first principal component (Figure 2.1C), suggesting that even under nutrient-replete conditions, these populations exhibit slight differences. To identify specific tRNAs responsible for population separation, differential expression (DE) analysis of tRNA abundances was performed. The initiator methionine tRNA (iMet-CAT) was found to be the most significantly enriched tRNA in the ribosome-embedded fraction, consistent with its essential role in translation initiation [108]. Notably, tryptophan (Trp-CCA) and selenocysteine tRNA (SeC-TCA) were significantly decreased in the ribosome-associated pool, which aligns

with the well-established scarcity of both amino acids in the human proteome [108, 109]. High reproducibility was observed across biological replicates with strong correlation between ribo-tRNA and total-tRNA samples (Figure 2.1D), supporting the robustness of the approach.

The tRNA modification landscape of ribo-tRNAs and total-tRNAs was then comparatively examined, revealing only minor differences in tRNA modification patterns. This suggests that tRNA modifications of ribo-tRNAs largely resemble those of total-tRNAs under nutrient-rich conditions. The nano-tRNAseq library preparation method relies on the 3′-CCA terminus of tRNA molecules for adapter ligation [54]. Specifically, the 5′ splint adapter carries an NTGG overhang that Watson-Crick base-pairs with the conserved CCA-3′ tail of each tRNA, enabling ligation of the adapter and, subsequently, capture of the tRNA molecule by the nanopore motor protein.Reads are therefore only generated from tRNAs whose 3′-CCA tail is intact and accessible for hybridisation. *CCA tail completeness* is thus a sequencing-level metric: it reflects the proportion of tRNA molecules in the input library that retain a full, intact 3′-CCA terminus and are consequently detectable by the nano-tRNAseq method. tRNAs with a truncated or absent CCA tail fail to ligate the splint adapter and are invisible to the assay, meaning that any partial CCA tail signal observed in the data directly reports on the integrity of the 3′ end of the sequenced molecules rather than on any biological property of CCA addition per se. Unexpectedly, our analysis also revealed that ribo-tRNAs and total-tRNAs differed in CCA tail completeness. While ribo-tRNAs displayed complete CCA tails for all tRNAs, required for aminoacylation, a partial loss in CCA tails was found in the total tRNA pool, in agreement with previous work [110]. Collectively, these results suggest that under normal elongation conditions, the total tRNA pool largely reflects the translationally active subset in terms of both tRNA abundances and modifications, indicating that tRNA supply is well matched to cellular translational demands.

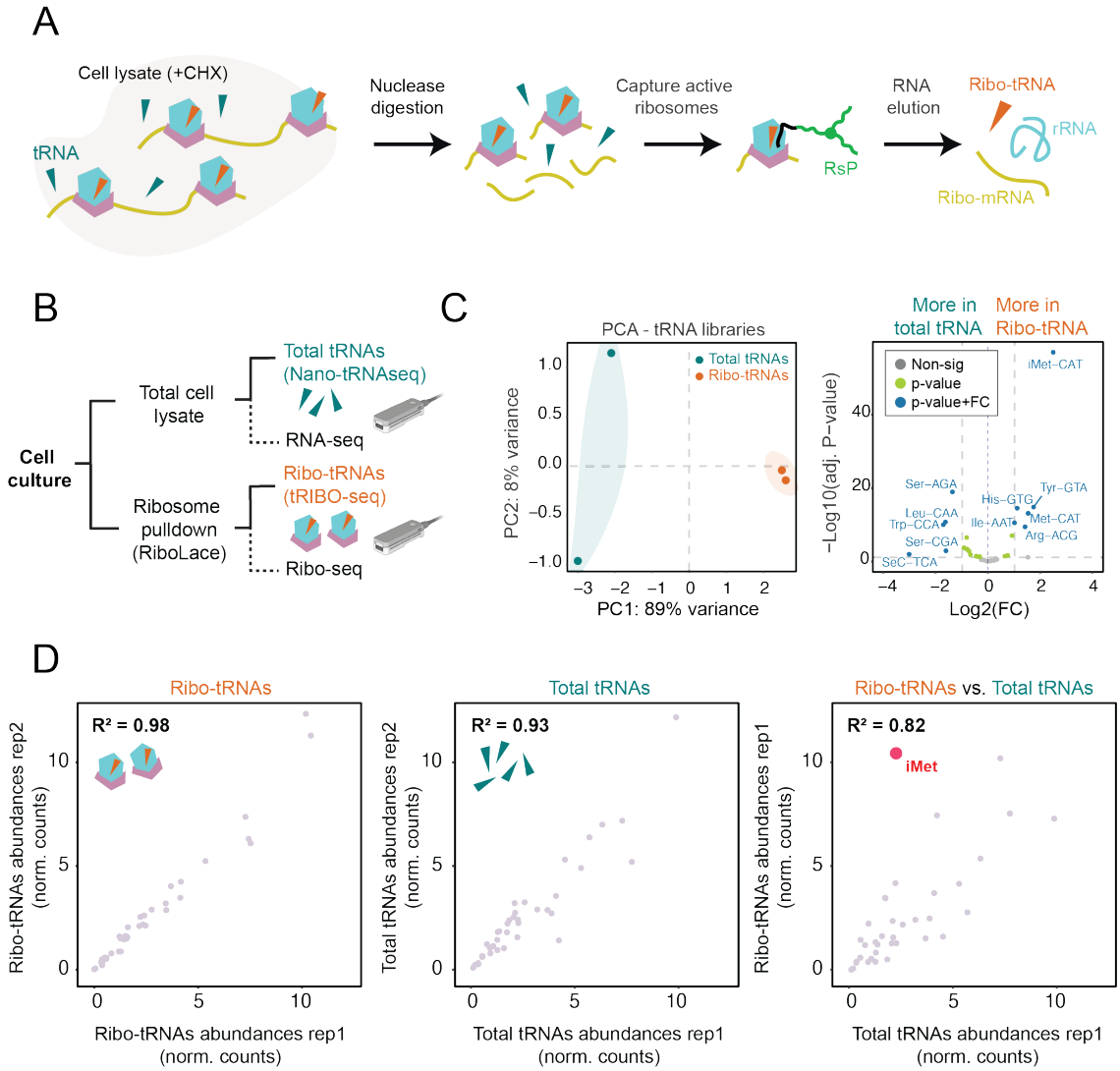

Figure 2.1: Experimental design and benchmarking of tRIBO-seq. (A) Schematic overview of the tRIBO-seq workflow. Translationally active ribosomes were isolated from cell lysates pre-treated with cycloheximide (CHX) to preserve ribosome-mRNA-tRNA complexes. Following nuclease digestion, ribosomes were purified using the RiboLace probe ('RsP') followed by RNA elution. The resulting RNA was used for obtaining ribo-tRNAs and mRNA footprints (Ribo-mRNAs), which can be sequenced using Ribo-seq. (B) Schematic representation of the experimental design: total tRNAs isolated from whole-cell lysates (total-tRNAs, captured using Nano-tRNAseq), and ribosome-associated tRNAs obtained from pulldown fractions (Ribo-tRNAs captured using tRIBO-seq). (C) Left: Principal component analysis (PCA) of tRNA abundances reveals clear separation between total and ribosome-associated tRNAs. Right: Volcano plot showing differential tRNA abundance when comparing ribo-tRNAs and total-tRNAs from HEK293T cell cultures (n = 2 independent biological replicates per condition). (D) Scatterplots depicting the replicability of tRNA abundances between independent biological replicates of HEK293T cultures, for ribo-tRNAs (left) and total-tRNAs (middle). Right: comparison between ribo-tRNA and total-tRNAs. The Pearson R^2 correlation is also shown.

2.3.2 tRIBO-seq Captures Translation Initiation Dynamics

To further validate that tRIBO-seq preferentially captures the ribosome-embedded tRNA fraction, performance was tested in cell cultures treated with harringtonine (HAR), a well-characterized translational inhibitor that binds to the ribosomal A-site and arrests elongation immediately after translation initiation, resulting in ribosome accumulation at start codons [111]. Ribosome profiling experiments were first performed to confirm that HAR exposure led to ribosome accumulation at ATG start codons, which was evidenced by strong accumulation of P-site reads at start codons (Figure 2.2A, bottom panel). Differential expression analysis of ribo-tRNA abundances between HAR-treated and control cells was then performed, revealing that ribo-tRNAs showed significant enrichment of initiator tRNA (iMet-CAT) in HAR-treated cells (Figure 2.2B left panel), consistent with the drug-induced block of ribosomes at start codons. By contrast, total-tRNAs did not show significant changes in iMet tRNA abundance, or of any other tRNA, confirming the specificity of tRIBO-seq for capturing ribosome-embedded tRNAs (Figure 2.2B, right panel).

Correlations between tRNA abundances of ribo-tRNAs and total-tRNAs with respect to codon usage were next examined (see Methods). Codon usage was quantified by extracting P-site codons from matched Ribo-Seq experiments on the same sample sets—specifically, the control (elongating) condition. Compared to total-tRNAs, ribosome-associated tRNAs exhibited stronger correlation with P-site codon usage. This was observed both at the amino acid level (Spearman's $\rho = 0.76$ for ribo-tRNAs vs. 0.57 for total-tRNAs) and with codon-anticodon matching using Percudani rules [104] ($\rho = 0.57$ vs. 0.31) (Figure 2.2C). This enhanced correlation validates the specificity of tRIBO-seq for capturing translationally active tRNAs.

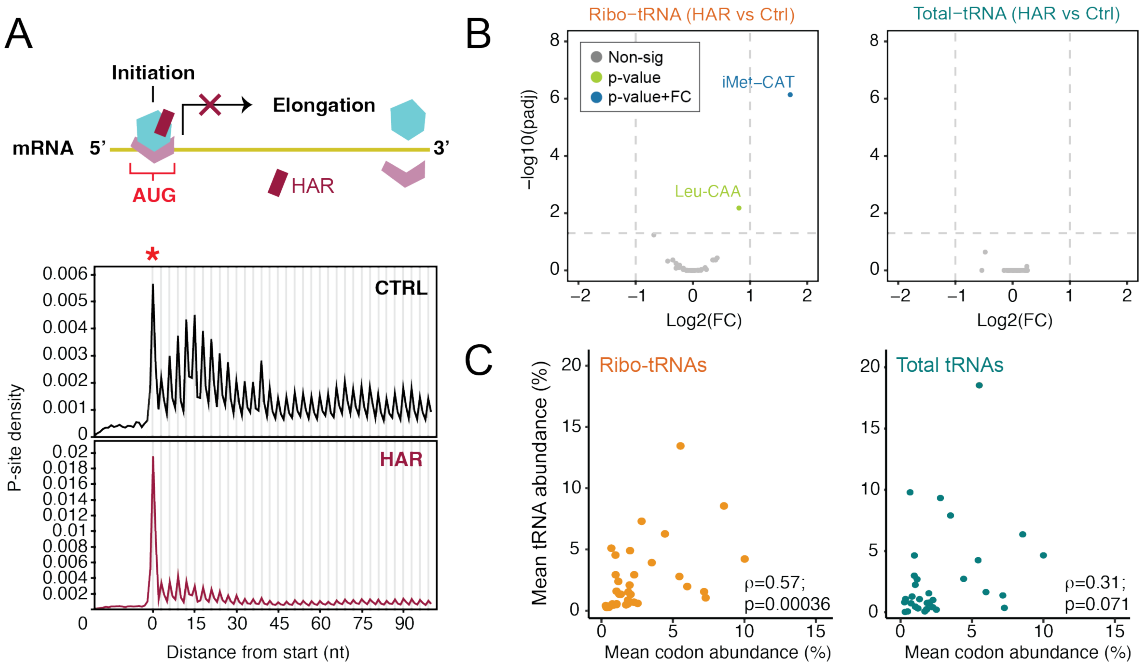

Figure 2.2: Validation of tRIBO-seq specificity using harringtonine. (A) Upper panel: Schematic representation of harringtonine (HAR) mechanism: HAR inhibits translation elongation, resulting in ribosome stalling at start codons. Bottom panel: Metagene plots obtained from ribosome profiling data in control (CTRL) and HAR-treated cells, showing accumulation of P-site reads at start codons under HAR treatment. (B) Volcano plot depicting differential tRNA abundance between HAR-treated and control MCF-7 cells, for both ribo-tRNAs (left) and total tRNAs (right). Blue dots indicate tRNAs meeting both statistical significance and fold-change thresholds; green dots represent tRNAs meeting only statistical significance; gray dots correspond to non-significant changes. (C) Scatterplots showing the correlation between mean codon abundance at the ribosomal P-site in control cells (x-axis) with tRNA abundances, using either ribo-tRNAs (left) or total tRNAs (right) populations (y-axis), respectively (n = 35). Spearman's rho and associated p-values are also shown. All experiments were performed with n = 3 independent biological replicates per condition.

2.3.3 Ribo-tRNAs are Sensitive to Diverse Amino Acid Deprivations

Whether tRIBO-seq would distinguish cellular responses to different modes of translational stress, such as diverse amino acid deprivations, was next investigated. Previous studies have shown that both arginine (Arg) and leucine (Leu) starvation can impair translation through distinct mechanisms elongation stalling in the case of Arg starvation, and initiation inhibition in the case of Leu starvation [112] (Figure 2.3A). To assess whether Arg or Leu deprivation would result in distinguishable ribo-tRNA profiles, HEK293T cells were treated for 3 hours with media lacking either amino acid, using fully supplemented medium as a matched control (Figure 2.3B).

Both Arg and Leu 3-hour deprivation conditions were first confirmed to significantly affect translational output using puromycin incorporation assays (Figure 2.4A. This was

further supported by polysome profiling, which showed significant disruption following both Arg and Leu deprivation already at 3 hours. Ribo-tRNA and total-tRNA populations from 3-hour Arg- and Leu-deprived cells, as well as their control conditions, were then sequenced. PCA of ribo-tRNAs revealed clear separation of starved versus control samples along the first principal component for both amino acid deprivation types (Figure 2.3C). By contrast, total-tRNAs showed weaker clustering, indicating that ribo-tRNA composition varies strongly upon amino acid deprivation, whereas the global tRNA pool is only modestly affected by amino acid deprivation (Figure 2.5. In both treated and untreated samples, strong enrichment in iMet-CAT tRNA was observed when comparing ribo-tRNAs and total-tRNAs, both in Leu- and Arg-deprived cells, in agreement with previous observations (Figure 2.1D and 2.2B).

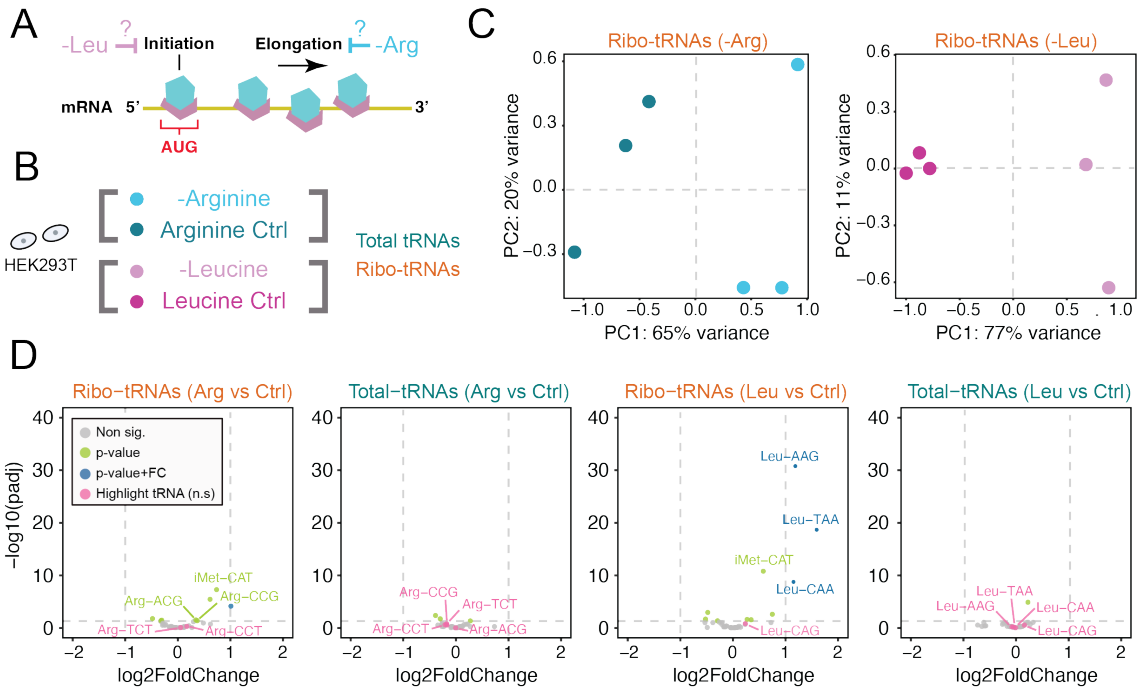

Figure 2.3: tRIBO-seq reveals distinct tRNA profiles under amino acid starvation. (A) Schematic of experimental rationale. Arg and Leu starvation perturb translation via different mechanisms elongation stalling and initiation inhibition, respectively. (B) Schematic for the experimental design, in which HEK293T cells were cultured in amino acid-deprived media for either Arg or Leu for 3 hours and their matching control media and subsequently analyzed by both tRIBO-seq (ribo-tRNAs) and Nano-tRNAseq (total-tRNAs). Each condition was performed with n = 3 independent biological replicates. (C) PCA of ribo-tRNAs shows separation of starved versus control samples along PC1, for both Arg (65% variance explained) and Leu (77%) starvation. Color coding corresponds to the legend in panel B. (D) Volcano plots showing differential abundance of tRNAs in ribosome-associated and total pools for both starvation conditions.

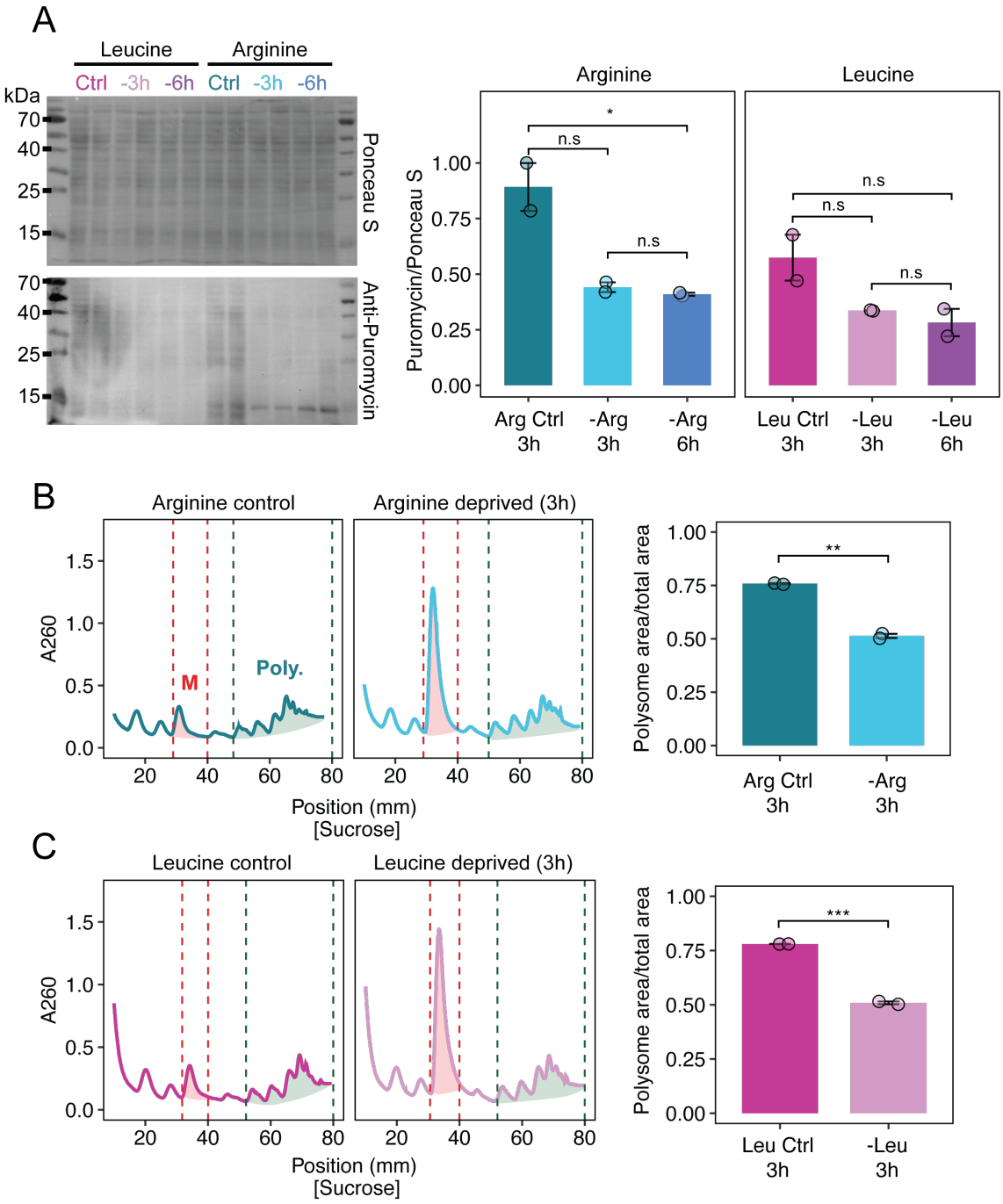

Figure 2.4: Validation of amino acid deprivation effects on cellular translation activity. (A) Global protein synthesis measured by puromycin incorporation assay in HEK293T cells under amino acid starvation (3h or 6h) compared to control. Western blot shows puromycin-labeled nascent peptides (lower panel) with Ponceau S staining as loading control (upper panel). Quantification (right) demonstrates significant reduction in protein synthesis at both timepoints. (B) Polysome profiling under arginine deprivation. Representative traces (left) show monosome (M, red) and polysome (Poly., green) peaks. Quantification (right) reveals significant decrease in polysome-associated ribosomes at 3h and 6h compared to control. (C) Polysome profiling under leucine deprivation. Representative traces (left) and quantification (right) confirm reduced active translation under starvation conditions. All error bars represent standard deviation from biological duplicates (n = 2). Statistical significance determined by two-sided Student's t-test: n.s., not significant; *P < 0.05; **P < 0.01; ***P < 0.001.

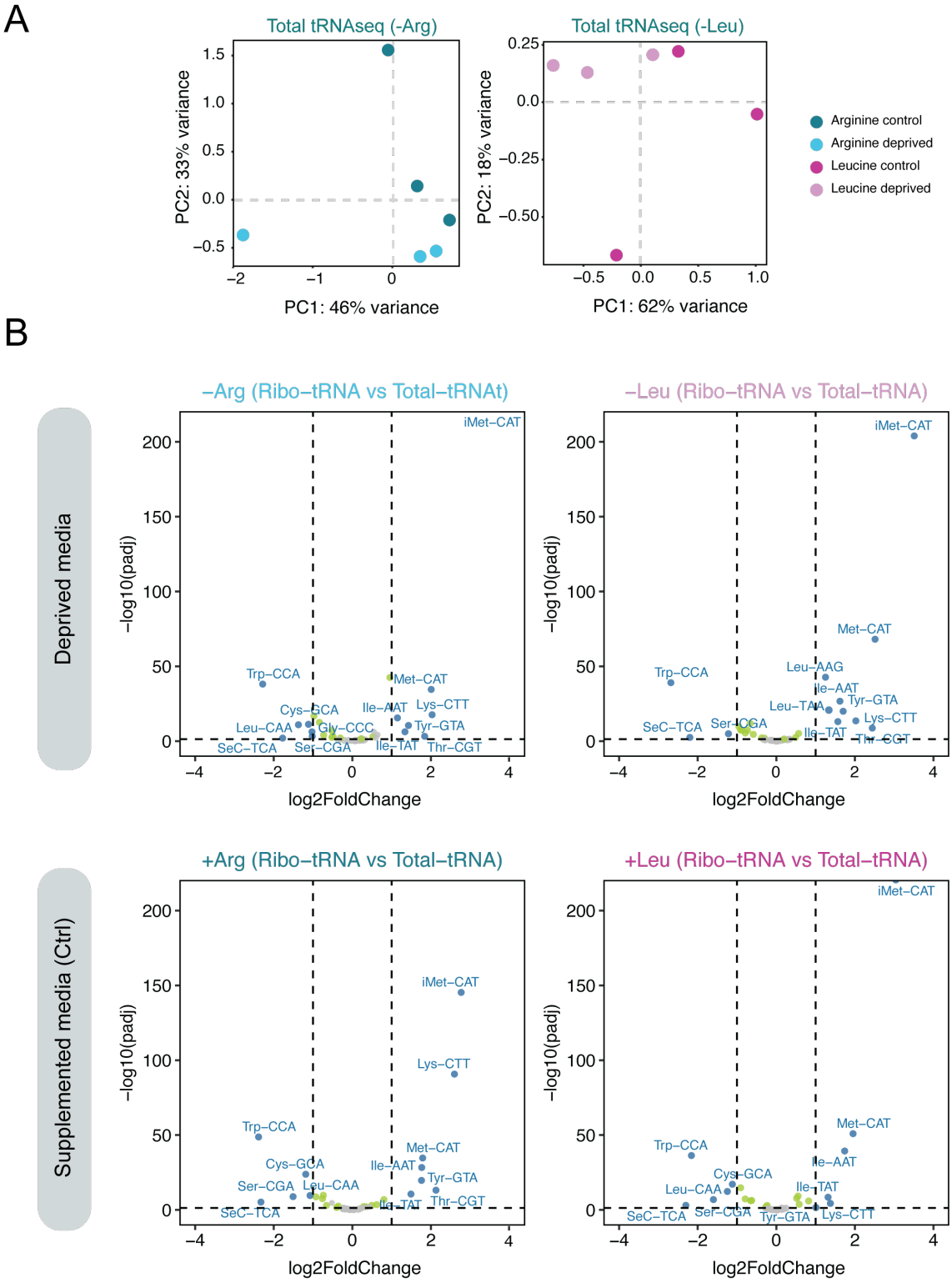

Figure 2.5: tRNA population analysis across amino acid deprivation conditions. (A) PCA of total-tRNA profiles across all conditions (n = 3 biological replicates per condition). (B) Volcano plots showing differential abundance between ribo-tRNAs and total-tRNAs for each condition. Positive fold changes indicate ribo-tRNA enrichment.

Differential tRNA expression analysis of ribo-tRNAs from Arg- and Leu-deprived compared to control samples was next performed. This analysis revealed significant enrich-

ment of initiator methionine tRNA (iMet) in ribosomes from both Arg- and Leu-starved cells (Figure 2.3D), suggesting a translational shift toward initiation pausing or slowed elongation. Unexpectedly, enrichment of cognate tRNAs in the ribosome-associated pool under starvation was also observed, contrasting with initial expectations of depletion of these tRNAs in the ribo-tRNA pool upon amino acid starvation. This trend was most pronounced upon leucine starvation, where several Leu tRNA isoacceptors were significantly enriched (Figure 2.3D). In the arginine condition, this pattern was more nuanced—only two isoacceptors, Arg-ACG and Arg-CCG, passed the significance threshold, but not the fold-change threshold.

2.3.4 Methionine Starvation Causes Differential Modification of Ribo-Embedded tRNAs

Amino acid deprivation can trigger distinct translational responses. We therefore investigated whether methionine starvation affects tRNA dynamics similarly to leucine or arginine deprivation, or produces a unique response. Methionine (Met) is the initiator amino acid, and its deprivation has been previously shown to cause drastic decreases in translation, assessed both by polysome profiling and puromycin incorporation assays [113]. Additionally, Met starvation has been shown to cause strong transcriptional responses [114]. However, the effect of methionine starvation on the tRNA pool has not been explored. MCF-7 cells were subjected to 6 and 16 hours of Met deprivation (Figure 2.6A), and their tRNA profiles in both ribo-embedded and total tRNA populations were examined, coupled with matched Ribo-seq and RNA-seq experiments.

Methionine deprivation was first confirmed to yield strong transcriptional responses both at 6 and 16 hours post-deprivation, in agreement with previous work [114]. Despite the strong transcriptional response, analysis of codon occupancy from Ribo-seq data revealed no significant changes at methionine codons upon Met deprivation, suggesting that most impairment occurs at the translation initiation stage rather than at the elongation stage. In agreement with this, ribo-tRNA and total-tRNA abundances remained relatively stable upon Met starvation, with no significant changes at 6 hours post-deprivation and only modest changes in ribo-tRNAs at 16 hours (Figure 2.6B).

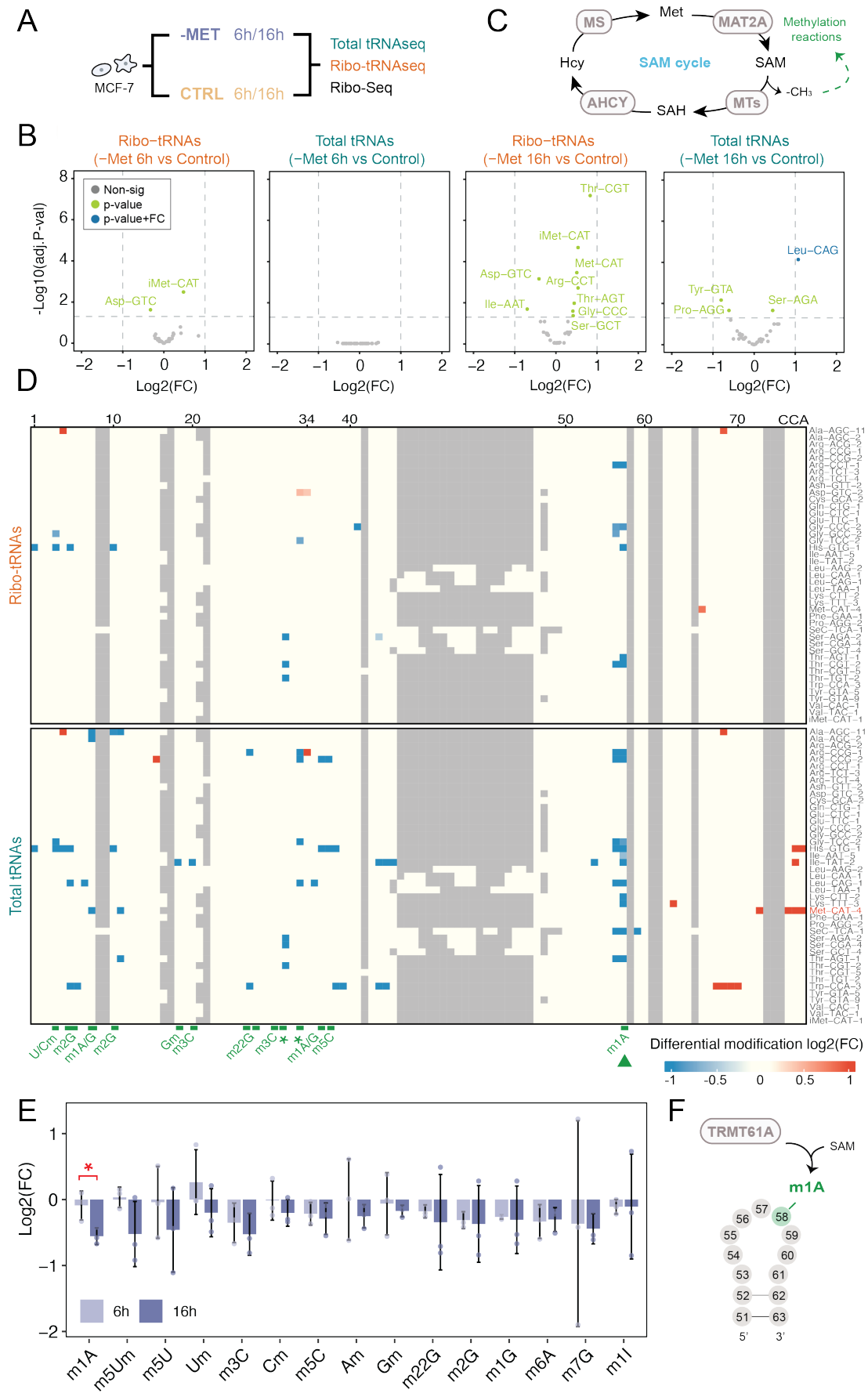
A
MCF-7
-MET 6h/16h
CTRL 6h/16h
Total tRNAseq
Ribo-tRNAseq
Ribo-Seq
C
MS
Met
MAT2A
Methylation reactions
Hcy
SAM cycle
SAM
AHCY
SAH
MTs
-CH3
B
Ribo−tRNAs (−Met 6h vs Control)
Total tRNAs (−Met 6h vs Control)
Ribo−tRNAs (−Met 16h vs Control)
Total tRNAs (−Met 16h vs Control)
Non-sig
p-value
p-value+FC
-Log10(adj.P-val)
Log2(FC)
iMet−CAT
Asp−GTC
Thr−CGT
Met−CAT
Arg−CCT
Thr+AGT
Gly−CCC
Ser−GCT
Ile−AAT
Leu−CAG
Tyr−GTA
Pro−AGG
Ser−AGA
D
Ribo-tRNAs
Total tRNAs
CCA
U/Cm
m2G
m1A/G
m2G
Gm
m3C
m22G
m3C
m1A/G
m5C
m1A
Differential modification log2(FC)
E
Log2(FC)
6h
16h
m1A
m5Um
m5U
Um
m3C
Cm
m5C
Am
Gm
m22G
m2G
m1G
m6A
m7G
m1I
F
TRMT61A
SAM
m1A
5'
3'

Figure 2.6: Methionine starvation causes loss of tRNA methylation on tRNAs at 16h. (A) Schematic of the experimental design. MCF-7 cells were cultured in Met-deprived media for 6 or 16 hours, alongside matched controls, with n = 3 biological replicates per condition for all experiments. Total and ribosome-associated tRNAs were analyzed using tRIBO-seq (ribo-tRNAs) and Nano-tRNA-seq (total-tRNAs). (B) Volcano plots showing differential abundance of tRNAs in the ribosome-associated (ribo-tRNAs) and total pools (total-tRNAs), for both 6-hour and 16-hour Met starvation conditions. (C) Schematic of the S-adenosylmethionine (SAM) biosynthesis and utilization cycle. (D) Heatmaps depicting differential modification (methionine-starved vs. control) for Ribo-tRNAs (top) and total tRNAs (bottom) at 16 hours. Differential basecalling errors are used as a proxy to quantify differential modifications across two conditions, as previously described [54], [3]. The x-axis represents nucleotide positions; the y-axis lists tRNA isoacceptors (alphabetically ordered). Common methylation sites are annotated in green. (E) LC–MS/MS quantification of modified nucleosides in ribo-tRNA isolated from methionine-starved and control cells. Bars represent mean ± s.d. from n = 3 biological replicates per condition. P-values were calculated using a paired Student's t-test comparing each condition to the corresponding control. Significance is indicated based on raw p-values. Given the high variability inherent to nucleoside MS measurements and the hypothesis-driven nature of the comparison, raw p-values are reported for exploratory interpretation. (F) Schematic of the tRNA T-loop structure, highlighting m^1A58 (in green), the methylation target of TRMT61A. Statistical significance is annotated as follows: *P < 0.05, **P < 0.01, ***P < 0.001.

The set of modestly-upregulated ribo-tRNAs upon 16-hour Met deprivation encompassed methionine, arginine, glycine, serine, and threonine (Figure 2.6B). Notably, all these amino acids are biochemically linked via the S-adenosylmethionine (SAM) production pathway. SAM acts as a universal methyl donor for numerous biological processes, including tRNA modification. In agreement with this, robust upregulation of MAT2A was observed, which encodes the enzyme responsible for synthesizing SAM from methionine (Figure 2.6C). Of note, MCF-7 cells are deficient in MTAP [114], a key enzyme in the methionine salvage pathway, making them reliant on de novo SAM synthesis via MAT2A.

Given the observed dysregulation of SAM production pathways upon Met starvation, whether tRNA modifications requiring methyl groups would be significantly altered upon Met starvation was investigated. While no significant changes were observed at 6 hours, prolonged deprivation (16 hours) led to strong and widespread hypomethylation of tRNAs, particularly in the total-tRNA pool, possibly reflecting reduced intracellular SAM availability. The hypomethylated positions overlapped with known methylation sites [115]. By contrast, the effect in ribo-tRNAs was more modest (Figure 2.6D), supporting the notion that proper tRNA modification is essential for decoding function, as previously described [39]. To further validate these observations, nucleoside liquid chromatography-tandem mass spectrometry (LC-MS/MS) was performed on control and Met-starved samples, revealing that most methylation marks showed reduced modification levels, with m^1A being significantly decreased upon Met starvation (Figure 2.6E). It should be noted that m^1A is predominantly found at position 58 in the tRNA T-loop and is deposited by TRMT61A, a SAM-dependent methyltransferase (Figure 2.6F).

2.3.5 Oxidative Stress Triggers tRF Formation in Ribosome-Associated tRNAs

Whether impairment of translation via chemical treatment would cause a distinct tRNA response compared to amino acid deprivation was next investigated. To this end, MCF-7 cells were treated with a high concentration of sodium arsenite (Ars), a well-characterized inducer of oxidative stress that acutely inhibits translation by triggering the integrated stress response (ISR) for a short time (50 minutes) [116, 117] (Figure 2.7A). The rationale of this condition was to move a translatomic response without a transcriptional rewiring.
Polysome profiling was first performed to confirm that arsenite exposure caused global shutdown of active translation (Figure 2.7A). This translational shutdown was accompanied by transcriptional upregulation of ISR effectors, such as HSPA, ATF3, and PPP1R15A, indicative of translational repression and stress adaptation. Indeed, 1-hour arsenite treatment resulted in 966 genes being exclusively translationally regulated, compared to 19 transcriptionally regulated genes and 13 affected at both levels, suggesting that post-transcriptional mechanisms dominate the cellular response to arsenic-induced stress, in contrast with the predominantly transcriptional response observed upon Met deprivation.
To assess how oxidative stress influenced tRNA dynamics, tRIBO-seq was performed in cells treated with arsenite and untreated controls (Figure 2.7A). PCA and correlation analysis revealed that ribo-tRNA abundances, but not total tRNA abundances, were strongly affected by arsenite treatment (Figure 2.7B-C). Differential expression analysis revealed that 14 ribo-tRNAs were significantly altered upon arsenite treatment, in contrast to only 2 changing in the total-tRNA pool. This suggests that oxidative stress primarily affects actively translating tRNAs, consistent with the strong translational phenotype induced by arsenite. Among the few differentially expressed total tRNAs, the initiator methionine tRNA (iMet) was significantly affected (Figure 2.7D, 'Total-tRNA (Arsenite vs Ctrl)'), potentially reflecting its sequestration in stalled pre-initiation complexes, in agreement with ISR activation [117]. Within the ribo-tRNAs, consistent depletion of tRNAs decoding cysteine, glutamate, and glycine was observed all of which are precursors of glutathione (GSH), a key antioxidant [118] (Figure 2.7D 'Ribo-tRNA (Arsenite vs Ctrl)'). This pattern suggests a stress-induced metabolic shift, possibly reflecting impaired tRNA charging in response to oxidative stress. Unfortunately, arsenite treatment at the condition tested ($1\mu M$) disrupted codon periodicity of ribosome footprints, limiting the ability to assess codon-level translation dynamics (Figure 2.8).

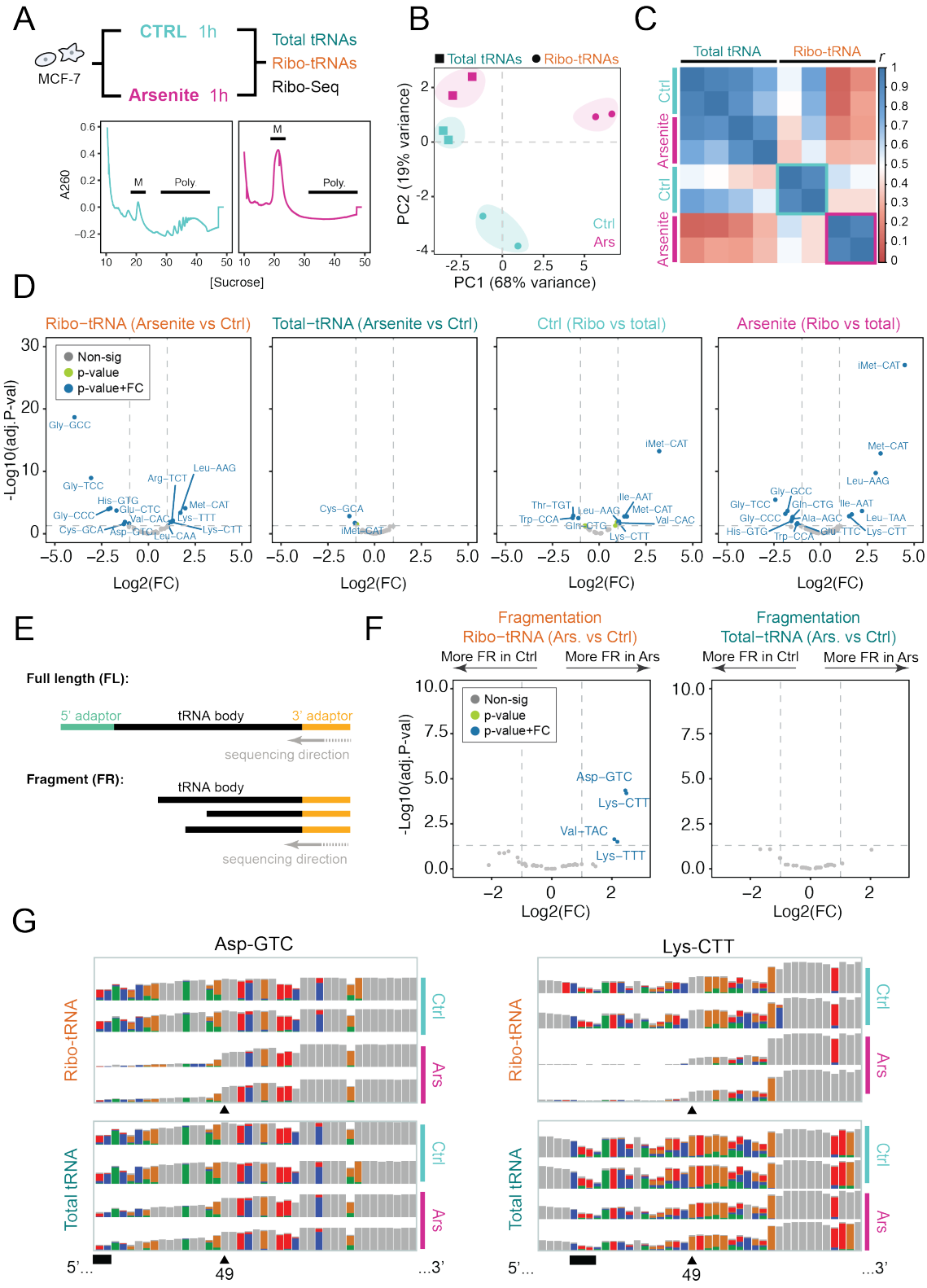
A
CTRL 1h
Arsenite 1h
MCF-7
Total tRNAs
Ribo-tRNAs
Ribo-Seq
A260
M
Poly
[Sucrose]
B
Total tRNAs
Ribo-tRNAs
PC2 (19% variance)
PC1 (68% variance)
Ctrl
Ars
C
Total tRNA
Ribo-tRNA
Ctrl
Arsenite
D
Ribo−tRNA (Arsenite vs Ctrl)
Total−tRNA (Arsenite vs Ctrl)
Ctrl (Ribo vs total)
Arsenite (Ribo vs total)
Non-sig
p-value
p-value+FC
-Log10(adj.P-val)
Log2(FC)
E
Full length (FL):
5' adaptor
tRNA body
3' adaptor
sequencing direction
Fragment (FR):
F
Fragmentation
Ribo−tRNA (Ars. vs Ctrl)
Total−tRNA (Ars. vs Ctrl)
More FR in Ctrl
More FR in Ars
G
Asp-GTC
Lys-CTT
Ribo-tRNA
Total tRNA
Ctrl
Ars
5'…
…3'
49

Figure 2.7: Selective fragmentation of Ribo-tRNAs in arsenite-treated samples. (A) Top: Schematic of the experimental design. MCF-7 cells were treated with 1 mM sodium arsenite for 50 minutes (Arsenite) or left untreated (CTRL). Total tRNAs and ribosome-associated tRNAs were profiled using Nano-tRNA-seq and tRIBO-seq, respectively. All experiments were performed with n = 2 independent biological replicates per condition. Bottom: Representative polysome profiles indicate translational repression following arsenite treatment. Peaks corresponding to monosomes (M) and polysomes (Poly.) are labeled. (B) PCA of total tRNAs and ribo-tRNAs. Arsenite treatment induces greater separation in ribo-tRNA profiles (circles) compared to total tRNA profiles (squares), indicating stronger effects on ribosome-associated tRNAs. (C) Pearson correlation matrix comparing tRNA abundance profiles across all conditions and library types (total vs ribo-tRNA). (D) Volcano plots showing differential abundance of tRNAs across all conditions and library types (total vs ribo-tRNA). (E) Schematic illustrating classification of full-length (FL) vs fragmented reads. FL reads contain both 5′ and 3′ adaptors; fragmented reads contain only a 3′ adaptor. (F) Volcano plots showing differential fragmentation (FR vs FL) between arsenite and control samples (see Methods). (G) IGV genome browser snapshots for two representative tRNAs showing arsenite-induced fragmentation. For each condition and tRNA pool, two independent biological replicates are shown. Arrow indicates a significant fragmentation site. Anticodon underlined in black.

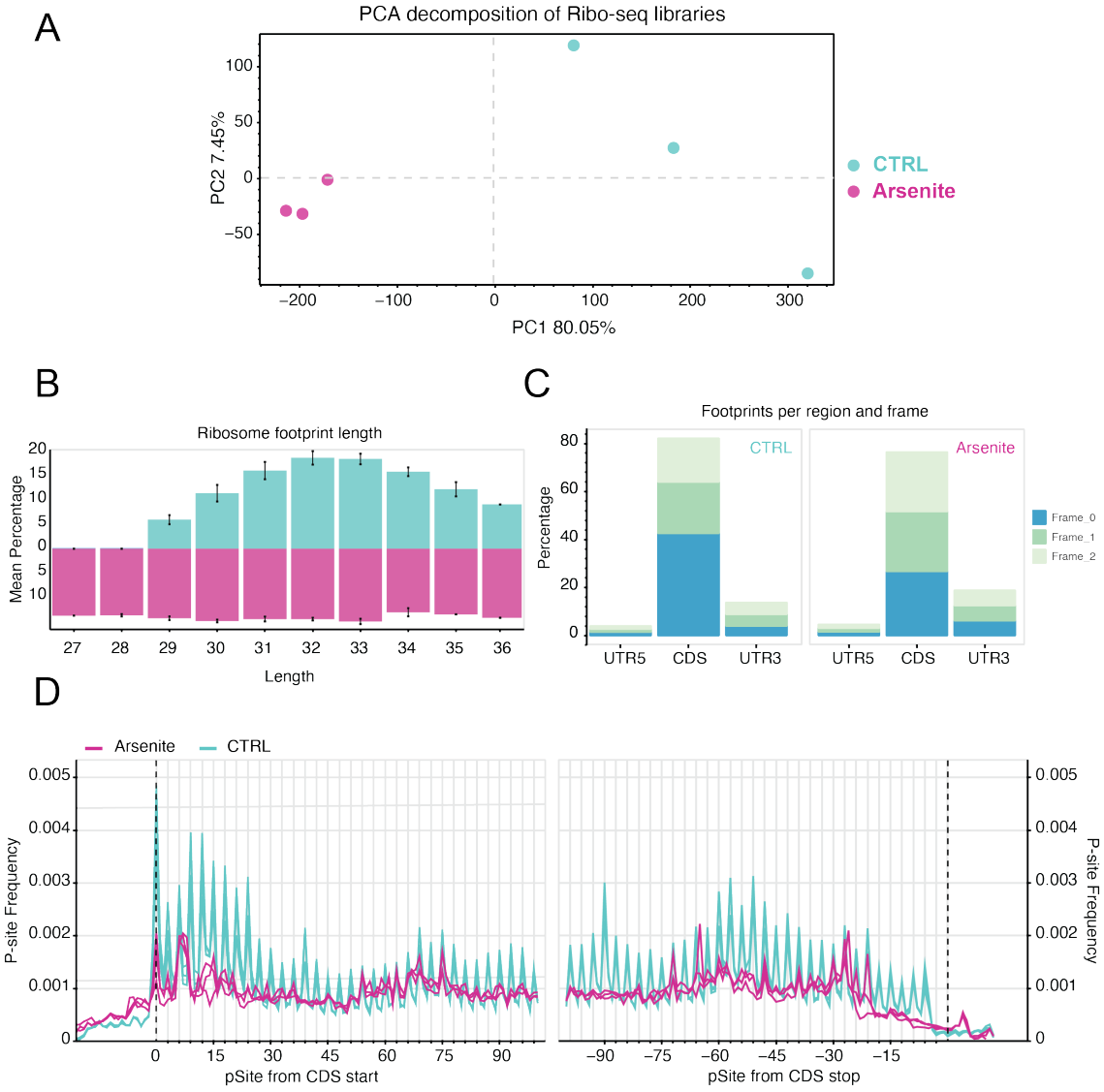

Figure 2.8: Ribosome profiling quality control for arsenite treatment. (A) PCA of Ribo-seq data showing separation between arsenite-treated and control samples (n = 3 per condition). (B) RPF length distribution after preprocessing (mean ± s.d., n = 3). (C) Distribution of RPFs across reading frames and genomic regions (5'UTR, CDS, 3'UTR). (D) Metagene profiles of ribosome occupancy at start and stop codons for both conditions (n = 3 biological replicates).

The tRNA modification profiles were then examined, revealing that, in contrast to methionine deprivation, arsenite stress did not significantly alter the tRNA modification landscape, neither in ribo-tRNAs nor total-tRNAs. However, closer examination of the IGV tracks revealed another phenomenon: a drastic increase in tRNA-derived fragments (tRFs) in ribo-tRNAs upon arsenite treatment relative to the control condition.

Arsenite is well-documented to induce tRNA fragmentation [119–121]. To determine whether tRIBO-seq was capturing tRNA-derived fragments, fragmentation analysis was performed, in which full-length tRNA counts were compared to tRNA fragment (tRF) counts for each tRNA species (Figure 2.7F and Figure 2.7E, see also Methods). This analysis revealed four tRNAs significantly enriched in the ribosome-associated fraction following arsenite treatment: Asp-GTC, Lys-CTT, Val-TAC, and Lys-TTT (Figure 2.7F). To exclude the possibility that the observed enrichment could result from nuclease digestion during tRIBO-seq preparation (Figure 2.1A), the same fragmentation analysis was applied across all datasets presented in this study. No significant enrichment was detected when performing fragmentation analysis in other treatment conditions, suggesting the observations are not an artifact of digestion and are largely unique to arsenite treatment.

To further dissect the nature of ribosome-associated tRNA fragments, whether specific cleavage sites could be identified along the tRNA body was next investigated. To this end, position-resolved fragmentation analysis was performed (see Methods). This revealed prominent cleavage sites in several tRNAs around position 49 (Figure 2.7G), suggesting a non-random, site-specific degradation pattern. Due to the nature of nanopore direct RNA sequencing, if the 5′ adapter is not ligated, which would be the case in tRNA fragments, the read will finish ~13-15 bases before the true end of the RNA molecule [122, 123], suggesting that the true cleavage site corresponds approximately to position ~34-36 of these tRNAs, which corresponds to the anticodon region (Figure 2.7G). In conclusion, tRIBO-seq captures arsenite-induced tRNA cleavage (leading to tRNA halves formation), which was found to occur preferentially in the ribo-embedded tRNA fraction (Figure 2.7F-G). This result is in agreement with previous work reporting that angiogenin activation occurs in the ribosome and thus likely preferentially degrades tRNAs in the ribosome [86].

2.4 Discussion and Conclusion

For decades, efforts have been made to understand and characterize how cells tune their translational output in response to stimuli. However, a simple and cost-effective approach to study the active tRNAome has been lacking. Consequently, previous work has largely

focused on charting the dynamics of total tRNA pools, which, as demonstrated here, can significantly differ from those tRNAs being actively used for translation.

The present work introduces tRIBO-seq, a novel and robust approach that enables multidimensional resolution of tRNA abundance, modifications, and fragmentomics. By specifically profiling tRNAs associated with ribosomes, this method offers a focused snapshot of the tRNAs actively engaged in translation, substantially enhancing the granularity of data obtained, particularly under stress conditions. Notably, mRNA populations can also be recovered and sequenced from the same ribosome-embedded samples (Ribo-Seq), allowing investigation of codon-anticodon interactions with minimal batch effects. The protocol can be completed within 5 hours and requires minimal input material of 3-5 µg of ribo-embedded RNA, which corresponds to approximately 5 million cells—depending on cellular translational status [16].

The ribo-tRNA and total-tRNA landscapes were examined under four different conditions. First, both tRNA pools under steady-state, nutrient-rich conditions were examined, revealing that the subset engaged with ribosomes largely mimics the total tRNA pool (Figure 1), indicating tight coordination between tRNA supply and the translational needs of the cell. This aligns with the concept that, under steady-state conditions, cells maintain an optimized tRNA repertoire to support efficient translation. Subsequently, how different stress-specific alterations reshaped this coordination was examined. Specifically, tRNAome behavior upon amino acid deprivation was observed. Upon Leu starvation—but not Arg starvation—the abundance of specific Leu-tRNAs was significantly altered in the ribo-tRNA pool (Figure 3). By contrast, methionine deprivation did not impact tRNA abundance but rather caused widespread hypomethylation of the total tRNA pool at multiple well-characterized tRNA-modified sites (Figure 4). The ribo-tRNA pool was also hypo-modified, albeit to a lesser extent, suggesting a coordinated cellular response to preserve translational fidelity, in which modified tRNAs are preferentially loaded onto ribosomes—possibly mediated by preferential aminoacylation of fully-modified tRNAs.

In recent years, it has been proposed that the tRNAome is rewired upon external environmental conditions [34, 124, 125]. Here, in agreement with previous work [76], for at least some cellular models and environmental conditions, the total tRNAome does not significantly change upon external stimuli; rather, analyses reveal that ribo-tRNA populations are often the ones altered in response to specific stimuli. This work highlights the importance of examining both ribo-embedded and total tRNAome population dynamics to comprehensively understand how the tRNAome is used in translation and rewired upon environmental cues. Notably, tRIBO-seq greatly simplifies the analysis of both ribo-tRNA and total-tRNA populations from a single sample.

Several caveats and limitations exist when applying this technique. First, the performance of tRIBO-seq will vary depending on cellular translational activity. For example, neuronal progenitors may require higher amounts of input material (e.g., 100-200 µg of total RNA)

compared to cancerous cell lines such as HEK or MCF-7 (10-20 µg total RNA). Second, the protocol has only been tested in mammalian cell lines; it could possibly be extended to non-mammalian eukaryotic samples with some optimizations and variations in the protocol, but this has not been tested. Finally, while tRIBO-seq can capture both full-length tRNAs and tRNA-derived fragments (tRFs), it should be noted that it will only capture 3′tRFs, but not 5′tRFs, due to selective CC(A) ligation. However, tRIBO-seq can be used to study tRF presence and formation while also capturing full-length tRNA abundances and modification patterns.

tRIBO-seq was developed collaboratively by me, the laboratory of Dr. Eva Maria Novoa (Centre for Genomic Regulation, Barcelona), and Immagina Biotechnology s.r.l. The contributions of each party are delineated as follows. The experimental protocol, including the adaptation of the RiboLace-based ribosome isolation procedure, the optimisation of splint adapter ligation for ribosome-embedded tRNAs, the decision to omit the deacylation step for CHX-treated libraries, and the design and execution of all cell culture treatments (arginine/leucine/methionine starvation and arsenite exposure), was led by collaborators at the Novoa laboratory and Immagina Biotechnology. My contributions were primarily computational and analytical: I was responsible for the development and implementation of all bioinformatics pipelines used for tRNA modification detection and quantification (including the modFinder algorithm described in Chapter 1), differential tRNA abundance analysis, and fragmentation analysis. I was the primary analyst for the complete arsenite treatment dataset, performing both the ribosome profiling analysis (using the Martian pipeline) and the nano-tRNA sequencing analysis from raw data through to biological interpretation. I contributed to the analysis of the methionine starvation, arginine deprivation, and leucine deprivation datasets in close collaboration with members of the Novoa laboratory. The integration of tRIBO-seq outputs with Ribo-seq codon usage data was also conducted by me. The manuscript arising from this work is currently under revision at *Nature Methods* and it is available on *BiorXiv doi.org/10.64898/2026.03.02.709006*.

Note: Some content in this chapter has been adapted from the manuscript in preparation for clarity and consistency within this thesis.

Chapter 3

Conclusions and Future Directions

Translation regulation represents a critical layer of gene expression control with profound implications for human health and disease. As outlined in the introduction, translation dysregulation affects the majority of human tumors, underlies multiple neurological disorders, and represents an emerging target for precision medicine approaches. This thesis addressed fundamental challenges in translational biology through the development of two complementary tools: Martian, an integrated computational pipeline for comprehensive ribosome profiling analysis, and contributions to tRIBO-seq, a novel method for profiling translationally active transfer RNAs. Together, these approaches advance our ability to interrogate translation regulation at both the ribosome and tRNA levels, providing new insights into the coordination between mRNA decoding and tRNA availability under diverse cellular conditions.

3.1 Summary of Key Findings

3.1.1 Martian: Integrated Ribosome Profiling Analysis

Benchmarking of Martian against established tools (riboWaltz and Plastid) revealed that these platforms represent fundamentally different design philosophies, each optimizing for distinct priorities. Martian achieved the fastest runtime (1,896 seconds, 37% improvement over riboWaltz) while maintaining competitive P-site detection accuracy (TIS score 0.91 versus riboWaltz's 1.00), demonstrating that comprehensive analysis and computational efficiency are not mutually exclusive goals. The stringent database curation approach implemented in Martian prioritizes high-confidence annotations over comprehensive coverage, resulting in reduced absolute transcript detection (32% fewer than riboWaltz) but enhanced reliability for downstream analyses including codon usage calculation, translation efficiency quantification, and regional assignment.

The integration of five analysis modules, database creation, ribosome profiling, condition comparison, translation efficiency calculation, and nano-tRNA sequencing, within

a unified framework directly addresses the fragmentation problem that affecting current ribosome profiling workflows. By eliminating the need to manually orchestrate multiple specialized tools with inconsistent data formats and normalization strategies, Martian reduces analytical bottlenecks while ensuring reproducible, standardized processing. The intermediate memory requirements and automated workflows accessible to users without programming expertise position Martian favorably for production analyses on standard computational infrastructure.

Importantly, Martian has demonstrated practical utility beyond this benchmarking study. Its analytical capabilities have been employed in collaborative research examining ribosome recycling mechanisms and translational control during cell state transitions, as detailed in recent work on ribosome behavior under specific cellular conditions [126]. This real-world application validates Martian's design choices and demonstrates its capacity to support cutting-edge translational research.

3.1.2 tRIBO-seq: Profiling Translationally Active tRNAs

The collaborative development of tRIBO-seq addressed a longstanding technical challenge: characterizing tRNA populations actively engaged in translation rather than relying on total cellular tRNA measurements that may not accurately reflect translational status, particularly under stress conditions. By combining RiboLace-based active ribosome isolation with nanopore direct RNA sequencing, tRIBO-seq enables simultaneous quantification of tRNA abundance, modification patterns, and fragmentation status from ribosome-associated populations in a single experiment requiring minimal input material (3-5 µg RNA, approximately 5 million cells) and completing within a single working day.

Application of tRIBO-seq across four distinct conditions revealed condition-specific tRNAome reprogramming mechanisms. Under steady-state conditions, ribosome-associated and total tRNA pools showed high similarity, indicating tight coordination between tRNA supply and translational demand. However, stress conditions revealed divergent responses. Amino acid deprivation (leucine starvation) induced changes primarily in ribo-tRNA abundances with enrichment of cognate Leu-tRNA isoacceptors, while methionine starvation caused widespread tRNA hypomethylation linked to S-adenosylmethionine depletion, with the ribo-tRNA pool showing reduced hypomethylation compared to total tRNAs—suggesting preferential loading of fully modified tRNAs onto ribosomes. Oxidative stress induced by arsenite exposure triggered both major ribo-tRNA abundance changes and selective tRNA fragmentation occurring predominantly in ribosome-embedded tRNAs, consistent with ribosome-localized endonuclease activation.

These findings demonstrate that ribosome-associated tRNAs respond distinctly from total tRNA pools under stress conditions, emphasizing the importance of studying translationally active tRNA populations to understand how cells coordinate tRNA availability

with translational demand. The ability to capture tRNA-derived fragments alongside full-length tRNAs provides a framework for investigating both canonical translation functions and emerging non-canonical roles of tRNA fragments in cellular signaling and stress responses.

3.2 Integration and Broader Impact

While developed independently, Martian and tRIBO-seq represent complementary approaches to a unified challenge: understanding how cells coordinate ribosome positioning, codon decoding, and tRNA availability to achieve precise translational control. The analytical framework established in Martian's Module 5 for integrating ribosome profiling with nano-tRNA sequencing data provides a foundation for future investigations linking codon-specific ribosome occupancy patterns with tRNA modification status and abundance. Such integrated analyses could reveal whether specific tRNA modifications influence ribosome dwell times at particular codons, whether stress-induced changes in tRNA pools correlate with altered translation elongation dynamics, and how cells balance tRNA supply with codon demand across different cellular states.

The broader impact of this work extends beyond the specific tools developed. By addressing fragmentation in ribosome profiling analysis workflows, improving accessibility through automated pipelines, and enabling streamlined profiling of translationally active tRNAs, these contributions lower barriers to comprehensive translational analysis. This democratization of analytical capabilities is particularly relevant as ribosome profiling transitions from a specialized research technique to a potential clinical diagnostic platform. The growing recognition of translation machinery components as therapeutic targets—including eIF4E inhibitors in acute myeloid leukemia and ribosome biogenesis factors as biomarkers for mTOR inhibitor response—underscores the clinical potential of robust, accessible tools for translational analysis.

3.3 Limitations and Considerations

The work presented in this thesis should be interpreted within the context of its scope and limitations. The Martian benchmarking evaluated performance across a limited dataset (11 samples from human cell lines), and performance characteristics on samples from different organisms, library preparation protocols, or biological conditions remain to be determined. Several potentially relevant metrics were not quantified, including reproducibility across biological replicates, sensitivity for detecting differential translation across varying sequencing depths, and computational scalability with increasing sample numbers. The stringent annotation filtering strategy, while improving analytical robustness, necessarily

excludes transcripts that may represent legitimate biology alongside annotation artifacts, a trade-off that warrants consideration when interpreting results from poorly annotated genomic regions.

For tRIBO-seq, performance will vary depending on cellular translational activity, with some cell types requiring substantially more input material than the typical 3-5 µg. The protocol has been tested exclusively in mammalian cell lines, and extension to other organisms may require optimization. The selective ligation strategy captures 3′ tRNA-derived fragments but not 5′ fragments, potentially limiting comprehensive tRNA fragmentomics analysis. Additionally, the reliance on nanopore sequencing means that modification detection accuracy depends on characteristic signal distortions, which may vary across modification types and sequence contexts.

3.4 Future Directions

Several directions for future development would enhance the capabilities and impact of the tools presented in this thesis.

3.4.1 Martian Development Priorities

Enhanced analytical capabilities. Incorporating machine learning approaches for P-site offset prediction could improve accuracy for challenging samples with heterogeneous fragment lengths or non-canonical ribosome occupancy patterns. Implementing adaptive database filtering that adjusts stringency based on data quality metrics could better balance annotation confidence with transcript coverage, potentially recovering legitimate biology from currently excluded transcripts while maintaining quality control.

Mature integration methodologies. Developing comprehensive statistical and computational frameworks for integrating ribosome profiling with nano-tRNA sequencing data would enable systematic analysis of how tRNA availability, modification status, and codon usage collectively influence translation elongation rates across the transcriptome. Such integration could reveal coordination mechanisms linking tRNA dynamics with ribosome behavior that are currently obscured by analyzing these data types independently.

Improved accessibility. Developing a graphical user interface would make Martian accessible to researchers without extensive bioinformatics expertise, while integration with cloud computing platforms could eliminate local hardware requirements, particularly addressing the 21 GB memory demands. Comprehensive documentation including video tutorials, example datasets, and detailed output specifications (available at `https://github.com/ImmaginaBiotechnology/Documents`) would facilitate broader adoption across diverse research communities.

Open source distribution. Plans to distribute Martian as free, open-source software under a permissive license would foster community-driven development and ensure reproducibility across laboratories. The modular architecture is designed to accommodate user contributions, enabling researchers to extend functionality for specialized applications while maintaining core pipeline stability. Publication of the source code repository with detailed developer documentation will encourage collaborative improvement of ribosome profiling analysis methodologies.

3.4.2 tRIBO-seq Applications and Extensions

Therapeutic applications. The speed, depth, and accessibility of tRIBO-seq provide opportunities for applications in therapeutic tRNA development. As tRNA-based therapeutics advance toward clinical evaluation [127, 128], methods enabling rapid assessment of therapeutic tRNA incorporation into actively translating ribosomes, modification status, and stability will become increasingly valuable. tRIBO-seq's capacity to simultaneously monitor tRNA abundance, modifications, and fragmentation from minimal input material positions it well for such applications.

Disease mechanisms. The identification of over 50 tRNA modification enzymes associated with human diseases, including upregulation correlating with poor survival in multiple tumor types, highlights opportunities for investigating tRNA-mediated disease mechanisms. Applying tRIBO-seq to disease models could reveal whether pathogenic mutations in tRNA modification enzymes alter ribosome-associated tRNA pools differently than total tRNA populations, potentially identifying new therapeutic intervention points.

Technical extensions. Adapting the protocol for non-mammalian systems would expand applicability across model organisms. Developing strategies to capture 5' tRNA fragments in addition to current 3' fragment detection would enable more comprehensive tRNA fragmentomics. Integrating tRIBO-seq with complementary approaches such as tRNA charging assays or aminoacyl-tRNA sequencing could provide additional dimensions of tRNA functional characterization.

3.4.3 Community-Level Challenges

Beyond tool-specific improvements, realizing the full potential of ribosome profiling and tRNA sequencing as research and clinical platforms will require addressing community-level challenges. The 47-fold variation in 3′UTR detection between tools, 40% peak detection reproducibility between replicates, and 2% agreement among ORF detection tools documented in this thesis and previous benchmarking studies reveal fundamental inconsistencies that undermine cross-study comparisons and meta-analyses. Addressing these challenges will require:

Standardized benchmarking. Development of community consensus datasets spanning diverse organisms, library preparation protocols, and biological conditions would enable systematic tool comparison and identification of best practices for specific applications.
Quality metrics. Definition of minimal quality thresholds for ribosome profiling experiments, analogous to established RNA-seq quality metrics, would facilitate objective assessment of dataset reliability and inform tool selection based on data characteristics.
Transparent reporting. Establishment of guidelines for reporting analytical choices, parameter settings, and quality control metrics would improve reproducibility and enable more meaningful comparison across studies.
Integration standards. Development of common data formats and metadata standards for multi-omics integration would reduce technical barriers to comprehensive translational analysis combining ribosome profiling, RNA sequencing, tRNA sequencing, and proteomics.

3.5 Closing Perspective

The central challenge motivating this thesis was the fragmentation of translational analysis workflows and the difficulty of coordinating measurements across different molecular layers, ribosomes, mRNAs, and tRNAs, that collectively determine protein synthesis. By developing an integrated computational framework for ribosome profiling analysis and contributing to methods for profiling translationally active tRNA populations, this work provides tools that reduce analytical barriers while maintaining scientific rigor. The practical demonstration of these tools in collaborative research projects validates their utility for addressing current biological questions while establishing a foundation for future methodological advances.
Translation regulation sits at the interface between genetic information and cellular function, integrating signals from multiple regulatory layers to achieve precise control over protein synthesis. As the field continues to uncover the complexity of this regulation, from codon-specific effects and tRNA modifications to ribosome heterogeneity and localized translation, the need for comprehensive, accessible analytical tools will only increase. The continued development of integrated platforms that connect experimental advances with robust computational analysis will be essential for translating the remarkable insights from ribosome profiling and tRNA sequencing into improved understanding of human biology and more effective therapeutic interventions for translation-related diseases.

Bibliography

1. Francis, C. Central Dogma of Molecular Biology. *Nature* **227,** 561–563. ISSN: 1476-4687 (1970).
2. Tahmasebi, S., Khoutorsky, A., Mathews, M. B. & Sonenberg, N. Translation Deregulation in Human Disease. *Nature Reviews Molecular Cell Biology* **19,** 791–807. ISSN: 1471-0080 (2018).
3. Liu, Y., Beyer, A. & Aebersold, R. On the dependency of cellular protein levels on mRNA abundance. *Cell* **165,** 535–550 (2016).
4. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics* **13,** 227–232 (2012).
5. Rodnina, M. V. The ribosome in action: Tuning of translational efficiency and protein folding. *Protein Science* **25,** 1390–1406 (2016).
6. Hershey, J. W., Sonenberg, N. & Mathews, M. B. Principles of Translational Control. *Cold Spring Harbor Perspectives in Biology* **11,** a032607. ISSN: 1943-0264 (2019).
7. Guertin, D. A. & Sabatini, D. M. Defining the Role of mTOR in Cancer. *Cancer Cell* **12,** 9–22. ISSN: 1535-6108 (2007).
8. Kapur, M., Monaghan, C. E. & Ackerman, S. L. Regulation of mRNA Translation in Neurons—A Matter of Life and Death. *Neuron* **96,** 616–637. ISSN: 08966273 (2017).
9. Assouline, S. *et al.* Molecular Targeting of the Oncogene eIF4E in Acute Myeloid Leukemia (AML): A Proof-of-Principle Clinical Trial with Ribavirin. *Blood* **114,** 257–260. ISSN: 0006-4971, 1528-0020 (2009).
10. Su, D. *et al.* Ribosome Profiling: A Powerful Tool in Oncological Research. *Biomarker Research* **12,** 11. ISSN: 2050-7771 (2024).

11. Mourksi, N.-E.-H. *et al.* Ribosome Biogenesis-based Predictive Biomarkers in Endocrine Therapy (Anastrozole) Combined with mTOR Inhibitor (Vistusertib) in Endometrial Cancer: Translational Study from the VICTORIA Trial in Collaboration with the GINECO Group. *Molecular Oncology* **17,** 27–36. ISSN: 1574-7891, 1878-0261 (2023).

12. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* **324,** 218–223. ISSN: 0036-8075, 1095-9203 (2009).

13. Santos, D. A., Shi, L., Tu, B. P. & Weissman, J. S. Cycloheximide Can Distort Measurements of mRNA Levels and Translation Efficiency. *Nucleic Acids Research* **47,** 4974–4985. ISSN: 0305-1048, 1362-4962 (2019).

14. Hussmann, J. A., Patchett, S., Johnson, A., Sawyer, S. & Press, W. H. Understanding Biases in Ribosome Profiling Experiments Reveals Signatures of Translation Dynamics in Yeast. *PLOS Genetics* **11** (ed Snyder, M.) e1005732. ISSN: 1553-7404 (2015).

15. Meindl, A. *et al.* A Rapid Protocol for Ribosome Profiling of Low Input Samples. *Nucleic Acids Research* **51,** e68–e68. ISSN: 0305-1048, 1362-4962 (2023).

16. Clamer, M. *et al.* Active Ribosome Profiling with RiboLace. *Cell Reports* **25,** 1097–1108.e5. ISSN: 22111247 (2018).

17. Del Piano, A. *et al.* Phospho-RNA Sequencing with circAID-p-seq. *Nucleic Acids Research* **50,** e23–e23. ISSN: 0305-1048, 1362-4962 (2022).

18. VanInsberghe, M., Van Den Berg, J., Andersson-Rolf, A., Clevers, H. & Van Oudenaarden, A. Single-Cell Ribo-seq Reveals Cell Cycle-Dependent Translational Pausing. *Nature* **597,** 561–565. ISSN: 0028-0836, 1476-4687 (2021).

19. Ozadam, H. *et al.* Single-Cell Quantification of Ribosome Occupancy in Early Mouse Development. *Nature* **618,** 1057–1064. ISSN: 0028-0836, 1476-4687 (2023).

20. Tomuro, K. *et al.* Calibrated Ribosome Profiling Assesses the Dynamics of Ribosomal Flux on Transcripts. *Nature Communications* **15,** 7061. ISSN: 2041-1723 (2024).

21. Schimmel, P. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nature Reviews Molecular Cell Biology* **19,** 45–58 (2018).

22. Phizicky, E. M. & Hopper, A. K. tRNA biology charges to the front. *Genes & Development* **24,** 1832–1860 (2010).

23. Berg, M. D. & Brandl, C. J. Transfer RNAs: diversity in form and function. *RNA Biology* **18,** 316–339 (2021).

24. Pinkard, O., McFarland, S., Sweet, T. & Coller, J. Quantitative tRNA-sequencing Uncovers Metazoan Tissue-Specific tRNA Regulation. *Nature Communications* **11,** 4104. ISSN: 2041-1723 (2020).

25. Novoa, E. M. & Ribas de Pouplana, L. Speeding with control: codon usage, tRNAs, and ribosomes. *Trends in Genetics* **28,** 574–581 (2012).

26. Pan, T. Modifications and functional genomics of human transfer RNA. *Cell Research* **28,** 395–404 (2018).

27. El Yacoubi, B., Bailly, M. & de Crécy-Lagard, V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annual Review of Genetics* **46,** 69–95 (2012).

28. Zhang, W., Foo, M., Eren, A. M. & Pan, T. tRNA modification dynamics from individual organisms to metaepitranscriptomics of microbiomes. *Molecular Cell* **82,** 891–906 (2022).

29. Suzuki, T. The expanding world of tRNA modifications and their disease relevance. *Nature Reviews Molecular Cell Biology* **22,** 375–392 (2021).

30. Chan, C. T. Y. *et al.* Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nature Communications* **3,** 937 (2012).

31. Zaborske, J. M. *et al.* A nutrient-driven tRNA modification alters translational fidelity and genome-wide protein coding across an animal genus. *PLoS Biology* **12,** e1002015 (2014).

32. Zinshteyn, B. & Gilbert, W. V. Loss of a conserved tRNA anticodon modification perturbs cellular signaling. *PLoS Genetics* **9,** e1003675 (2013).

33. Kirchner, S. & Ignatova, Z. Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nature Reviews Genetics* **16,** 98–112 (2015).

34. Chan, C. T. Y. *et al.* A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS Genetics* **6,** e1001247 (2010).

35. Watkins, C. P., Zhang, W., Wylder, A. C., Katanski, C. D. & Pan, T. A multiplex platform for small RNA sequencing elucidates multifaceted tRNA stress response and translational regulation. *Nature Communications* **13,** 2491 (2022).

36. Pang, Y. L. J., 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 Research* **42,** e170 (2014).

37. Berg, M., Li, C. & Kaiser, S. NAIL-MS reveals tRNA and rRNA hypomodification as a consequence of 5-fluorouracil treatment. *Nucleic Acids Research* **53** (2025).

38. Nedialkova, D. D. & Leidel, S. A. Optimization of Codon Translation Rates via tRNA Modifications Maintains Proteome Integrity. *Cell* **161,** 1606–1618. ISSN: 00928674 (2015).

39. Gu, C., Begley, T. J. & Dedon, P. C. tRNA modifications regulate translation during cellular stress. *FEBS Letters* **588,** 4287–4296 (2014).

40. Chujo, T. & Tomizawa, K. Human Transfer RNA Modopathies: Diseases Caused by Aberrations in Transfer RNA Modifications. *The FEBS Journal* **288,** 7096–7122. ISSN: 1742-464X, 1742-4658 (2021).

41. Zhou, M. *et al.* tRNA-derived Small RNAs in Human Cancers: Roles, Mechanisms, and Clinical Application. *Molecular Cancer* **23,** 76. ISSN: 1476-4598 (2024).

42. Thüring, K., Schmid, K., Keller, P. & Helm, M. Analysis of RNA modifications by liquid chromatography-tandem mass spectrometry. *Methods* **107,** 48–56 (2016).

43. Cozen, A. E. *et al.* ARM-seq: AlkB-facilitated RNA Methylation Sequencing Reveals a Complex Landscape of Modified tRNA Fragments. *Nature Methods* **12,** 879–884. ISSN: 1548-7091, 1548-7105 (2015).

44. Gogakos, T. *et al.* Characterizing Expression and Processing of Precursor and Mature Human tRNAs by Hydro-tRNAseq and PAR-CLIP. *Cell Reports* **20,** 1463–1475. ISSN: 22111247 (2017).

45. Hauenschild, R. *et al.* The Reverse Transcription Signature of *N* -1-Methyladenosine in RNA-Seq Is Sequence Dependent. *Nucleic Acids Research,* gkv895. ISSN: 0305-1048, 1362-4962 (2015).

46. Zheng, G. *et al.* Efficient and Quantitative High-Throughput tRNA Sequencing. *Nature Methods* **12,** 835–837. ISSN: 1548-7091, 1548-7105 (2015).

47. Lin, S., Liu, Q., Jiang, Y.-Z. & Gregory, R. I. Nucleotide Resolution Profiling of m7G tRNA Modification by TRAC-Seq. *Nature Protocols* **14,** 3220–3242. ISSN: 1754-2189, 1750-2799 (2019).

48. Behrens, A., Rodschinka, G. & Nedialkova, D. D. High-Resolution Quantitative Profiling of tRNA Abundance and Modification Status in Eukaryotes by Mim-tRNAseq. *Molecular Cell* **81,** 1802–1815.e7. ISSN: 10972765 (2021).

49. Zhang, W., Xu, R., Matuszek, Ż., Cai, Z. & Pan, T. Detection and quantification of glycosylated queuosine modified tRNAs by acid denaturing and APB gels. *RNA* **26,** 1291–1298 (2020).

50. Hernandez-Alias, X. *et al.* Single-read tRNA-seq analysis reveals coordination of tRNA modification and aminoacylation and fragmentation. *Nucleic Acids Research* **51,** e17 (2023).

51. Cui, J., Liu, Q., Sendinc, E., Shi, Y. & Gregory, R. I. Nucleotide resolution profiling of m3C RNA modification by HAC-seq. *Nucleic Acids Research* **49,** e27 (2021).

52. Marchand, V., Bourguignon-Igel, V., Helm, M. & Motorin, Y. Mapping of 7-methylguanosine (m7G), 3-methylcytidine (m3C), dihydrouridine (D) and 5-hydroxycytidine (ho5C) RNA modifications by AlkAniline-Seq. *Methods in Enzymology* **658,** 25–47 (2021).

53. Marchand, V. *et al.* HydraPsiSeq: a method for systematic and quantitative mapping of pseudouridines in RNA. *Nucleic Acids Research* **48,** e110 (2020).

54. Lucas, M. C. *et al.* Quantitative Analysis of tRNA Abundance and Modifications by Nanopore RNA Sequencing. *Nature Biotechnology* **42,** 72–86. ISSN: 1546-1696 (2024).

55. Smith, A. M., Abu-Shumays, R., Akeson, M. & Bernick, D. L. Capture, Unfolding, and Detection of Individual tRNA Molecules Using a Nanopore Device. *Frontiers in Bioengineering and Biotechnology* **3,** 91 (2015).

56. Thomas, N. K. *et al.* Direct Nanopore Sequencing of Individual Full Length tRNA Strands. *ACS Nano* **15,** 16642–16653 (2021).

57. Shaw, E. A. *et al.* Combining Nanopore direct RNA sequencing with genetics and mass spectrometry for analysis of T-loop base modifications across 42 yeast tRNA isoacceptors. *Nucleic Acids Research* **52,** 12074–12092 (2024).

58. White, L. K., Strugar, S. M., MacFadden, A. & Hesselberth, J. R. Nanopore sequencing of internal 2'-PO4 modifications installed by RNA repair. *RNA* **29,** 847–861 (2023).

59. White, L. K. *et al.* Nanopore Sequencing of Intact Aminoacylated tRNAs. *Nature Communications* **16,** 7781. ISSN: 2041-1723 (2025).

60. Lauria, F. *et al.* riboWaltz: Optimization of ribosome P-site positioning in ribosome profiling data. *PLOS Computational Biology* **14,** 1–20 (2018).

61. Chothani, S. *et al.* deltaTE: Detection of Translationally Regulated Genes by Integrative Analysis of Ribo-seq and RNA-seq Data. *Current Protocols in Molecular Biology* **129,** e108. ISSN: 1934-3639, 1934-3647 (2019).

62. Dana, A. & Tuller, T. Mean of the Typical Decoding Rates: A New Translation Efficiency Index Based on the Analysis of Ribosome Profiling Data. *G3 Genes|Genomes|Genetics* **5,** 73–80. ISSN: 2160-1836 (2015).

63. Iosub, I. A., Wilkins, O. G. & Ule, J. Riboseq-Flow: A Streamlined, Reliable Pipeline for Ribosome Profiling Data Analysis and Quality Control. *Wellcome Open Research* **9,** 179. ISSN: 2398-502X (2024).

64. He, J. *et al.* Deep Learning Prediction of Ribosome Profiling with Translatomer Reveals Translational Regulation and Interprets Disease Variants. *Nature Machine Intelligence* **6,** 1314–1329. ISSN: 2522-5839 (2024).

65. Shao, B. *et al.* Riboformer: A Deep Learning Framework for Predicting Context-Dependent Translation Dynamics. *Nature Communications* **15,** 2011. ISSN: 2041-1723 (2024).

66. White, L. K. *et al. Comparative Analysis of 43 Distinct RNA Modifications by Nanopore tRNA Sequencing* pre-published.

67. Padhiar, N. H., Katneni, U., Komar, A. A., Motorin, Y. & Kimchi-Sarfaty, C. Advances in Methods for tRNA Sequencing and Quantification. *Trends in Genetics* **40,** 276–290. ISSN: 01689525 (2024).

68. Wang, Y., Tang, Y., Xie, Z. & Wang, H. RPFdb v3.0: An Enhanced Repository for Ribosome Profiling Data and Related Content. *Nucleic Acids Research* **53,** D293–D298. ISSN: 0305-1048, 1362-4962 (2025).

69. Tierney, J. A. S. *et al.* RiboSeq.Org: An Integrated Suite of Resources for Ribosome Profiling Data Analysis and Visualization. *Nucleic Acids Research* **53,** D268–D274. ISSN: 0305-1048, 1362-4962 (2025).

70. Cope, A. L. *et al.* Riboviz 2: A Flexible and Robust Ribosome Profiling Data Analysis and Visualization Workflow. *Bioinformatics* **38** (ed Valencia, A.) 2358–2360. ISSN: 1367-4803, 1367-4811 (2022).

71. Legrand, C. & Tuorto, F. RiboVIEW: A Computational Framework for Visualization, Quality Control and Statistical Analysis of Ribosome Profiling Data. *Nucleic Acids Research* **48,** e7–e7. ISSN: 0305-1048, 1362-4962 (2020).

72. Gelhausen, R. *et al.* RiboReport - Benchmarking Tools for Ribosome Profiling-Based Identification of Open Reading Frames in Bacteria. *Briefings in Bioinformatics* **23,** bbab549. ISSN: 1467-5463, 1477-4054 (2022).

73. Diament, A. & Tuller, T. Estimation of Ribosome Profiling Performance and Reproducibility at Various Levels of Resolution. *Biology Direct* **11,** 24. ISSN: 1745-6150 (2016).

74. Tong, G., Hah, N. & Martinez, T. F. Comparison of Software Packages for Detecting Unannotated Translated Small Open Reading Frames by Ribo-seq. *Briefings in Bioinformatics* **25,** bbae268. ISSN: 1467-5463, 1477-4054 (2024).

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

76. Pavon-Eternod, M. *et al.* Vaccinia and influenza A viruses select rather than adjust tRNAs to optimize translation. *Nucleic Acids Research* **41,** 1914–1921 (2013).

77. Chen, C.-W. & Tanaka, M. Genome-wide translation profiling by ribosome-bound tRNA capture. *Cell Reports* **23,** 608–621 (2018).

78. Hasin, Y., Seldin, M. & Lusis, A. Multi-Omics Approaches to Disease. *Genome Biology* **18,** 83. ISSN: 1474-760X (2017).

79. Yu, Y., Mai, Y., Zheng, Y. & Shi, L. Assessing and Mitigating Batch Effects in Large-Scale Omics Studies. *Genome Biology* **25,** 254. ISSN: 1474-760X (2024).

80. Xiao, Z., Zou, Q., Liu, Y. & Yang, X. Genome-Wide Assessment of Differential Translations with Ribosome Profiling Data. *Nature Communications* **7,** 11194. ISSN: 2041-1723 (2016).

81. Zhong, Y. *et al.* RiboDiff: Detecting Changes of mRNA Translation Efficiency from Ribosome Footprints. *Bioinformatics* **33** (ed Birol, I.) 139–141. ISSN: 1367-4803, 1367-4811 (2017).

82. Pickholz Berliner, Y. & Sharan, R. PEANUT: Pathway Enrichment Analysis through Network UTilization. *Bioinformatics* **41** (ed Alkan, C.) btaf410. ISSN: 1367-4811 (2025).

83. Reimand, J. *et al.* Pathway Enrichment Analysis and Visualization of Omics Data Using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nature Protocols* **14,** 482–517. ISSN: 1754-2189, 1750-2799 (2019).

84. Kanwal, S., Khan, F. Z., Lonie, A. & Sinnott, R. O. Investigating Reproducibility and Tracking Provenance – A Genomic Workflow Case Study. *BMC Bioinformatics* **18,** 337. ISSN: 1471-2105 (2017).

85. Grüning, B. *et al.* Practical Computational Reproducibility in the Life Sciences. *Cell Systems* **6,** 631–635. ISSN: 24054712 (2018).

86. Loveland, A. B., Koh, C. S., Ganesan, R., Jacobson, A. & Korostelev, A. A. Structural mechanism of angiogenin activation by the ribosome. *Nature* **630,** 769–776 (2024).

87. Rodriguez, J. M. *et al.* APPRIS: Selecting Functionally Important Isoforms. *Nucleic Acids Research* **50,** D54–D59. ISSN: 0305-1048 (2021).

88. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLOS Computational Biology* **9,** 1–10 (2013).

89. Ross, G. J. S. & Preece, D. A. The Negative Binomial Distribution. *Journal of the Royal Statistical Society. Series D (The Statistician)* **34,** 323–335. ISSN: 00390526, 14679884 (1985).

90. Wald, A. Sequential Tests of Statistical Hypotheses. *The Annals of Mathematical Statistics* **16,** 117–186. ISSN: 00034851 (1945).

91. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57,** 289–300. ISSN: 0035-9246 (1995).

92. Kolberg, L. *et al.* G:Profiler—Interoperable Web Service for Functional Enrichment Analysis and Gene Identifier Mapping (2023 Update). *Nucleic Acids Research* **51,** W207–W212. ISSN: 0305-1048 (2023).

93. Sprinzl, M., Steegborn, C., Hübel, F. & Steinberg, S. Compilation of tRNA Sequences and Sequences of tRNA Genes. *Nucleic Acids Research* **24,** 68–72. ISSN: 0305-1048 (1996).

94. Heinemann, I. U., Nakamura, A., O'Donoghue, P., Eiler, D. & Söll, D. tRNAHis-guanylyltransferase Establishes tRNAHis Identity. *Nucleic Acids Research* **40,** 333–344. ISSN: 0305-1048 (2011).

95. Seetharaman, M., Williams, C., Cramer, C. J. & Musier-Forsyth, K. Effect of G-1 on Histidine tRNA Microhelix Conformation. *Nucleic Acids Research* **31,** 7311–7321. ISSN: 0305-1048 (2003).

96. Begik, O. *et al.* Quantitative Profiling of Pseudouridylation Dynamics in Native RNAs with Nanopore Sequencing. *Nature Biotechnology* **39,** 1278–1291. ISSN: 1546-1696 (2021).

97. Ichihara, K. *et al.* Combinatorial Analysis of Translation Dynamics Reveals eIF2 Dependence of Translation Initiation at Near-Cognate Codons. *Nucleic Acids Research* **49,** 7298–7317. ISSN: 0305-1048 (2021).

98. Wangen, J. R. & Green, R. Stop Codon Context Influences Genome-Wide Stimulation of Termination Codon Readthrough by Aminoglycosides. *eLife* **9** (eds Sonenberg, N., Manley, J. L., Bedwell, D., Puglisi, J. D. & Weissman, J. S.) e52611. ISSN: 2050-084X (2020).

99. Dmitriev, S. E., Vladimirov, D. O. & Lashkevich, K. A. A quick guide to small-molecule inhibitors of eukaryotic protein synthesis. *Biochemistry (Moscow)* **85,** 1389–1421 (2020).

100. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34,** 3094–3100 (2018).

101. Pryszcz, L. P. *et al.* Rapid and accurate demultiplexing of direct RNA nanopore sequencing data with SeqTagger. *Genome Research* **35,** 956–966 (2025).

102. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15,** 550 (2014).

103. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32,** 2847–2849 (2016).

104. Percudani, R., Pavesi, A. & Ottonello, S. Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. *Journal of Molecular Biology* **268,** 322–330 (1997).

105. Schmidt, E. K., Clavarino, G., Ceppi, M. & Pierre, P. SUnSET, a nonradioactive method to monitor protein synthesis. *Nature Methods* **6,** 275–277 (2009).

106. Chen, Y., Chen, L., Lun, A. T. L., Baldoni, P. L. & Smyth, G. K. edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Research* **53** (2025).

107. Yurekten, O. *et al.* MetaboLights: open data repository for metabolomics. *Nucleic Acids Research* **52,** D640–D646 (2024).

108. Wang, J. *et al.* Structural basis for the transition from translation initiation to elongation by an 80S-eIF5B complex. *Nature Communications* **11,** 5003 (2020).

109. Lobanov, A. V., Hatfield, D. L. & Gladyshev, V. N. Eukaryotic selenoproteins and selenoproteomes. *Biochimica et Biophysica Acta* **1790,** 1424–1428 (2009).

110. Czech, A. Deep sequencing of tRNA's 3'-termini sheds light on CCA-tail integrity and maturation. *RNA* **26,** 199–208 (2020).

111. Eisenberg, A. R. *et al.* Translation initiation site profiling reveals widespread synthesis of non-AUG-initiated protein isoforms in yeast. *Cell Systems* **11,** 145–160.e5 (2020).

112. Darnell, A. M., Subramaniam, A. R. & O'Shea, E. K. Translational control through differential ribosome pausing during amino acid limitation in mammalian cells. *Molecular Cell* **71,** 229–243.e11 (2018).

113. Mazor, K. M. *et al.* Effects of single amino acid deficiency on mRNA translation are markedly different for methionine versus leucine. *Scientific Reports* **8** (2018).

114. Tang, X. *et al.* Comprehensive profiling of amino acid response uncovers unique methionine-deprived response dependent on intact creatine biosynthesis. *PLoS Genetics* **11,** e1005158 (2015).

115. Cappannini, A. *et al.* MODOMICS: a database of RNA modifications and related information. 2023 update. *Nucleic Acids Research* **52,** D239–D244 (2024).

116. Pizzinga, M. *et al.* The cell stress response: extreme times call for post-transcriptional measures. *Wiley Interdisciplinary Reviews RNA* **11,** e1578 (2020).

117. Andreev, D. E. *et al.* Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *eLife* **4,** e03971 (2015).

118. Cnubben, N. H., Rietjens, I. M., Wortelboer, H., van Zanden, J. & van Bladeren, P. J. The interplay of glutathione-related processes in antioxidant defense. *Environmental Toxicology and Pharmacology* **10,** 141–152 (2001).

119. Su, Z., Kuscu, C., Malik, A., Shibata, E. & Dutta, A. Angiogenin generates specific stress-induced tRNA halves and is not involved in tRF-3-mediated gene silencing. *Journal of Biological Chemistry* **294,** 16930–16941 (2019).

120. Sanadgol, N., König, L., Drino, A., Jovic, M. & Schaefer, M. R. Experimental paradigms revisited: oxidative stress-induced tRNA fragmentation does not correlate with stress granule formation but is associated with delayed cell death. *Nucleic Acids Research* **50,** 6919–6937 (2022).

121. Saikia, M. *et al.* Genome-wide identification and quantitative analysis of cleaved tRNA fragments induced by cellular stress. *Journal of Biological Chemistry* **287,** 42708–42725 (2012).

122. Yan, B., Tzertzinis, G., Schildkraut, I. & Ettwiller, L. Comprehensive determination of transcription start sites derived from all RNA polymerases using ReCappable-seq. *Genome Research* **32,** 162–174 (2022).

123. Mulroney, L. *et al.* Identification of high-confidence human poly(A) RNA isoform scaffolds using nanopore sequencing. *RNA* **28,** 162–176 (2022).

124. Jungfleisch, J. *et al.* CHIKV infection reprograms codon optimality to favor viral RNA translation by altering the tRNA epitranscriptome. *Nature Communications* **13,** 4725 (2022).

125. Li, S. & Mason, C. E. The pivotal regulatory landscape of RNA modifications. *Annual Review of Genomics and Human Genetics* **15,** 127–150 (2014).

126. Miluzio, A. *et al.* Recycling of ribosomes at stop codons drives the rate of translation and the transition from proliferation to REStart. *Molecular Cell.* ISSN: 1097-2765 (2025).

127. Coller, J. & Ignatova, Z. tRNA therapeutics for genetic diseases. *Nature Reviews Drug Discovery* **23,** 108–125 (2024).

128. Anastassiadis, T. & Köhrer, C. Ushering in the era of tRNA medicines. *Journal of Biological Chemistry* **299,** 105246 (2023).