

**CHARACTERIZING HUMAN TRANSFER RNAs BY HYDRO-TRNASEQ AND
PAR-CLIP**

A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfillment of the Requirements for
the degree of Doctor of Philosophy

by
Tasos Gogakos

June 2017

©Copyright by Tasos Gogakos 2017

Abstract

CHARACTERIZING HUMAN TRANSFER RNAs BY HYDRO-TRNASEQ AND PAR-CLIP

Tasos Gogakos, Ph.D.

The Rockefeller University 2017

The participation of tRNAs in fundamental aspects of biology and disease necessitates an accurate, experimentally confirmed annotation of tRNA genes, and curation of precursor and mature tRNA sequences. This has been challenging, mainly because RNA secondary structure and nucleotide modifications, together with tRNA gene multiplicity, complicate sequencing and read mapping efforts. To address these issues, I developed hydro-tRNAseq, a method based on partial alkaline RNA hydrolysis that generates fragments amenable for sequencing. To identify transcribed tRNA genes, I further complemented this approach with Photoactivatable Crosslinking and Immunoprecipitation (PAR-CLIP) of SSB/La, a conserved protein involved in pre-tRNA processing. My results show that approximately half of all predicted tRNA genes are transcribed in human cells, suggesting that the tRNA genomic space is more contracted than previously thought as a result of regulated expression. I also report predominant nucleotide modification sites, their order of incorporation, and identify tRNA leader, trailer and intron sequences. By using complementary sequencing-based methodologies I present a human tRNA atlas, and determine expression levels of mature and processing intermediates of tRNAs in human cells.

The technical advances provide by hydro-tRNAseq are applied towards the molecular diagnosis of a genetic neurodevelopmental syndrome, caused by mutations in the tRNA processing factor, CLP1. Since then, it has also been widely

used on multiple other fronts, some which are outlined in the appendix of this thesis.

Finally, I harness this novel experimental and computational expertise towards the identification of the endonuclease complex C3PO as a novel processing factor of human tRNAs. I carry out a transcriptome-wide analysis of C3PO targets, identify its binding sites and motifs, and provide insights into its biochemical and biological functions.

Acknowledgments

First, I would like to thank my

contents

Table of Contents

List of Figures	ix
List of Tables	xi
List of Abbreviations	xiii
List of Terms	xiv
1 Introduction	1
1.1 Overview	1
1.1.1 tRNA and disease	3
1.1.2 tRNA biogenesis	3
1.1.3 tRNA sequencing	5
1.1.4 Previous efforts for genome-wide tRNA annotation	7
1.1.5 Small RNA sequencing protocol	8
2 Results	11
2.1 Hydrolysis-based tRNA sequencing	11
2.2 Hierarchical sequence read mapping	12
2.3 Iterative manual tRNA curation	14
2.4 Protocol and pipeline outputs	15

2.4.1	Mature tRNA alignment	15
2.4.2	Pre-tRNA alignment	16
2.5	Justification for using precursor reads	17
2.6	Composition of hydro-tRNAseq libraries	18
2.7	Need for pre-tRNA enrichment	18
2.8	PAR-CLIP methodology for the study of RNA-RBP interactions	19
2.9	SSB PAR-CLIP	20
2.10	tRNA gene annotation	22
2.11	Comparison with other methods	24
2.12	Applications and biological insights	26
2.13	Mature tRNA abundance does not correlate with pre-tRNA abundance	28
2.14	tRNA transcription initiation and termination	29
2.15	Ribonucleotide modifications	32
2.16	Annotation of intron-containing tRNA genes	35
2.17	CLP1	36
2.18	CLP1 figures	37
2.18.1	Plausible pathomechanisms of CLP1 mutations	37
2.18.2	hydro-tRNAseq on CLP1	39
2.19	tRNA enzyme screen	41
2.20	Discussion	42
2.21	Summary	45
3	Materials and methods	46
3.1	Hydro-tRNAseq	46
3.2	SSB PAR-CLIP	47
3.3	Bioinformatic analysis	48

3.4	Accession codes	49
4	C3PO	50
4.1	Introduction	50
4.2	From many	53
4.3	TSN binds tRNAs	55
4.4	C3PO possesses a length- and structure-dependent endonucleolytic activity	56
4.5	Biochemical characterization of C3PO's tRNA processing activity . .	56
4.6	Functional validation of C3PO's targets	56
4.7	C3PO summary report 2014	57
4.8	C3PO summary from annual report 2015	59
4.9	C3PO summary from annual report 2016	60
5	Things that didn't work	63
	References	83

List of Figures

1.1 tRNA structure	3
1.2 tRNA biogenesis	5
1.3 Small RNA sequencing protocol	9
2.1 Experimental and bioinformatic pipeline for tRNA annotation and reference transcript curation by hydro-tRNAsq	13
2.2 Mature tRNA alignment	16
2.3 Pre-tRNA alignment	17
2.4 Information entropy in pre-tRNA segments and mature body	17
2.5 Composition of hydro-tRNAsq libraries	18
2.6 SSB crosslinking to RNA	19
2.7 PAR-CLIP	21
2.8 figure2cd	22
2.9 figure2ef	22
2.10 figure3	23
2.11 figure4	24
2.12 supp6	25
2.13 supp7	26
2.14 figure4A	27
2.15 figure4D	28

2.16 figure4Drot	29
2.17 supp4	30
2.18 clp1 bar	30
2.19 figure6	31
2.20 figure6de	31
2.21 supp5	32
2.22 paper7a	33
2.23 paper7b	35
2.24 figure5	37
2.25 intrno-containing tRNA	38
2.26 clp1 alignments	38
2.27 clp1 bar	39
2.28 clp1 bar	39
2.29 NB	40
2.30 splicing	40
4.1 parclips	53

List of Tables

1.1 RNA types recovered by small RNA sequencing protocol	10
--	----

Glossary

List of Abbreviations

ChIP-seq	chromatin immunoprecipitation sequencing.
ncRNA	noncoding RNA.
OH	hydroxyl.
POLR3	human RNA polymerase III.
pre-tRNA	precursor tRNA.
RNP	ribonucleoprotein.
RT	reverse transcriptase.
tRNA	transfer RNA.

List of Terms

tRNA isoacceptor tRNA molecules that decode synonymous codons.

Chapter 1

Introduction

1.1 Overview

transfer RNAs (tRNAs) are essential factors for the expression of genetic information, serving as the adaptor molecules that decode the genetic code during protein synthesis [1], and are among the earliest studied noncoding RNA (ncRNA) non-coding RNA molecules [2, 3]. The biological importance of tRNAs and their associated proteins is underscored by the pathologic conditions that are related to aberrations in their expression and function [4–7]. Despite their highly conserved participation in the translational machinery, tRNAs have received new attention in recent years in the context of codon-resolved translational control [8–13], and due to the involvement of their metabolic byproducts in regulation and cross-talk with processing and effector functions of other classes of non-coding RNAs (ncRNAs) [14–18].

Nevertheless, the lack of reliable methods for tRNA quantification has hampered such analyses, and necessitated the use of predicted tRNA gene copy number as a surrogate index of expression [12, 19, 20]. This hinged on the

assumption that predicted tRNA gene loci are all expressed constitutively and equally, even though there has been experimental evidence against it [21]. Similarly, experimental tRNA gene annotation in the past had to focus on human RNA polymerase III (POLR3) chromatin immunoprecipitation sequencing (ChIP-seq) [22–24] or hybridization-based approaches [25, 26]. The former, however, were impeded by their restricted genomic resolution and the assumption that POLR3 binding always leads to productive tRNA expression followed by complete processing, while the latter fell short of providing absolute counts and did not address the discovery of new transcripts and genes, assuming also normal hybridization rules for modified nucleosides.

An improvement in tRNA quantification has arisen from recent efforts that employed modification-reverting enzymes prior to sequencing, in order to minimize stalling of reverse transcriptase at modified sites [27, 28]. However, an extensive annotation of human genes and transcripts was foregone because the focus was either on mature tRNAs only [28] or on tRNA fragments not inclusive of full-length precursor precursor tRNA (pre-tRNA) transcripts [27]. Thus, to-date an experimentally validated list of curated mature and pre-tRNA sequences and annotating tRNA genes in human is still missing.

To address this lack of experimentally-validated tRNA reference, I combined complementary high-throughput techniques for obtaining the sequence composition and abundance of tRNAs in human embryonic kidney cells (HEK293). First, I developed hydro-tRNAseq, a modified small RNA sequencing protocol based on partial alkaline hydrolysis of input RNA that succeeded in identifying and quantifying tRNAs.

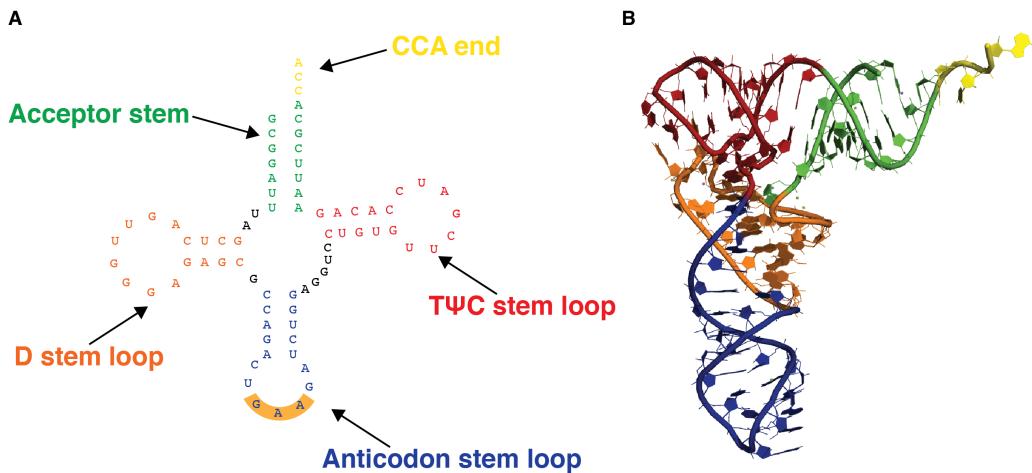


Figure 1.1: tRNA structure. (A) tRNA transcripts, such as the phenylalanine tRNA shown here, adopt the typical "cloverleaf" secondary, which in turns adopts an L-shaped tertiary structure as shown in (B). The structurally conserved stems and stemloops are indicated in A, color-coded, and their coordinates are reflected in the 3-dimensional structure in B (PDB 1EHZ).

1.1.1 tRNA and disease

talk about javier's papers, schimmel and dreyfuss review

1.1.2 tRNA biogenesis

tRNA genes are transcribed by POLR3 that uses promoters internal to the DNA sequence of the tRNA gene (tDNA), resulting in a primary transcript with a 5' triphosphate. In humans, a minority of tRNA transcripts harbor introns (see section refintrons). A dedicated tRNA splicing complex composed of core and accessory proteins carries out tRNA splicing [29–33]. Pre-tRNAs comprise the mature tRNA sequence, and 5' leader and 3' trailer extensions, which are trimmed in a coordinated manner by endonucleases and other processing factors. The ribonucleoprotein (RNP) complex RNase P removes the 5' leaders, leaving a 5' monophosphate, and ELAC2, the human homolog of tRNase Z trims the 3' trailer,

leaving a 3' hydroxyl (OH). Next, the universally conserved 3' terminal CCA tail is added by the tRNA nucleotidyl transferase 1 (TRNT1), and acts as the acceptor of the amino acid. tRNAs are further modified by chemical nucleotide modifications (2.15), exported from the nucleus to the cytoplasm where they can undergo further modifications, are aminoacylated with their cognate amino acid by aminoacyl tRNA synthetases, and are finally presented to the ribosome by translation factors to participate in protein synthesis (**Fig. 1.2**)[34–36].

Although these processes allow for multiple levels of regulation, variation in tRNA expression across tissues or between normal and pathologic conditions has not been studied extensively, mainly for two reasons. First, until recently there was the assumption that their essentiality obviated a need for any specialized transcriptional or post-transcriptional control. Second, the lack of an extensively curated and experimentally validated tRNA profile prevented quantitative and systematic studies. Nevertheless, it is now clear that the expression of tRNAs can be dynamic and can indeed exhibit tissue specificity [21, 37]. Importantly, abnormal tRNA expression levels have been correlated and causally associated with pathologic conditions, such as cancer [21, 26].

Overview of tRNA expression

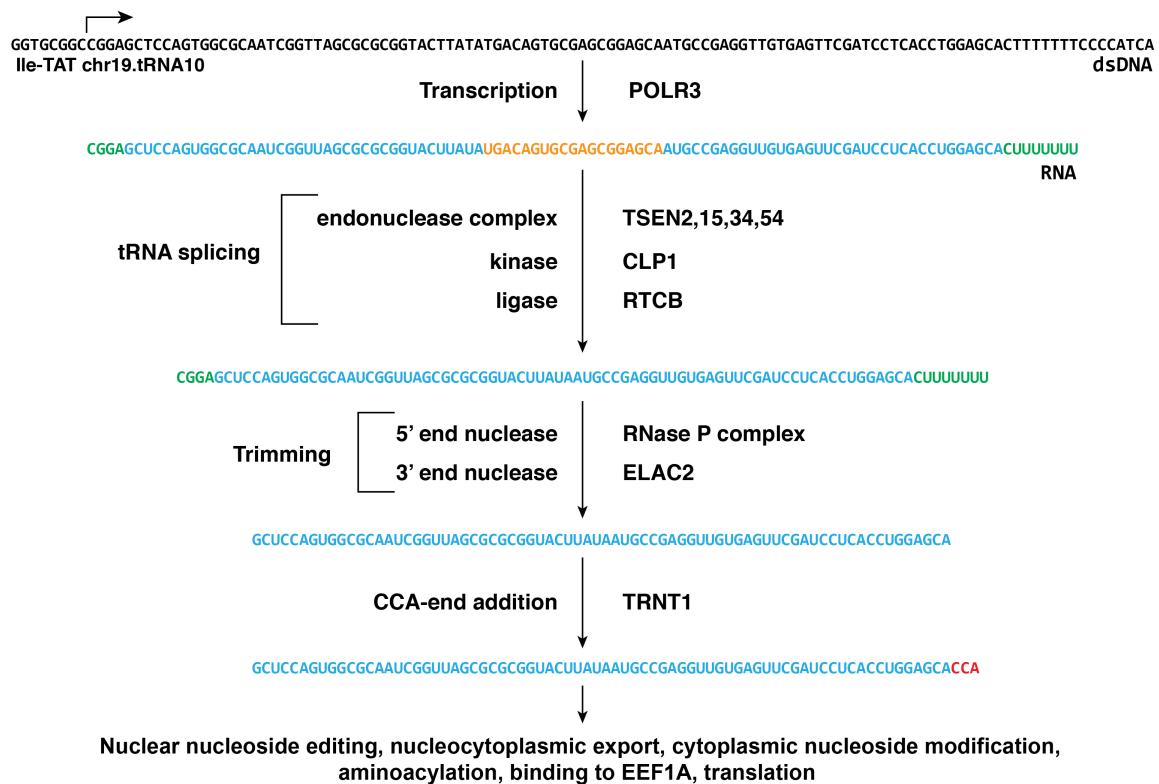


Figure 1.2: Overview of tRNA biogenesis and processing. tRNAs are transcribed by POLR3. If present, tRNA introns are removed by the tRNA splicing complex, and mature halves are ligated by the tRNA ligase (RTCB). Pre-tRNA leaders are trimmed by the RNase P complex, and 3' trailer by ELAC2. The 3' terminal CCA tail is added by TRNT1. tRNAs are further modified by nucleoside editing in the nucleus and the cytoplasm, are aminoacylated by cognate tRNA synthetases and are presented to the ribosome by translation factors.

1.1.3 tRNA sequencing

This complex biogenesis and processing pathway adds multiple layers of difficulty to the analysis of tRNAs. Obtaining data for tRNAs is hindered by multiple obstacles:

- i) sequencing of tRNAs is technically arduous due to their relatively small size, and their stable structure that impedes enzymes used in cDNA library prepa-

- rations, such as RNA ligases and reverse transcriptase (RT)
- ii) numerous (>100) tRNA pseudogenes are interspersed in the human genome [38, 39]
 - iii) all tRNAs undergo extensive post-transcriptional processing (see 1.2, while some involve extra processing steps (intron removal, addition of a 5' guanosine to all histidine tRNAs [40]))
 - iv) tRNAs are subjected to extensive chemical modifications on numerous nucleosides, which lead to mismatches upon the reverse transcription step of the RNA cloning protocols [41, 42]. Some modifications are universally conserved and required for proper tRNA function (e.g. adenosine to inosine deamination at the wobble position of the anticodon and methylation of adenosine in the T Ψ C loop) [41, 43]. Since alignment algorithms cannot tolerate multiple mismatches, it is likely that significant numbers of tRNA reads are excluded even if non-default mapping parameters are used.
 - v) tRNA isoacceptors share a large degree of sequence similarity that makes the distinction between alternative isoacceptors and editing products equivocal.
 - vi) eukaryotic cells harbor two distinct populations of tRNAs, nuclear and mitochondrial, whose length, structure, genomic organization, and processing differ considerably, and thus call for customized annotation procedures.

Owing to all these hurdles, the normal genetic makeup and variation of the tRNA population in human cells has not been probed adequately with RNA sequencing (RNA-seq) tools. Instead information about tRNA sequences and genes comes from bioinformatic predictions [38, 39]. Such approaches take into account base-pair covariation, secondary structure predictions of the classical cloverleaf

fold of tRNAs, and the tRNA promoter and termination architecture, and scan the human genome in order to identify sequences that are likely to obtain the typical tRNA structure. These analyses have resulted in the most comprehensive standard for whole-genome, predictive annotation of tRNAs so far, and the sequences they have predicted have been used extensively as bona fide tRNAs.

1.1.4 Previous efforts for genome-wide tRNA annotation

Even though no direct and rigorous experimental validation of tRNA sequences has been carried out, there has been indirect experimental evidence for tRNA expression:

- i) ChIP-seq studies focusing on the occupancy of genomic locations by POLR3 and/or its transcription factors [22–24]
- ii) tRNA microarrays that use the predicted tRNA sequences as the reference for the creation of array probes [25]

These methods, though, have several limitations. ChIP-seq, for example, uses chromatin occupancy as a proxy for productive RNA synthesis. Conversely, tRNA microarrays have limited sensitivity and specificity thresholds due to off-target hybridization that is potentiated by nucleoside modifications²¹, while their dynamic range is considerably narrower than RNA-seq. Finally, neither method is appropriately equipped to determine definitively pre-tRNAs or their transcription start and termination sites. This is an important limitation, as pre-tRNA fragments have been associated with neurodegenerative diseases [**Hanada:2013bk**, **Weitzer:2014bi**, **Karaca:2014em**].

1.1.5 Small RNA sequencing protocol

In order to obtain RNA-seq reads, I decided to first apply the well established protocol for sequencing small RNAs, established in the Tuschl lab [44] (**Fig. 1.3**). The experimental procedure takes advantage of the 5' monophosphate (p) and 3' OH groups present in small RNAs, such as micro RNAs (miRNAs), in order to enrich for such RNA species over other abundant RNA molecules. The use of a truncated and mutated RNA ligase (T4 Rnl2(1-249)K227Q) that requires 5' preadenylated adapter (App-(5')-adapter) prevents on one hand the formation of secondary, circularized byproducts, and allows on the other for exclusive ligation at the 3' end of the target RNA. Rnl1 is used to ligate an adapter with a different sequence at the 5' end, by activating the 5' monophosphate of the small RNA. cDNA is obtained by RT and amplified at non-saturated levels by PCR. The derived small RNA cDNA library is submitted to high-throughput on Illumina instruments using sequencing primer sites present in the adapter sequences. The different sequences of the 3' and 5' adapters preserves the strandedness of the original RNA sequence, enhancing ncRNA discovery and curation. Adding short (5-nt) barcode sequences at the 3' adapter also allows for pooling of several multiplexed samples, reducing costs, processing time, and batch variability.

Even though the utility of this protocol has been documented for the discovery and quantification of miRNAs, it was reasonable to apply towards tRNA sequencing because:

1. tRNAs, which are on average 75 nucleotides (nts) long, are closer in length than most other highly abundant ncRNAs (typically longer than 150 nts)
2. mature tRNAs and miRNAs both have a 5' monophosphate and 3' OH, which are employed at different steps of library preparation

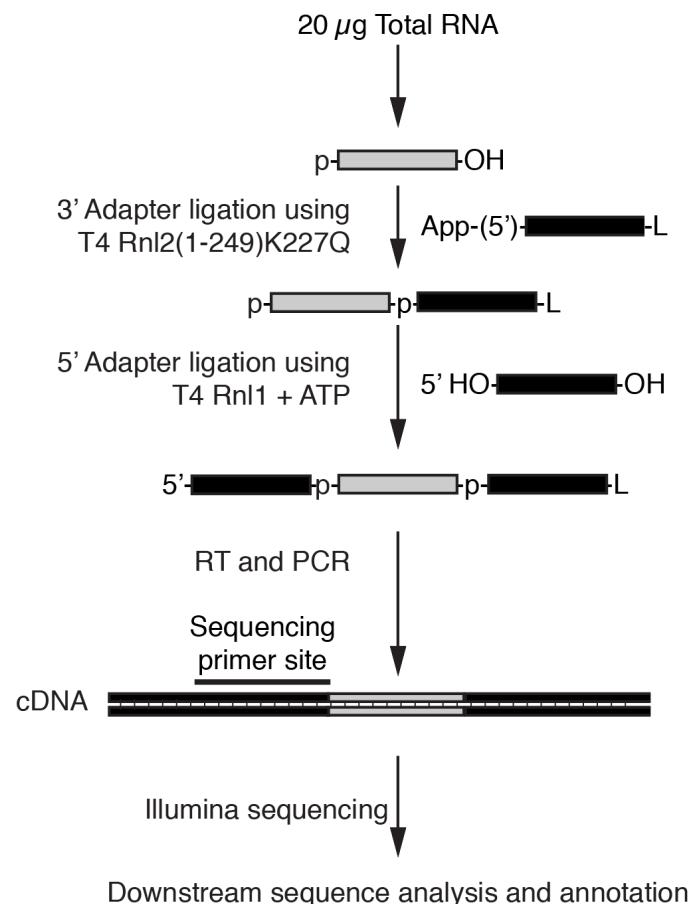


Figure 1.3: Small RNA sequencing protocol. Schematic overview of the conventional small RNA sequencing protocol, as it has been described previously [44]

The application of this protocol for tRNA sequencing, though, resulted in RNA-seq datasets with only ~2% tRNA content, with an average length of 59 nts (Table 1.1). These suggested that tRNAs were refractory to the small RNA sequencing protocol, and necessitated the development of a novel sequencing protocol.

RNA type	% Total reads	Mean length (nt)
rRNA	35.8%	60.5
no match	24.1%	76.2
no annotation	17.8%	64.2
snRNA/snoRNA	15.1%	62.5
repeat	3.8%	59.1
tRNA	2.0%	59.1
miscRNA	1.3%	63.1
miRNA	0.1%	22.2

Table 1.1: RNA types recovered by small RNA sequencing protocol Percentage of reads per RNA type over total depth of library, and mean length of reads mapped to RNAs of each type are shown. snRNA: small nuclear RNA; snoRNA: small nucleolar RNA; repeat: repetitive DNA sequence; miscRNA: all other ncRNAs

Chapter 2

Results

2.1 Hydrolysis-based tRNA sequencing

We wanted to obtain an RNA-seq validated list of human nuclear and mitochondrial tRNA gene, and their processing intermediates. This was my goal. To design a method for sequencing and a

To address the lack of a global and unbiased analysis of the human tRNA profile, I will develop an experimental and computational methodology for the generation of a reference tRNA atlas. To overcome existing experimental challenges, I will use a customized RNA-Seq technique (HydroRNAseq). To efficiently analyze the sequencing data in silico, I will develop a systematic and iterative bioinformatics platform

In order to overcome the problems associated with tRNA sequencing, we tried to identify the minimal number of simplest steps that could tackle the maximal number of problems. Thus, to curate and quantify human tRNAs, we isolated 60-100 nt-sized total RNA from HEK293 cells comprising both precursor and mature tRNAs, but being devoid of most other abundant RNAs and short tRNA

turnover products [16]. Full-length tRNAs have thermodynamically stable secondary and tertiary structures and are heavily modified by RNA editing, all of which compromise reverse transcription (RT) and RNAseq analysis. To overcome these problems, we implemented a limited alkaline hydrolysis step, which generates shorter RNA fragments with less structure and fewer modifications per fragment, and consequently more amenable to small RNA cDNA library preparation and deep sequencing. Furthermore, basic conditions also cleave the aminoacyl-tRNA bond, freeing the 3' terminal hydroxyl group required for 3' adapter ligation during RNA cDNA library preparation. This approach increased the tRNA read content to >40% in our deepest dataset **Table S1**. We named this procedure **hydro-tRNAseq** (**Fig. 2.1**). In summary, partial hydrolysis of tRNAs overcame technical limitations of suboptimal adapter ligation and RT and albeit this resulted in shorter reads, it also resulted in fewer errors at sites of modification per sequenced read, which ultimately improved the performance of the mapping algorithm.

2.2 Hierarchical sequence read mapping

In parallel with the tRNA annotation procedure, we had to build a bioinformatic pipeline for processing the obtained sequence information. We developed an iterative, hierarchical approach for mapping and annotating our sequence reads.

We mapped the reads to reference tRNA genes (hg19, <http://gtrnadb.ucsc.edu/>) using an iterative and hierarchical protocol (Fig. 1B). We started by mapping only to mature tRNAs, which included the 3' CCA aminoacyl acceptor terminus introduced posttranscriptionally by tRNA nucleotidyl transferase, and the G-1 nucleotide added posttranscriptionally to histidine tRNAs (Gu, 2003; Juhling et al.,

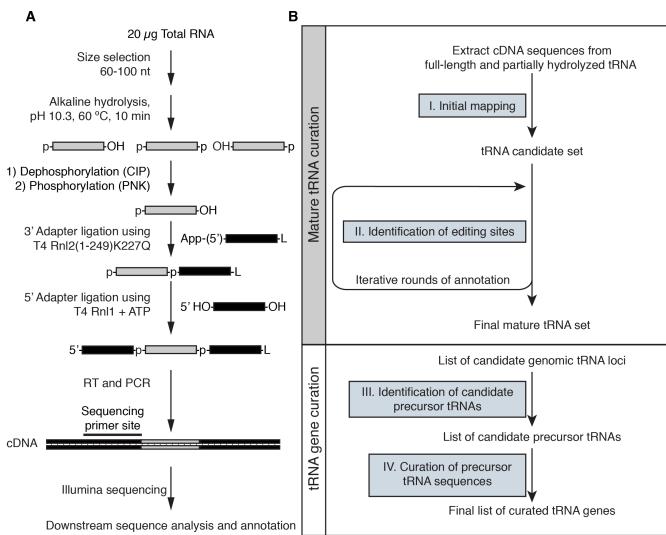


Figure 2.1: Experimental and bioinformatic pipeline for tRNA annotation and reference transcript curation by hydro-tRNAsseq. (A) tRNAs and pre-tRNAs were size-selected from HEK293 total RNA and subjected to limited alkaline hydrolysis, followed by dephosphorylation, rephosphorylation and conventional small RNA sequencing as described previously (Hafner et al., 2012). (B) An iterative mapping and annotation protocol was used to first annotate and curate fully processed and nucleotide-modified mature tRNAs. Leftover reads that spanned the mature-precursor junctions were used to identify transcribed tRNA genes.

2009), but excluded tRNA introns. Starting with two most abundant tRNA transcripts per isotype (tRNAs encoding the same amino acid) as indicated after the first mapping round, except for selenocysteine, where only one mature tRNA sequence could be identified, we performed iterative rounds of mapping and manual reference transcript selection, focusing in every step on transcripts that collected more reads with an error distance of 1-2 than 0. If these reads with mismatches could be assigned to other tRNA isoacceptors (tRNA accepting the same amino acid), these were included in our candidate reference set. Otherwise, we reasoned that the mismatches were the results of nucleotide-modification-induced errors of RT. In those cases, we accounted for the modified nucleoside signatures by introducing a new, edited reference transcript in our set. For tRNAs that exhibited multiple positions with high modification rates (>10% compared to reference),

we compiled reference sequences with all possible combinations of modified signatures at all detectably modified positions, aiming to account for the maximum possible number of mapped sequence reads. We ended the curation cycles when there was no observed modified position that exhibited a mismatch frequency greater than or equal to 10% compared to the reference. By performing this iterative process of curation, we obtained an experimentally validated reference set of mature tRNAs accounting for modified-nucleotide-induced sequence variation upon reverse transcription (Table S2).

Now, I can automatically and fast obtain relative read counts for all classes of ncRNAs in one experiment. Applying this pipeline to the tRNA-seq experimental results shows that indeed the majority of reads map to tRNAs. The depth of the library is extensive and thus allows for a reliable annotation of tRNA genes and transcripts.

2.3 Iterative manual tRNA curation

In order to identify possible tRNA gene loci, we mapped the curated tRNA sequences back to the genome, allowing for gaps to accommodate tRNA introns, as well as up to 7 mismatches to accommodate terminal and internal RNA editing events. By appending 40 nt 5' and 3' of the location of genomic mapping, we obtained a candidate pre-tRNA gene set. We mapped non-annotated residual reads to these candidates to identify 5' leader- and 3' trailer-comprising pre-tRNA reads, which also distinguished actively transcribed tRNA genes from silent ones or pseudogenes. Leader- and trailer-comprising tRNA genes show higher sequence variation, as evidenced by higher information entropy values, across the leader and trailer nucleotides than internal sequences within the mature tRNA

suggesting that even short precursor sequences with read coverage are sufficient for the annotation of non-redundant tRNA genes (Fig. S1). At the end of our analysis we accounted for 93% of the 114,367,140 reads in our deepest library (Table S1). Given the depth of sequencing, we are confident that we accounted for the vast majority of precursor and mature tRNAs. Indeed, *a posteriori* we looked for genomic regions that collected at least 50 overlapping reads throughout their whole length, fell within the 60- to 100-nt size window, and adopted a cloverleaf structure, in an effort to detect any tRNAs that might have been overlooked by our approach or in prior literature. The only sequences that we identified were U1 snRNA (pseudo)genes (Fig. S2), suggesting that our analysis was exhaustive, at least for tRNAs in HEK293 cells.

2.4 Protocol and pipeline outputs

2.4.1 Mature tRNA alignment

Representative alignment of reads to a mature tRNA reference., where the reference is shown at the bottom, the reads that map to it, and the depth of read abundance, as well as the reference locations at which its read maps.

Due to the intentional fragmentation of input RNA, we observe that the majority of reads were shorter than the full length tRNA, but there were enough long reads (like the one shown here in red) to bridge together separate segments of the obtained sequence .

Figure X. Alignment of tRNA-seq reads from HEK293 to mature tRNA transcript TRNAE5. Shorter reads are shown in black; a longer read, bridging the two halves of the tRNAs is shown in red. The frequency of each read (count) and the

number of locations that it maps within the tRNA reference with no mismatches are indicated. Vertical lines represent the relative frequency of binned, normalized read count in log₄ increments.

(Fig. 2.2)

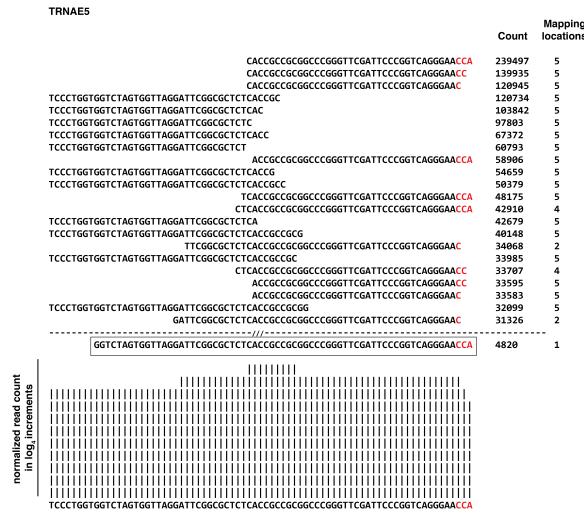


Figure 2.2: Mature tRNA alignment. The figure shows a table of tRNA sequences and their mapping counts, followed by a vertical bar chart representing the normalized read count in log₄ increments.

2.4.2 Pre-tRNA alignment

The reads that after a first pass were mapped to a mature tRNA were set aside, and the leftover reads were then mapped to tRNA genomic locations that were extended 40 nts up- and downstream from all mature tRNA boundaries. The reads that mapped to such tRNA precursors (that I will refer to as pre-tRNAs) were used to identify actively transcribed tRNA loci in our cell system of HEK293 cells (Fig. 2.3).

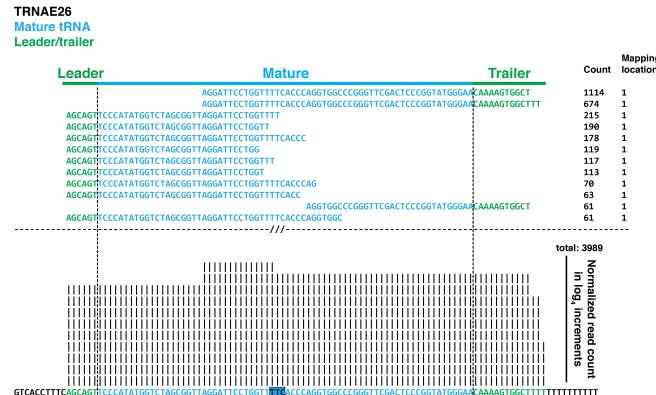


Figure 2.3: Pre-tRNA alignment!. Placeholder

2.5 Justification for using precursor reads

Entropy (Fig. 2.4)

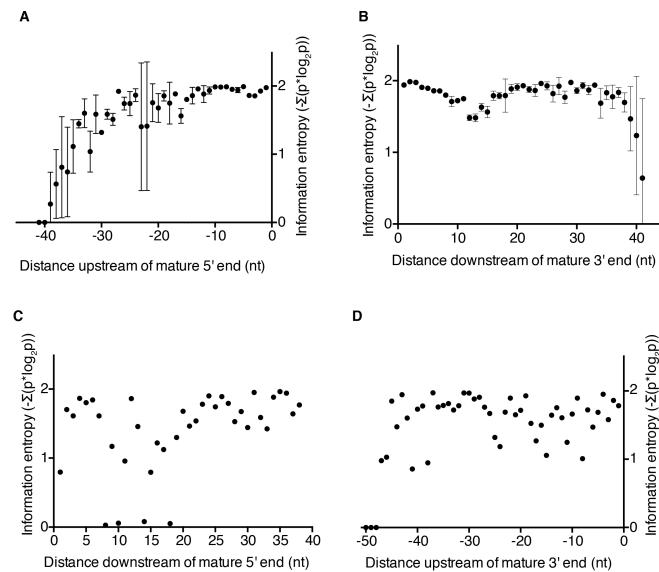


Figure 2.4: Information entropy in pre-tRNA segments and mature body (A,B)
 Information entropy $H = -\sum_{i=1}^n p(i) * \log(p(i))$, (where p is the frequency of each nucleotide at a given position, i, and n the total number of transcripts) was calculated using read evidence from hydro-tRNaseq (four replicates) for the 5' leader and 3' trailers of all pre-tRNAs with positions centered at the 5' and 3' ends of mature tRNAs. (C,D) Same as before, but using the reference sequence of mature tRNAs.

2.6 Composition of hydro-tRNAseq libraries

The majority of our reads obtained from 60-100 nt size-fractionated total RNA were assigned to mature tRNAs. The improvement we observed in recovering tRNA reads was considerable, as 2/3 of our reads mapped to either mature or pre-tRNAs or mitochondrial tRNAs (**Fig. 2.5**).

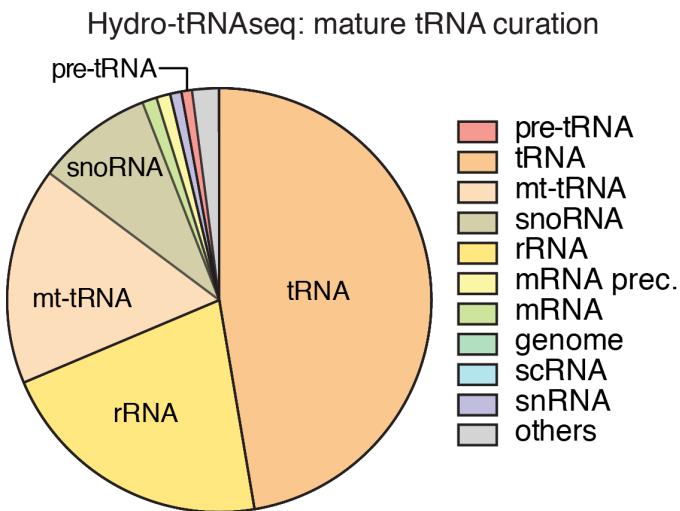


Figure 2.5: Composition of hydro-tRNAseq libraries. Total RNA composition of the 60-100 nt size fraction from hydro-tRNAseq according to RNA classes.

2.7 Need for pre-tRNA enrichment

Even though the majority of our reads obtained from 60-100 nt size-fractionated total RNA were assigned to mature tRNAs, only 1% of the reads comprised sequences overlapping with pre-tRNA leader or trailer sequences (**Fig. 2.5, Table S1**). This raised the possibility that we might have missed reads corresponding to lowly expressed or very rapidly processed pre-tRNAs.

We did this by performing PAR-CLIP (photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation), a technique developed in our lab to identify

RNA targets of RNA binding proteins at nucleotide level resolution and with high specificity.

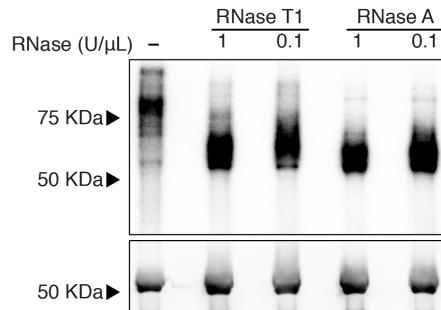


Figure 2.6: SSB crosslinking to RNA. Phosphorimage of SSB-crosslinked to radiolabeled RNA. PAR-CLIP was performed using RNase A or RNase T1, at two different concentrations to account for possible biases of RNase treatment conditions. Libraries from PAR-CLIP using 1 U/μL of RNase A and RNase T1 were prepared and submitted for sequencing. Western blot against HA, shown in the bottom, confirmed the immunoprecipitation of SSB.

2.8 PAR-CLIP methodology for the study of RNA-RBP interactions

A series of techniques have been developed for the study of RNA-RBP interactions on a genomic scale [27] from TRP. Our lab developed Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP), coupled with deep sequencing, which is a cell-based approach that allows the determination of RBP binding sites on RNA targets at nucleotide-level resolution (Fig. 2.7[28] from TRP). To enable efficient RNA-RBP crosslinking using long wavelength

UV, 4-thiouridine (**4SU**) is added to culture medium, taken up by cells and incorporated into nascent transcripts. The crosslinked ribonucleoprotein complex is submitted to partial RNase digestion, immunopurification and size-fractionated. Crosslinked RNA is recovered, converted into small RNA cDNA libraries, and sequenced. Importantly, crosslinking introduces a structural change in the thiouridine base, which allows pinpointing the position of crosslinking by scoring for characteristic T-to-C transitions in the sequenced cDNA. In addition, the abundant background derived from non-crosslinked fragments of co-purifying cellular RNAs do not contain these T-to-C transitions and can be filtered out. Thus, PAR-CLIP has a very low rate of false positive target identification, since the nucleotide transition signature reliably marks true crosslinking sites. PAR-CLIP has so far been applied successfully to the study of mRNA- and miRNA-binding proteins, but not tRBPs.

2.9 SSB PAR-CLIP

Therefore, we decided to complement our efforts with PAR-CLIP-sequencing of **SSB** (**Fig. 2.6** All tRNA genes are transcribed by POLR3, which terminates upon decoding an oligo-uridine (**oligoU**) region [45]. SSB binds to the short pre-tRNA 3' oligoU tail [46] prior to removal of the entire 3' trailer sequence. Therefore, we reasoned that SSB should bind all tRNA precursors, and that if we could isolate its targets, we would be able to reliably identify transcribed tRNA loci.

SSB exhibited a striking binding preference for pre-tRNAs and showed a drastic enrichment in precursor tRNAs compared to hydro-tRNAseq (**Fig. 2.8**), which confirmed our hypothesis, as well as previous observations [47]. We performed PAR-CLIP using two different nucleases to control for sequence biases at the nu-

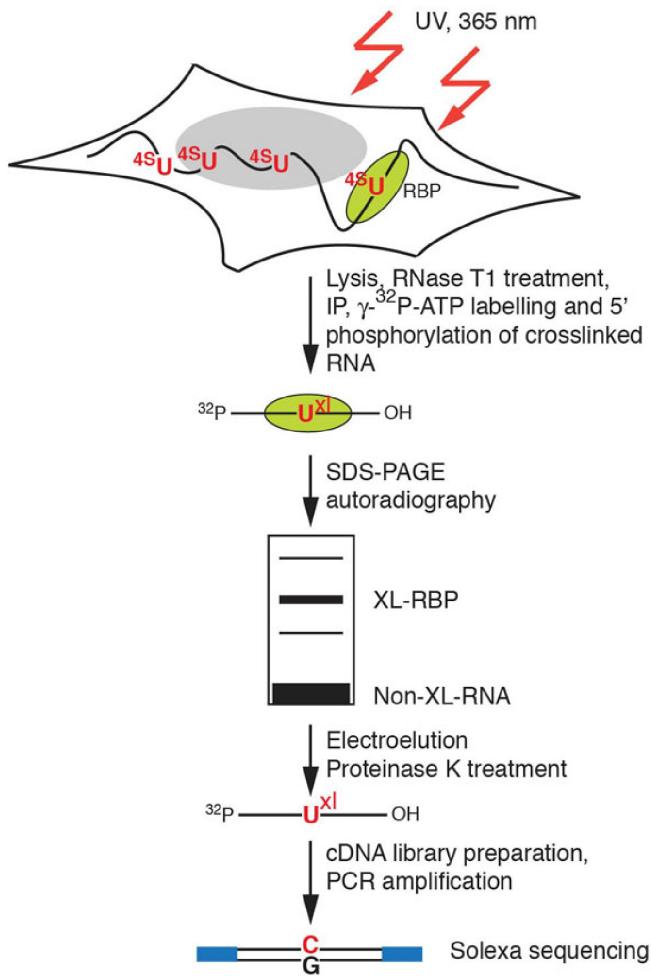
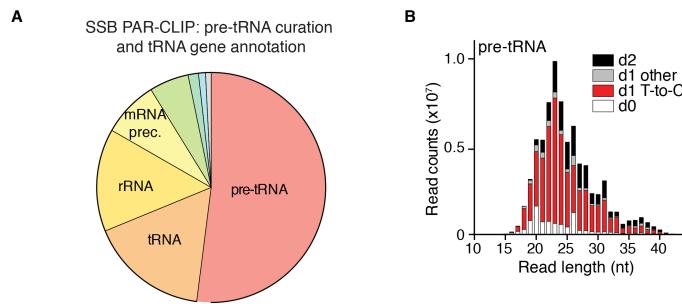


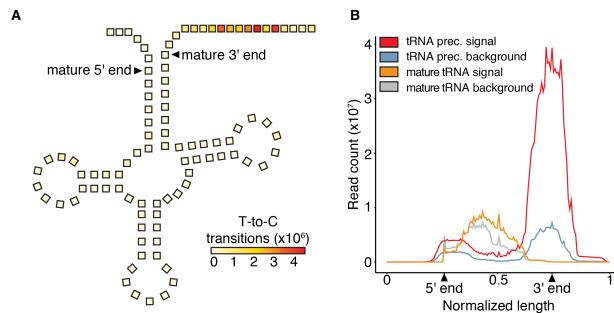
Figure 2.7: PAR-CLIP. Outline

please digestion step. RNase T1 resulted in longer precursor tRNA trailer sequences than RNase A, due the latter's preference for cleaving 3' to pyrimidines, which are highly abundant in the 3' trailer sequences. Overall, 46% of all PAR-CLIP reads mapped to pre-tRNAs (Fig. 2.8), the overwhelming majority of which showed the characteristic T-to-C transition, indicative of crosslinking (Fig. 2.8, table S3).

The vast majority of crosslinking sites in pre-tRNAs were concentrated, as expected, in the oligoU tract of the 3' trailer sequence (Fig. 2.9A,B). We also found that SSB crosslinked to the 5' segment of the mature tRNA body at conserved

**Figure 2.8: figure2cd. placeholder**

sites in the D-stemloop (**Fig. 2.9B**), which is a novel finding, hinted at by a report proposing that the affinity of SSB for a full-length pre-tRNA cannot be explained solely by its binding to the 3' oligoU tract [47]. The other major target of SSB was 5S ribosomal RNA (rRNA), which is the only POLR3-transcribed rRNA, and as such also terminates with an oligoU stretch to which SSB crosslinked (**Fig. S3**).

**Figure 2.9: figure2ef. placeholder**

2.10 tRNA gene annotation

We combined hydro-tRNAsseq and SSB PAR-CLIP to identify actively transcribed tRNA genes (genomic locations that give rise to a supported pre-tRNAs). We confidently identified 288 tRNA genes as the intersection of 4 replicates of hydro-tRNAsseq (**Fig. 2.10A**), and 349 tRNA genes as the intersection of two SSB PAR-CLIP experiments. Of note, SSB PAR-CLIP confirmed the expression of an addi-

tional 7 tRNA genes that were not supported in hydro-tRNAsq replicate (e.g. **Fig. 2.10B**), further showcasing the complementarity of the two approaches. We observed a strong correlation of pre-tRNA abundances between SSB PAR-CLIP and hydro-tRNAsq (Pearson $R = 0.72$; **Fig. 2.10D**), providing confidence that SSB PAR-CLIP quantitatively detected pre-tRNAs, without introducing biases (e.g. artificially enriching for lowly expressed pre-tRNAs). Instead, we observed no strong correlation between precursor and mature tRNA read counts in either of the two techniques ($R < 0.2$; **Fig. S4**). The correlation of identified isoacceptor counts between SSB PAR-CLIP and hydro-tRNAsq was virtually perfect (Pearson $R = 0.99$; **Fig. 2.15C**), ruling out the introduction of a pronounced systematic bias from our hydrolysis-based protocol. Some anticodons seemed to be served by multiple isodecoders (e.g. 19 isodecoders for tRNA^{Ser}_{GCA}), while others only from one (e.g. tRNA^{Ser}_{ACT};**(Fig. 2.15B, Fig. 2.15)**). Selenocysteine was the only amino acid that, in our data, was decoded by only one tRNA gene.

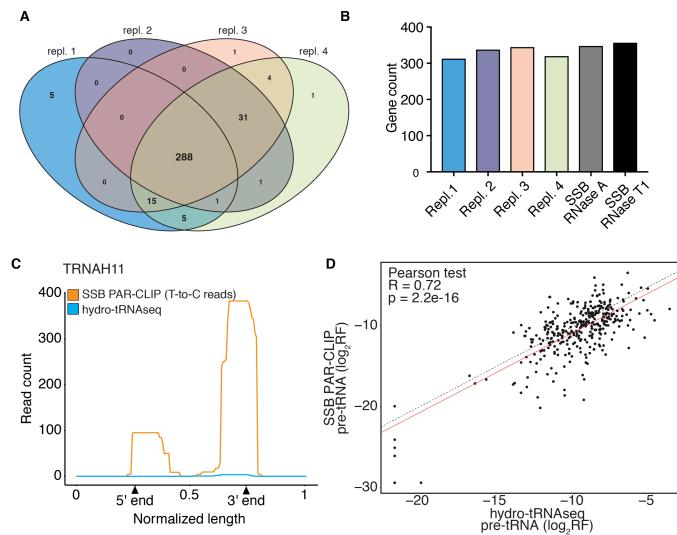


Figure 2.10: figure3. placeholder

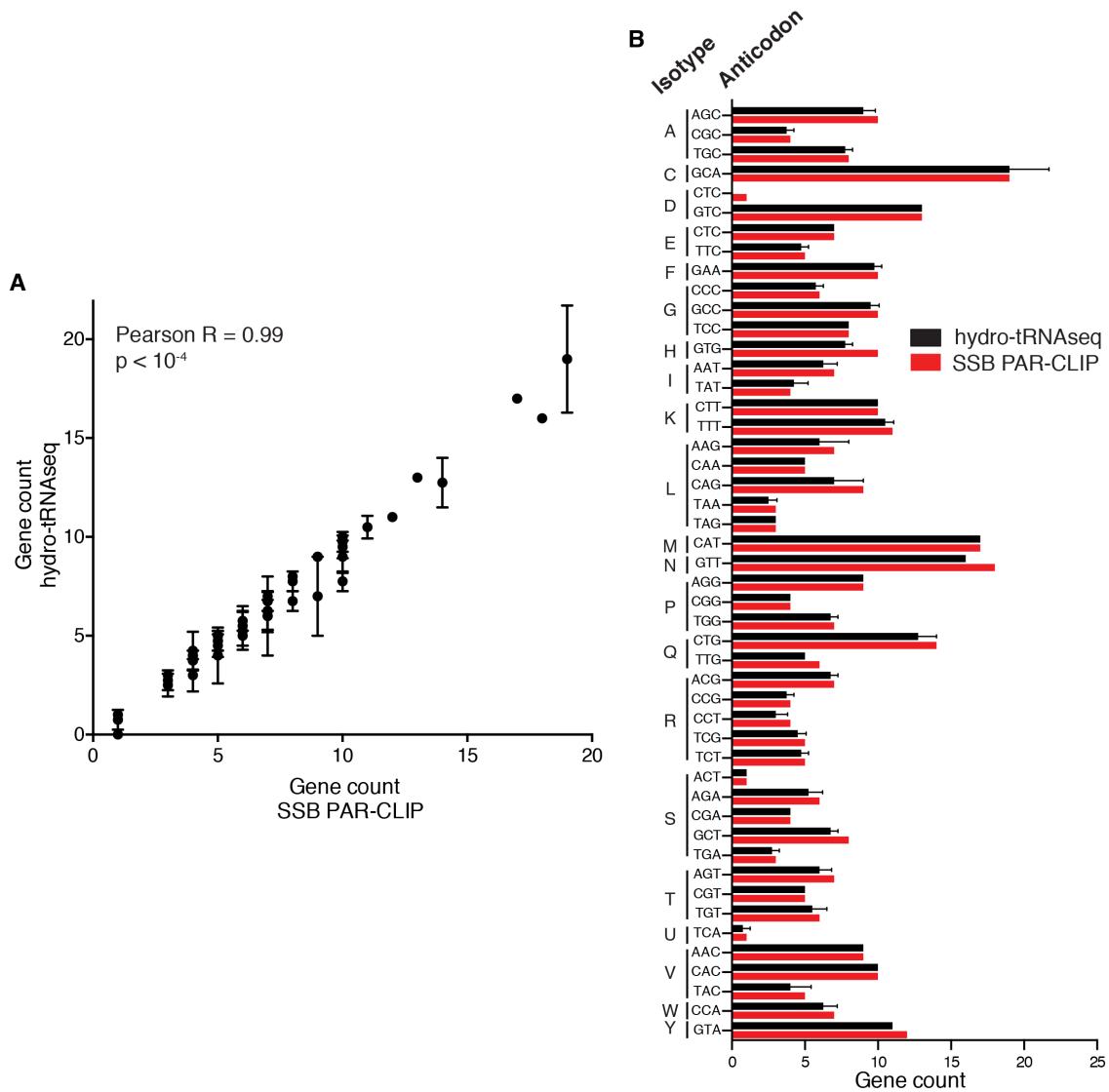


Figure 2.11: figure4. placeholder

2.11 Comparison with other methods

Recently tRNA sequencing methods have been developed that employ dealkylating enzymes and/or highly thermostable reverse transcriptase to overcome respectively the hurdles of modifications and stable structures that impede tRNA sequencing [27, 28]. However, they both have specific limitations that we tried to address. We size-selected at a higher size window to avoid contamination

by tRNA-derived fragments (as compared to [27]). Also, we used two sequential adapter ligation methods to make sure that only full-length fragments were sequenced and the sequencing reads were not results of blocks during RT (as compared to [28]), which allowed us to differentiate RT stops from fragment ends. Additionally, we did not bias our sequencing protocol towards mature tRNAs, but instead we captured more precursors by both RNAseq and more importantly PAR-CLIP methods, allowing us to perform a deeper precursor tRNA curation. Importantly, despite the reportedly high processivity conferred by dealkylating methyl modifications, in the previous studies only a small fraction of reads mapped at a given transcript were full-length reads (<1% of all reads), with a marginal increase compared to untreated controls (**Fig. 2.12**).

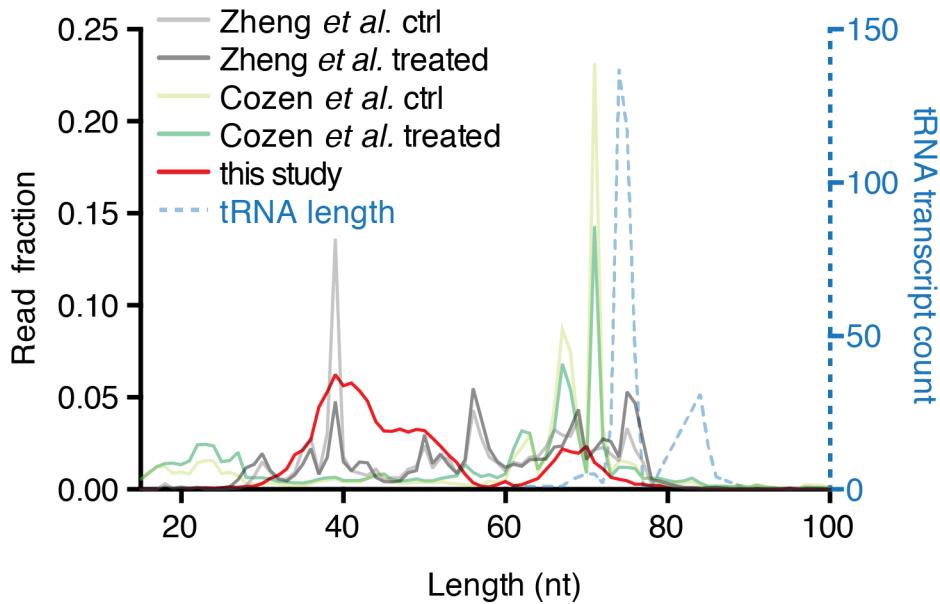


Figure 2.12: supp6. placeholder

In contrast, hydro-tRNAseq yielded a higher cumulative fraction of mature tRNAs with read evidence across their whole length with a mean read coverage of 0.99 of the full length (compared to 0.95 and 0.87 in previous studies; **Fig. 2.13**). Also, SSB PAR-CLIP was more sensitive in identifying tRNA genes, detecting 349

genes as compared to 159 and 212 in the other methods. Also, SSB PAR-CLIP was more sensitive in identifying tRNA genes, detecting 349 genes as compared to 159 and 212 in the other methods.

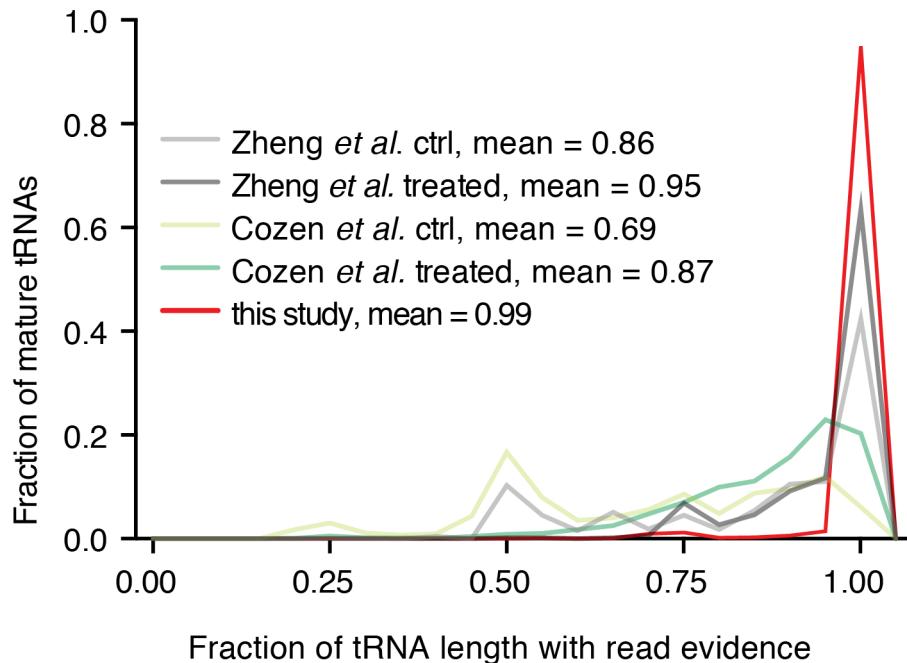


Figure 2.13: supp7. placeholder

2.12 Applications and biological insights

tRNA gene abundance does not correlate with tRNA gene count on the iso-type level

There is no monotonic relationship between number of tRNA genes per amino acid and the abundance of each class/family of tRNAs. This lies in contrast with prior publications that had assumed that the number of tRNA predicted tRNA genes can be used as a proxy of tRNA expression.

This was assumed in the absence of tRNA sequencing data and because in yeast it seems that all tRNA genes are expressed and seem to contribute equally

to the mature tRNA pool. So, this result underscore the need for caution when reporting tRNA abundance measurements and estimates.

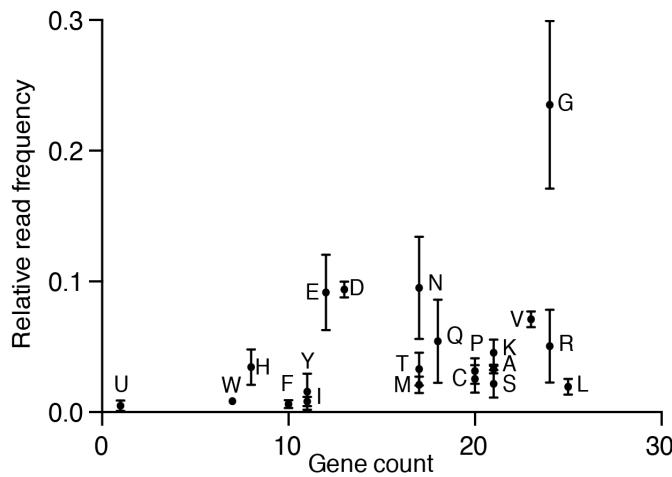


Figure 2.14: figure4A. placeholder

Although tRNA isotypes with higher relative abundances generally tend to have higher tRNA gene numbers, we did not observe a clear linear correlation between read frequency and gene count ($R = 0.12$; **Fig. 2.14**), like it has been reported before [12]. We then focused on the number of tRNA isoacceptors per amino acid, and isodecoders (tRNAs with the same anticodon sequence) per anticodon. We noticed a wide range of pre-tRNA counts per isoacceptor (**Fig. 2.15B**), with our data providing read evidence for 47 out of 62 coding codons (61 canonical and 1 selenocysteine TAG).

tRNA gene abundance does not correlate with tRNA gene count on the isoacceptor (same amino acid, different anticodon) level

The same non-monotonic relationship seems to be true also on the level of isoacceptors, that is tRNAs with different anticodons that decode the same amino acid.

On this graph the data are broken down by aminoacid, which you can see

as headers at the top, and then by anticodon which you can see on the bottom. The y-axis represents tRNA gene count, and the size of every disc the relative abundance of all mature tRNAs with a given anticodon.

Thus, even though, for example, Cys GCA is the tRNA with the highest gene count, Glycine GCC, is the tRNA group with the highest abundance.

Also, if you take a look at Proline, the group with highest gene count is the one with the lowest total abundance which is completely opposite to what was assumed before.

But how about correlation between individual pre- and mature tRNA levels

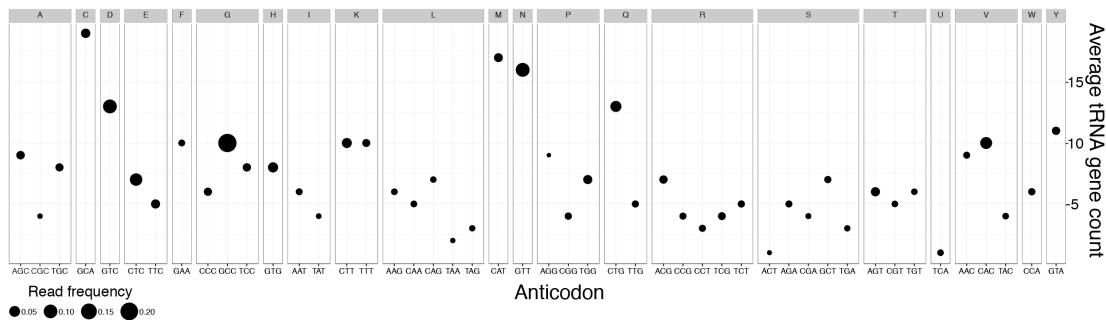


Figure 2.15: figure4D. placeholder

2.13 Mature tRNA abundance does not correlate with pre-tRNA abundance

No good correlation (pearson coefficients < 0.2) as identified by our 2 separate techniques. (Fig. 2.17)

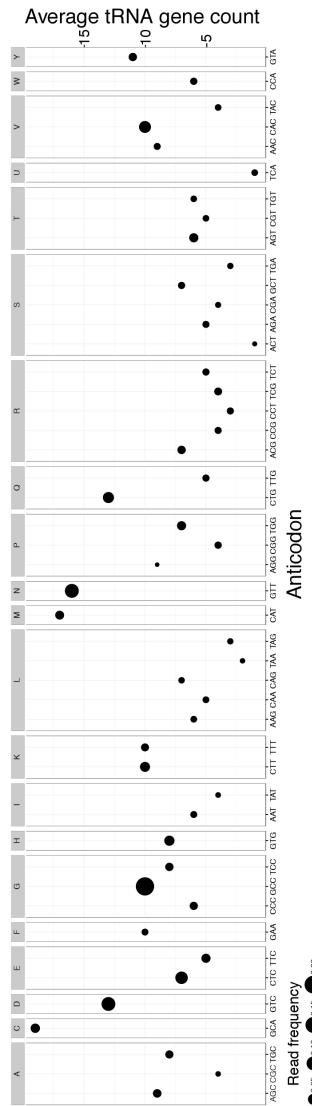


Figure 2.16: figure4Drot. placeholder

2.14 tRNA transcription initiation and termination

Besides tRNA gene annotation and quantification, our approach yielded insights about pre-tRNA 3' trailer sequences. Based on hydro-tRNAseq, we determined the median 5' leader and 3' trailer lengths to be 6 and 10 nt, respectively, with the trailer lengths showing a broader distribution (**Fig. 2.19A,B**). Interestingly, SSB PAR-CLIP revealed a subset of much longer trailers (**Fig. 2.19C**), suggesting that

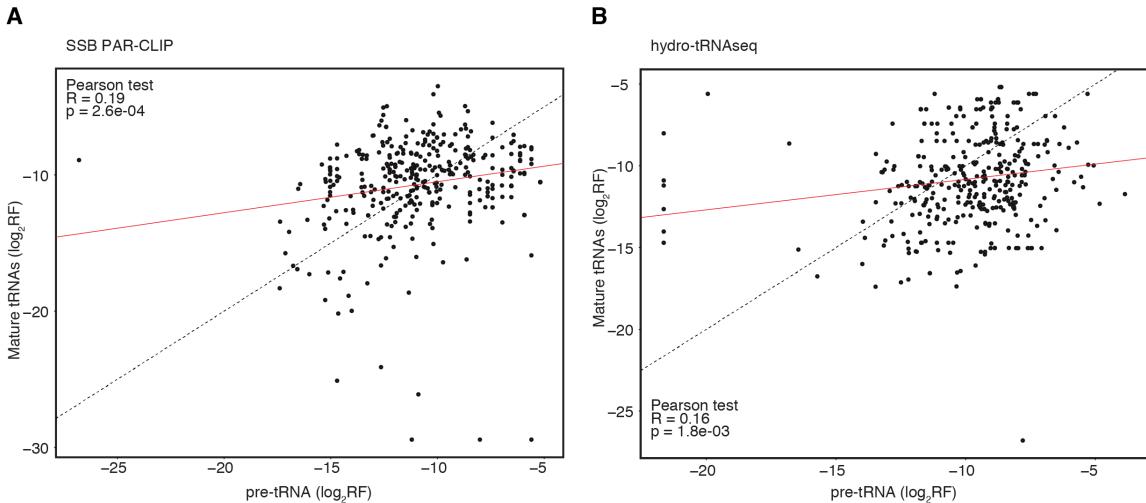


Figure 2.17: supp4. placeholder

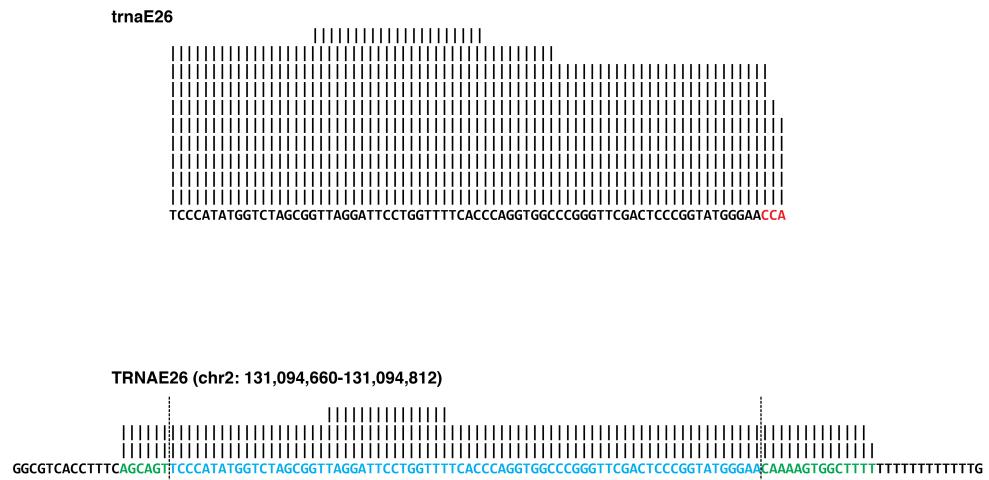


Figure 2.18: clp1 bar. placeholder

SSB PAR-CLIP captured the very initial steps of precursor tRNA processing, and accordingly that hydro-tRNAseq captures pre-tRNAs partially trimmed, either by ELAC2 (tRNase Z) or some other nuclease [34].

We next focused on the POLR3 oligoU termination signals. Various reports in the past have focused on the oligoU requirements for transcription termination in different species [48, 49]. SSB protected consistently a 3' 4 to 5 nt oligoU stretch, which was also confirmed by hydro-tRNAseq (Fig. 2.20). This is in agreement with previous *in vitro* results [46, 47, 50]. We also addressed the proposed re-

quirement for a stem-loop immediately upstream of the oligoU termination signal [49]. Secondary structure predictions for the trailer sequences with documented sequence evidence in hydro-tRNAseq and SSB PAR-CLIP did not detect predicted stable stem-loop structures for approximately half of all pre-tRNAs (**Fig. 2.21**). This argued against a formal requirement for a stem-loop in the termination process of POLR3, at least on tRNA genes, in accordance with previous biochemical evidence [48].

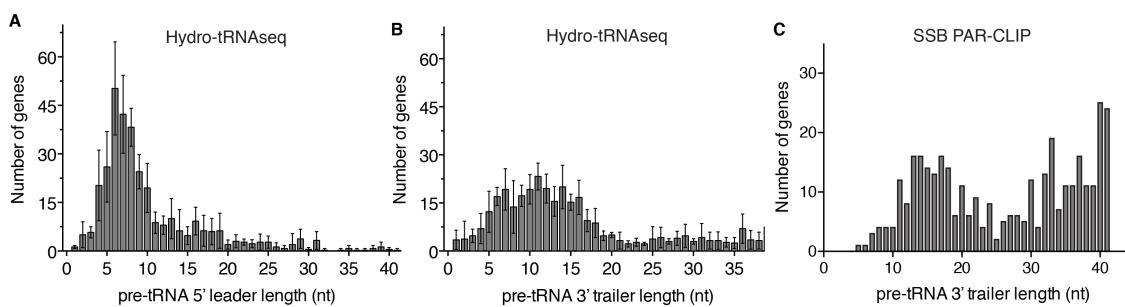


Figure 2.19: figure6. placeholder

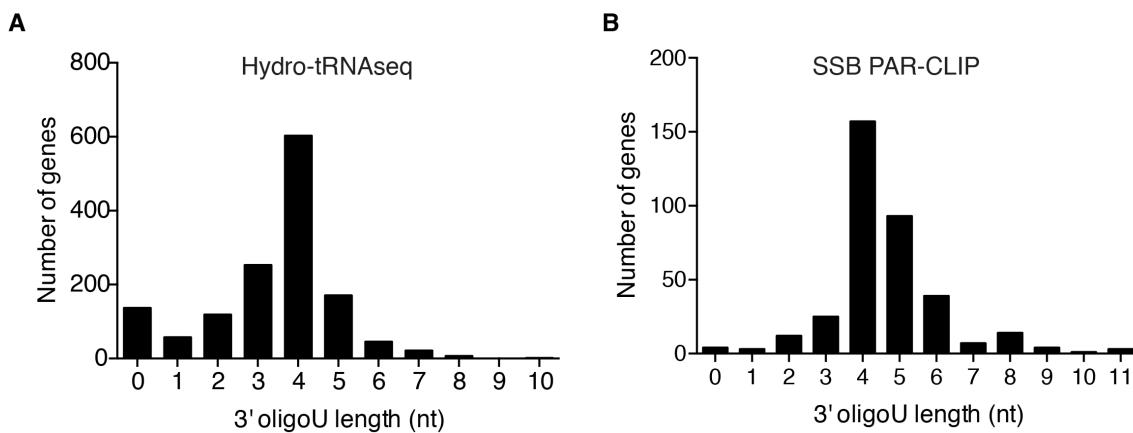


Figure 2.20: figure6de. placeholder

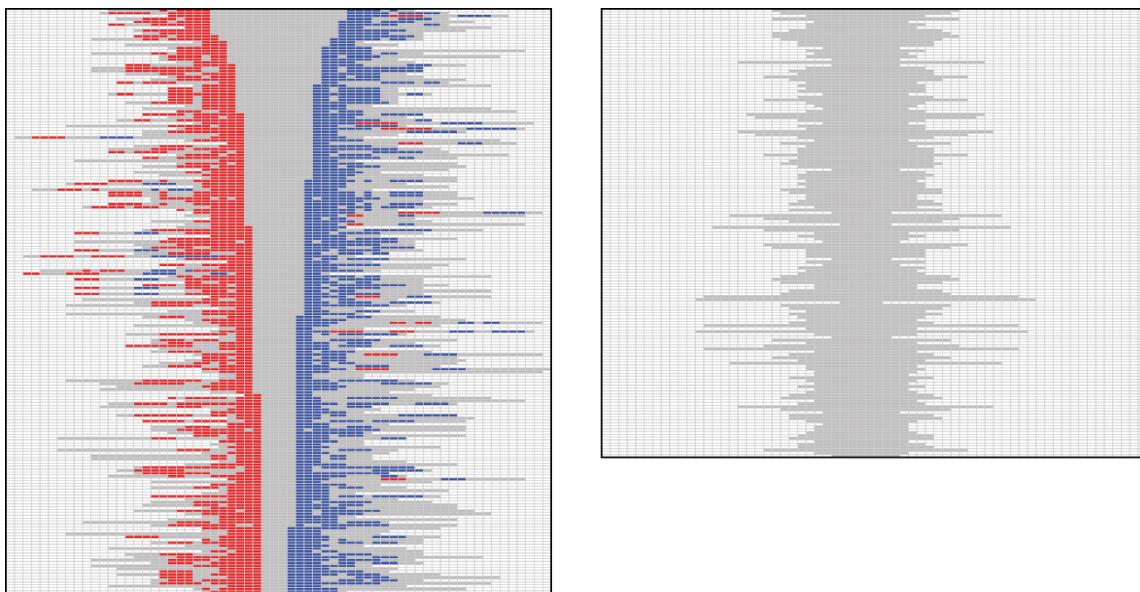


Figure 2.21: supp5. placeholder

2.15 Ribonucleotide modifications

RT across modified nucleotide-containing RNA leads to errors in cognate deoxynucleotide incorporation, revealed by mismatches in sequence reads upon mapping to reference genomic sequence. Read coverage across regions with a high degree of modifications may result in incomplete or largely uneven coverage. Therefore, we included in our mature tRNA reference the combination of all frequent mismatch signatures in all heavily modified positions. We reported the most frequently modified positions per tRNA gene ([Table S5](#)), and computed the frequencies of every nucleoside change per position across all tRNA genes ([Fig. 2.22](#)).

The majority of editing events were A-to-G transitions at the first position of the anticodon and at the position 3' to the anticodon (usually position 37). Both positions are known to be heavily modified, the former being deaminated to inosine, and the latter further modified (e.g. 1-methylinosine) [51]. In our data the majority

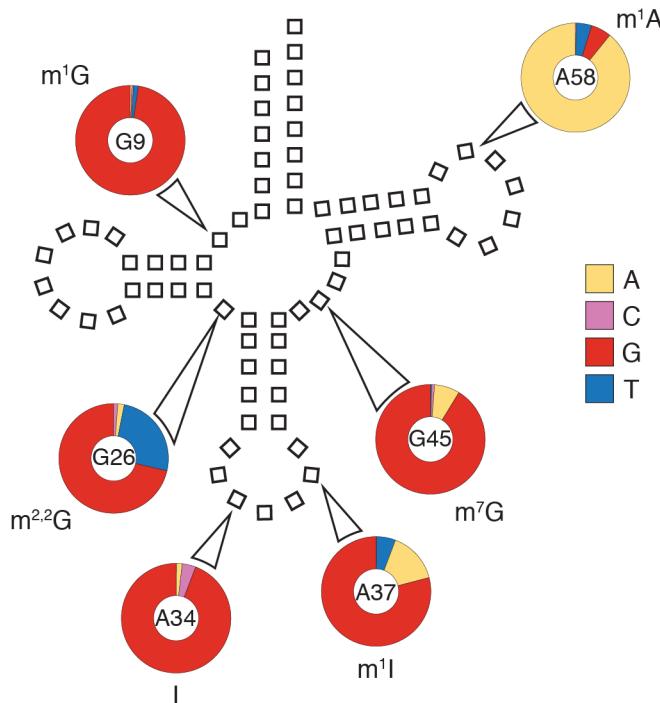


Figure 2.22: paper7a. placeholder

of the reads that mapped to the anticodon of the modified tRNAs contained the mismatches. To a lesser extent we could also detect 1-methyladenosine in the pseudouridine loop (returned as A-to-T or A-to-G), and various guanosine modifications at positions 9, 26, and 45, which most likely correspond to 1-methyl-, N₂,N₂-dimethyl-, and 7-methyl-guanosine, respectively [51].

The temporal resolution of tRNA modifications by RNAseq has begun to be addressed recently [52], however at a single modification level (inosine 34), and by using libraries relative poor in tRNA reads (<1% of total reads). We were appropriately poised to address this issue since our very deep sequencing set, in combination with our hierarchical annotation pipeline, offered the advantage of dissecting multiple modifications simultaneously. We focused on the inosine modifications, since they represented the majority of modified nucleosides. By inspecting read alignments with error distance 1 to the reference pre-tRNA, we noticed A-to-G

transition mismatches at position 34 in reads that retained the leader and trailer sequences of the precursor tRNA (**Fig. 2.23, top**). This confirmed that A34 deamination takes place at the precursor level, and therefore is a nuclear modification, as it has been previously reported [52]. Next, we noticed that 1-methyl-inosine at position 37 also appears at the precursor stage. Of note the A37 modification became apparent prior to A34, as the majority of the error distance 1 reads contained a mismatch at A37. Reads with two mismatches contained both modifications (**Fig. 2.23, bottom**)

In many instances we observe specific and highly abundant mismatchsignatures in pre-tRNAs, which can possibly allow us to determine and monitor by RNA-seq the editing hierarchy of tRNAs and the order of function of their modifying enzymes.

We observe editing in the D1 alignments. Those reads do not map at error distance 1, since we are following a hierarchical mapping, neither to they map back to the genome. This suggests that we can observe nucleotide editing at the precursor stage of tRNAs.

What we can say is that we observe tRNA editing leading to mismatches in RNA-seq specifc positions edited many times more abundant than unedited This slide brings up another point...

We are far from experts on tRNA editing and modifications, but our protocols do allow for a careful annotation and overview of specific modifications. Those that lead to characteristic signatures of mismatches upon RNAseq In many instances we observe specific and highly abundant mismatchsignatures in pre-tRNAs, which can possibly allow us to determine and monitor by RNA-seq the editing hierarchy of tRNAs and the order of function of their modifying enzymes.

We observe editing in the D1 alignments. Those reads do not map at error

distance 1, since we are following a hierarchical mapping, neither to they map back to the genome. This suggests that we can observe nucleotide editing at the precursor stage of tRNAs.

TRNAA6 d1 alignment



TRNAA6 d2 alignment



Figure 2.23: paper7b. placeholder

2.16 Annotation of intron-containing tRNA genes

Intron-containing tRNAs represent a particularly interesting set of tRNA genes, as mutations in their evolutionarily conserved, yet distinct, processing machinery have emerged recently as causes of severe neurodevelopmental syndromes,

such as pontocerebellar hypoplasia [53]. Therefore, there is documented need for a comprehensive annotation of human intron-containing tRNAs, which should be revisited as markers or disease-causing candidates in phenotypically similar conditions. We confirmed 26 out of 32 predicted intron-containing tRNAs by hydro-tRNAseq (**Fig. 2.24A**). Excluding any unknown biologically redundant mechanism, this suggests that the integrity of the tRNA splicing complex is essential for survival. To further confirm our observations, we coupled hydro-tRNAseq results with previously published PAR-CLIP data on the human tRNA ligase, RTCB [54]. Despite the shallow read depth of the dataset, we identified a crosslinked read peak at the anticodon loop of all intron-containing tRNAs annotated by our approaches. (**Fig. 2.24B,C**). One tRNA isotype and three isoacceptors are fully dependent on tRNA splicing, which suggests that despite functional redundancies that are often observed in tRNA processing steps, tRNA splicing is an essential process for viability, at least in the cell system that we studied.

2.17 CLP1

This came at an opportune moment because of the studies of CLP1

- Recessive mutations in children with neurodevelopmental disorders were localized to tRNA splicing factor CLP1
- Exome sequencing identified R140H recessive mutation
- R140H reduced interaction between CLP1 and the TSEN splicing complex

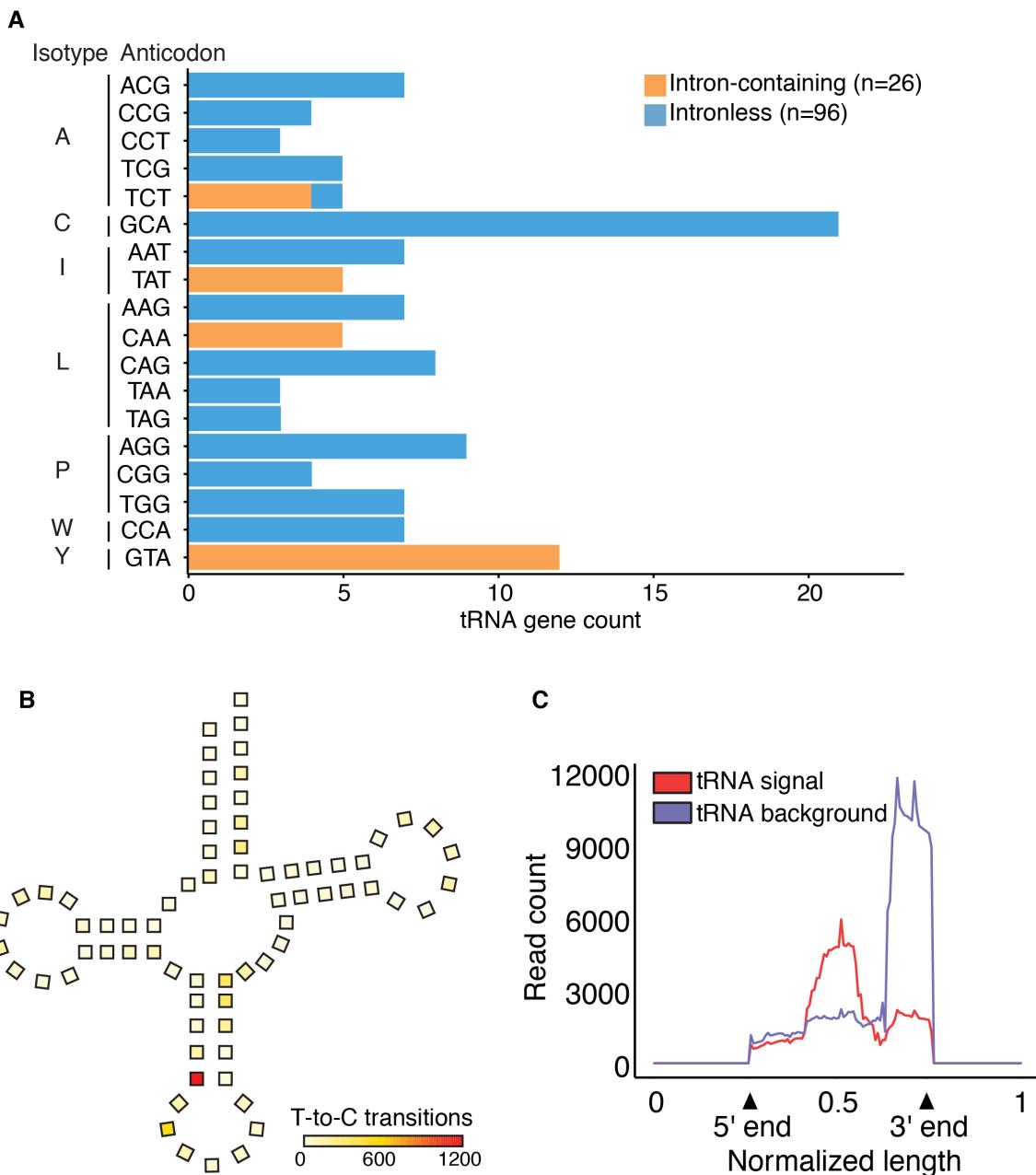


Figure 2.24: figure5. placeholder

2.18 CLP1 figures

2.18.1 Plausible pathomechanisms of CLP1 mutations

- i) Reduced tRNA splicing in sensitive tissues (e.g. neurons)

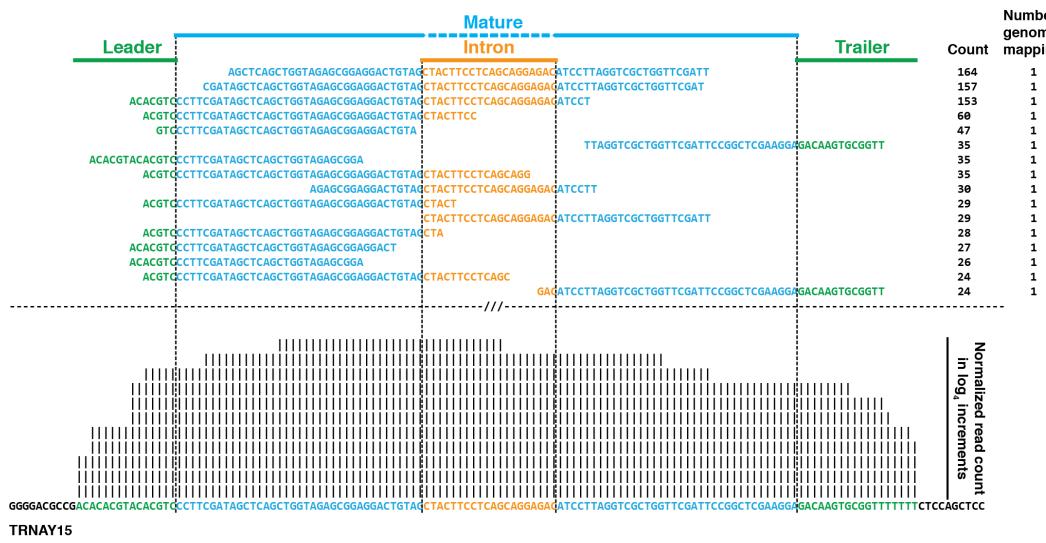


Figure 2.25: intron-containing tRNA. placeholder

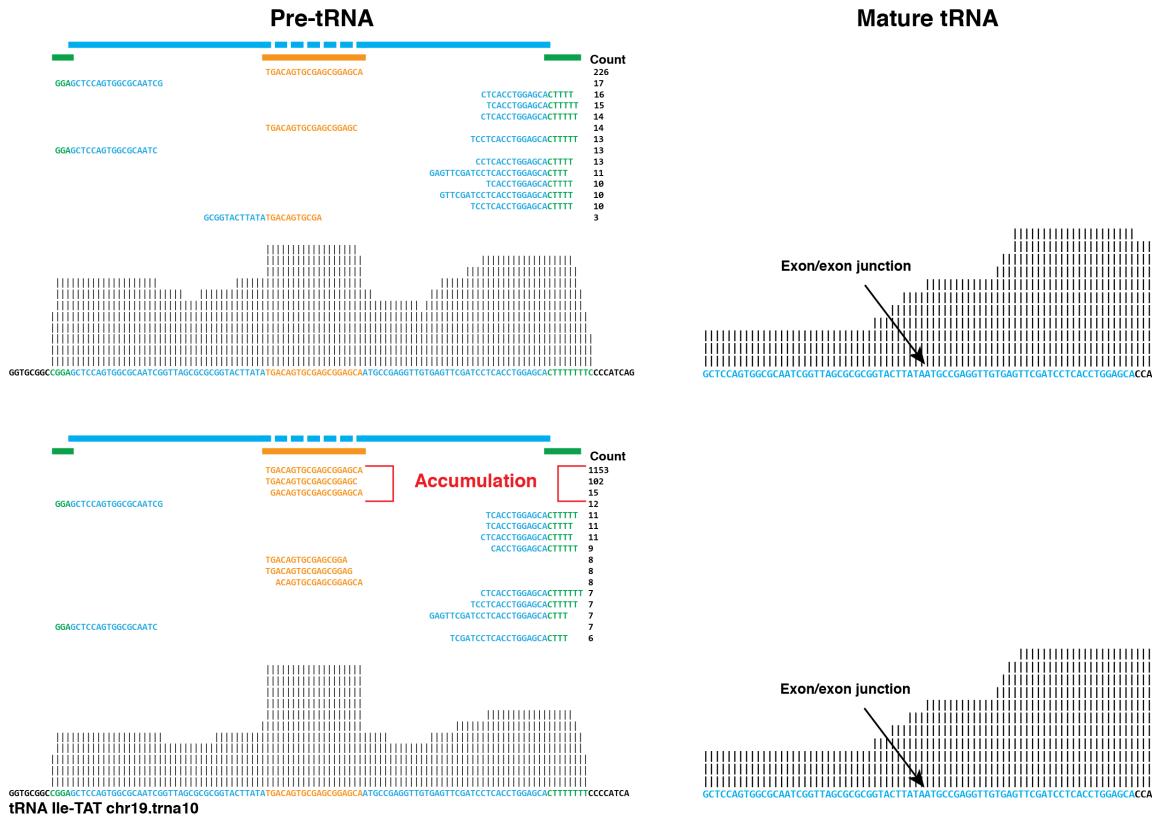


Figure 2.26: clp1 alignments. placeholder

- Innate immunity activation by tRNA intron accumulation
- tRNA-unrelated effects (e.g. CLP1 participates in mRNA polyadenylation)

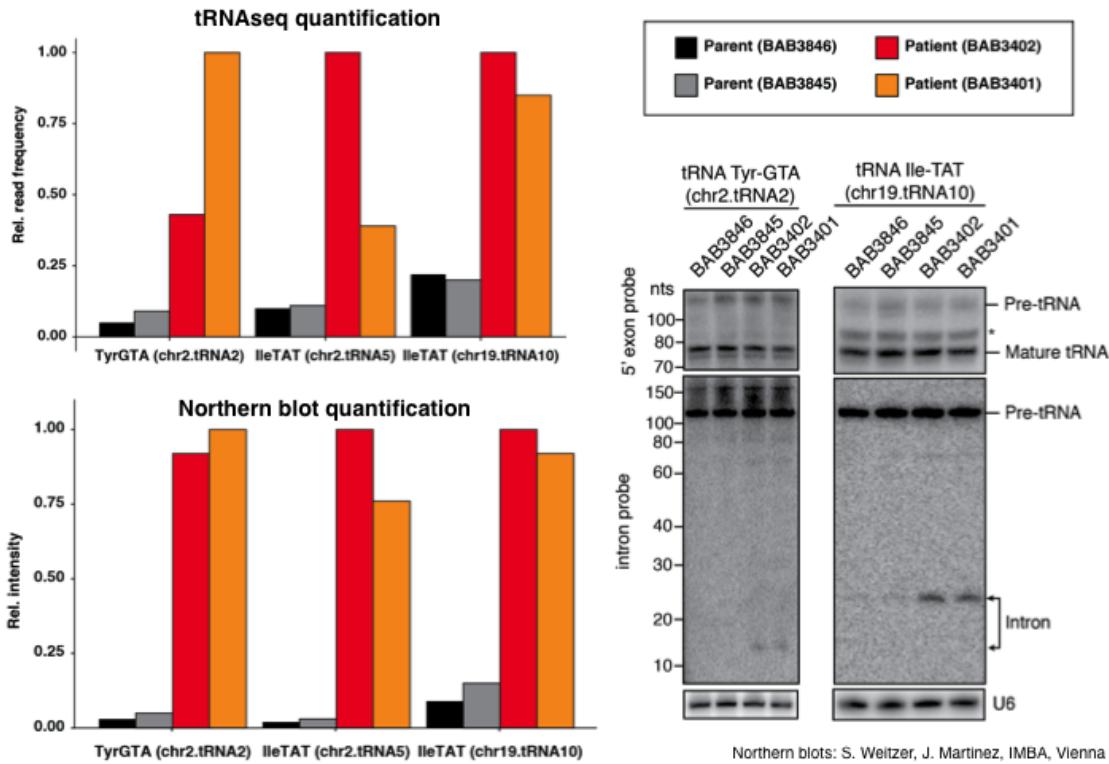


Figure 2.27: clp1 bar. placeholder

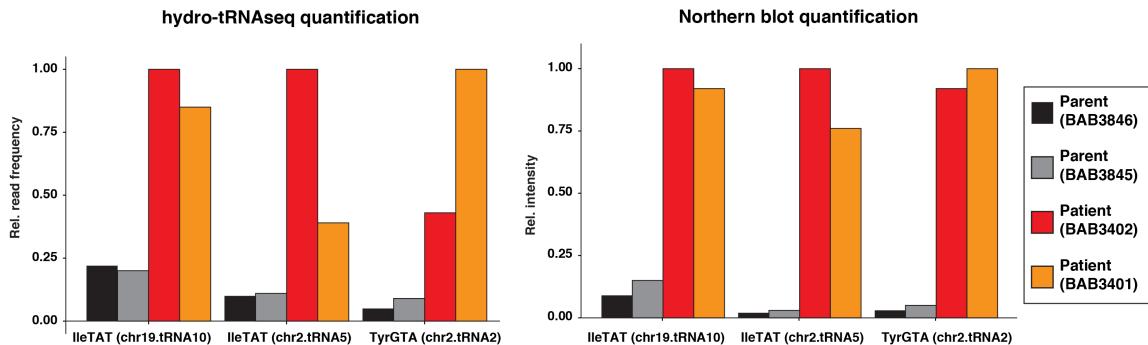
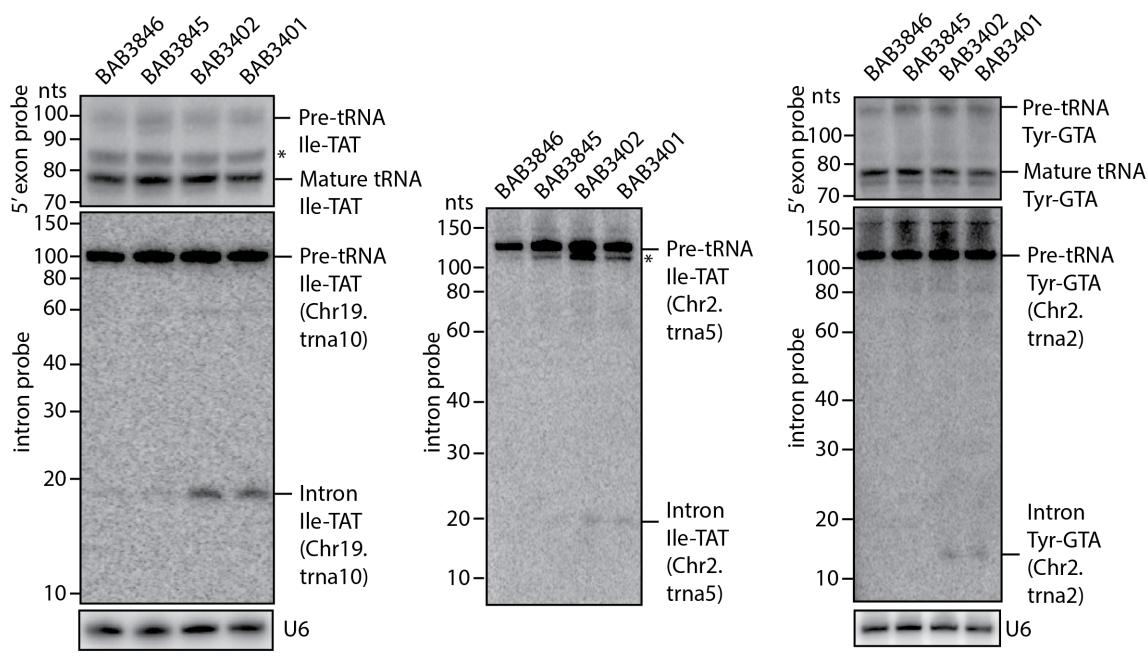


Figure 2.28: clp1 bar. placeholder

2.18.2 hydro-tRNAseq on CLP1

tRNA intron accumulation in patient with CLP1 recessive mutation compared to heterozygous parent



S. Weitzer, J. Martinez, IMBA, Vienna, Austria

Figure 2.29: NB. placeholder

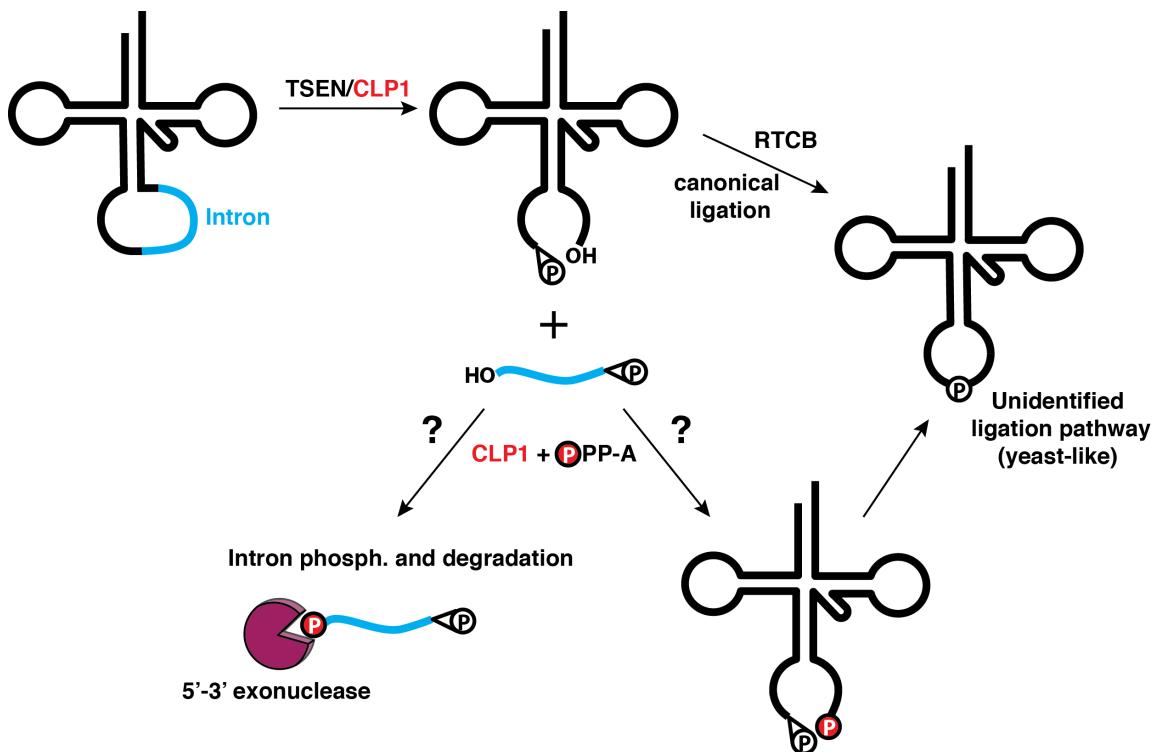


Figure 2.30: splicing. placeholder

2.19 tRNA enzyme screen

With respect to more functional aspects of tRNA biology, I focused on the results from knockdowns from a series of tRNA modifying and processing enzymes. Surprisingly, even though I targeted key components of the tRNA biogenesis and processing pathway, I could not observe major defects in tRNA processing or abundance patterns with the exception of the tRNA ligase (RTCB) that led to a very modest accumulation of 5' end tRNA halves. This suggested either that the residual enzymatic activity after knockdown is sufficient for proper tRNA processing or that there is unknown redundancy in several steps of tRNA biogenesis.

This publication steered interest in the connection of tRNA and disease, a link that was previously not well examined. To further showcase the utility and usefulness of my method in this field I decided to use it in a proof of principle example. I have carried out an siRNA-mediated gene silencing screen targeting a series of tRNA processing and modifying enzymes (i.e. the CCA terminal transferase, the tRNA methyltransferase TRMT10A, the tRNA splicing ligase HSPC117, a member of the RNase P complex, and also TRANSLIN, a putatively novel tRNA processing enzyme – see aim 2). In the cases where a reliable antibody was available, the knockdowns were validated by western blot analysis. In all other cases, knockdowns are currently being validated by mRNASeq. RNA from the validated knockdowns will be subjected to tRNASeq, with the expectation to identify tRNA processing defects that prior to our method would have been overlooked. In addition, I have performed tRNASeq from spinal cord-derived fibroblasts of knockout mice for the tRNA ligase. There I was already able to show tRNA processing is affected characterized by an accumulation of unspliced tRNA halves.

2.20 Discussion

We have combined two complementary transcriptome-wide approaches to provide experimentally validated annotation for mature and pre-tRNA transcripts and their respective genes, in addition to furnishing an accurate quantification of tRNA abundance in human cells.

First, we developed hydro-tRNAsseq, a fragmentation-based protocol for overcoming hurdles of tRNA sequencing, and obtained deep sequencing sets that enabled the annotation of tRNA genes and derivation of mature tRNA reference sequences for accurately assigning sequence reads to the otherwise edited and nucleotide-modified original tRNA. Alkaline hydrolysis of the tRNA-containing pool relieved thermodynamically stable structural constraints that impair ligation steps in the cDNA library preparations, reduced the number of modified nucleosides per sequenced fragment, resulting in high read-through in the RT step, and release of the 3' hydroxyl group of the otherwise aminoacylated tRNA 3' end.

Then, we took advantage of the pre-tRNA binding properties of SSB protein, which coordinates posttranscriptional processing and maturation of tRNAs [45], to enrich for tRNA precursors and allow for a comprehensive curation of pre-tRNA transcripts and annotation of tRNA genes. Of note, since SSB interacts with pre-tRNAs and other small nuclear RNA U-rich 3' ends in all organisms examined so far [50, 55], our approach can be adapted towards tRNA annotation in other species.

Our data suggest that, at least in our experimental system, the tRNA gene space is considerably more contracted than it has been previously predicted by bioinformatics, evidenced by the fact that almost half of the predicted tRNA loci were transcriptionally silent, presumably representing retrointegrated tRNA pseu-

dogenes. It would be interesting to examine whether such an observation holds for various cell types and at different stages of development or disease, in order to confirm the differential expression and regulation of tRNA gene expression that has been reported before [21, 26]. Our analysis shows that selenocysteine seems to be the only monogenic tRNA species, suggesting an increased sensitivity to mutations due to the lack of functional redundancy.

Our approach allowed the elucidation of relevant issues regarding POLR3 transcription such as the length of pre-tRNA leaders and trailers, the length of oligoU required for recognition by SSB, both of which have shown species specificity [56]. We also detected that 4 sequential Us act as the transcription termination signal for POLR3, confirming similar predictions based on genomic sequences that suggested a requirement for 4 Us for efficient termination in vertebrates [57, 58] as well as structural data documenting the capacity of SSB to accommodate 4 Us in its binding site [46, 50]. At the same time, the length distribution of the oligoU tract identified in our experiments reflects the heterogeneity of termination signal lengths that has been noted as an intrinsic property of pol III *in vitro* [56].

We could also confirm a second binding site of SSB in the 5' half of the mature tRNA sequence, in support of previous observations proposing the presence of additional pre-tRNA binding sites besides the 3' tail [46, 47]. It has been previously noted that the binding mode of SSB to tRNA is more complex than the recognition alone of the 3' tails, and that one of the RNA recognition motifs present in SSB, RRM1, could bind elsewhere in the tRNA. This stems from two observations:

- i) SSB has a higher affinity for precursor over mature tRNAs,
- ii) structural data show that RRM1 is unoccupied when SSB is bound to UUU-3'-OH substrate.

Our data seem to validate this observation, and could shed light into new modes of SSB-mediated processing of pre-tRNAs into either mature tRNAs or other kinds of ncRNAs [14].

Moreover, we were able to carry out a careful overview and tRNA modifications that result in characteristic mismatch signatures. We introduced all possible combinations of all “mutated” nucleotides at the most prominently modified positions in every tRNA in order to collect as many reads that could be having RT misincorporations at the modified positions. This created a large number of similar tRNA sequences, and therefore we allowed for extensive multimapping, but split the read counts in order to avoid artificial read count inflation. By making use of our hierarchical annotation pipeline, we were able to dissect the temporal order of inosine modifications at the tRNA anticodon, confirming that A34 deamination occurs in the nucleus prior to the nucleolytic processing of the pre-tRNA, and establishing that the same holds true for A37 modifications, which in fact precede A34 deamination. Accounting for modification signatures was also important for the reason that CLIP-seq, and especially PAR-CLIP, depends on apparent mismatches (in the case of PAR-CLIP, T-to-C conversions) for the identification of RBP binding sites on target RNAs. Since we used PAR-CLIP of SSB for the annotation of pre-tRNAs and tRNA genes, we first examined uridine modifications that result in T-to-C conversions. Only a small minority of modification signatures were T-to-C transitions, suggesting that it is highly unlikely that our PAR-CLIP data were artificially inflated. Collectively, our results give a census of the tRNA transcriptome in human cells. We provide a method and computational flowchart for the analysis of tRNAs. We expect that these results will inform studies of tRNA-related human disease.

2.21 Summary

- i) Hydro-tRNAseq is a facile and efficient method for sequencing tRNAs
- ii) PAR-CLIP of SSB/La informs the annotation tRNA genes and curation of pre-tRNAs
- iii) Combined together, they can be used to probe long-standing questions of tRNA biogenesis, processing and function
- iv) Hydro-tRNAseq can be applied for studies of human genetic diseases

Chapter 3

Materials and methods

3.1 Hydro-tRNAseq

Total RNA from HEK293 (Flp-In T-Rex, Invitrogen) was isolated using TRIzol (Invitrogen). For each sample 20 μ g of total RNA were resolved on 12% urea-polyacrylamide gels and recovered within a size window of 60-100 nt. The eluted fraction was subjected to limited alkaline hydrolysis in a 15 μ L buffer of 10 mM Na₂CO₃ and 10 mM NaHCO₃ at pH 9.7 either at 65 °C for 10 min (replicate 1) or 1 h (replicates 2-4).

The partially hydrolyzed RNA was dephosphorylated with 10 U of calf intestinal phosphatase (NEB) in a 50 μ L reaction of 100 mM NaCl, 50 mM Tris-HCl, pH 7.9 at 25 °C, 10 mM MgCl₂, 1 mM DTT, 3 mM Na₂CO₃ and 3 mM NaHCO₃, at 37 °C for 1 h. The resulting RNA was re-phosphorylated with 10 U of T4 polynucleotide kinase (NEB) in a 20 μ L reaction of 70 mM Tris-HCl, pH 7.6, 10 mM MgCl₂, 5 mM DTT and 1 mM ATP, at 37 °C for 1 h. Fragments of 19-35 nt were converted into barcoded small RNA cDNA libraries, as previously described [44], and sequenced on an Illumina HiSeq 2500 instrument. Adapters were trimmed using cutadapt

(<http://journal.embnet.org/index.php/embnetjournal/article/view/200/458>). Sequencing read alignments were performed using the Burrows-Wheeler aligner against an in-house curated and annotated list of mature and precursor tRNAs containing predicted tRNA sequences for human genome version hg19 (<http://grnadb.ucsc.edu>). Sequencing reads were first mapped against mature tRNAs. Remaining reads were mapped against genomic tRNA sequences that included 5' leader and 3' trailer sequences, as well as tRNA introns.

3.2 SSB PAR-CLIP

Flp-In T-Rex HEK293 cells (Invitrogen) were grown in high glucose DMEM supplemented with 10% (v/v) FBS, 1 mM sodium pyruvate, 100 U/mL penicillin, 100 µg/mL streptomycin, 100 µg/mL zeocin and 15 µg/mL blasticidin. Cell lines stably expressing FLAG/HA- (FH-) SSB were generated as described previously [59]. Expression of FH-SSB was induced by addition of 1 µg /mL doxycycline for 24 h. 4SU PAR-CLIP was performed as described previously, using either RNase T1 or RNase A [60]. PAR-CLIP cDNA libraries were sequenced on an Illumina HiSeq 2500 instrument. Adapter extracted reads were aligned against an in-house curated and annotated list of mature and precursor mRNAs and ncRNAs. Bioinformatic analysis was performed using a analysis pipeline based on a curated and annotated reference RNA collection, which we organized into categories, such as rRNA, tRNA, snoRNA, mRNAs, etc. This pipeline is available at https://rnaworld.rockefeller.edu/PARCLIP_suite/. T-to-C conversion frequency, indicative of binding, was calculated for each annotated category of RNA.

3.3 Bioinformatic analysis

Reads were mapped to our transcriptomic database with error distance 0 (d_0), 1 (d_1) or 2 (d_2), allowing mismatches, insertions and deletions. Assignment of reads with more than one mapping locations that belong to different RNA classes followed a hierarchical procedure reflective of the cellular abundance of each RNA class. Mature RNA sequences (e.g. fully processed tRNAs) received priority compared to precursors (e.g. pre-tRNAs), thus minimizing multimapping events. A tRNA gene was considered to be expressed when there were reads spanning the precursor/mature junctions, including exon/intron junctions for intron-containing tRNAs. For abundance reports, multimapping reads were split equally over the number of their mapping locations, and all reads mapping to edited and non-edited versions of the same tRNA transcript were summed for quantification of a given tRNA. Naming of tRNAs followed HUGO guidelines, with edited variants of reference tRNAs exhibiting the edited position and the identity of induced mismatch in their naming. The same analysis pipeline was applied to mitochondrial tRNAs, with the exception that no tRNA precursors were annotated due to continuous transcription of the mitochondrial genome. Remaining reads that did not map to any annotated transcript were mapped to the human genome. The mapped read annotation process was based on a hierarchical procedure that assigned priority to reads mapping in their entirety to mature sequences, followed by reads that spanned the precursor-mature junctions. Bioinformatic analysis was performed by custom Perl and Python scripts, all available upon request. Graphs were created in R and Prism (Graphpad).

3.4 Accession codes

The RNAseq and PAR-CLIP sequence data have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive under accession numbers GSE95683.

Chapter 4

C3PO

4.1 Introduction

Finally, recent studies provide growing evidence for the participation of a novel endoribonuclease complex in the lifecycle of tRNAs. C3PO (component 3 promoter of RISC) is a multimeric complex of the RNA-binding protein TRANSLIN (TSN) and the nuclease TRANSLIN-ASSOCIATED PROTEIN X (TSNAX), originally shown to localize to DNA break points^{36,37}. Recently, however, it was shown to promote the activity of the RNA-induced silencing complex in drosophila. Simultaneously, it was reported that TSNAX possesses RNA endonucleolytic activity both in vitro and in vivo, and the crystal structure of the C3PO apo-complex from various organisms was reported³⁸⁻⁴⁰. Its biological importance is underscored by the severe neurodegenerative phenotypes observed in mice lacking components of the complex⁴¹. Interestingly, it has also been observed that C3PO mutations in the fungus *N.crassa* and in mouse fibroblasts result in accumulation of tRNA fragments (tRFs), elevated levels of mature tRNAs and protein translation, and increased resistance to cell death-inducing agents¹⁰. This unexpected finding

suggests that C3PO might be a novel tRNA processing enzyme. Despite these studies, though, the biological targets and the details of the C3PO's biochemical activity remain elusive. Moreover, it is not yet clear whether C3PO is important for the biogenesis of tRNAs, the generation of competent stress or non-stress related tRNA fragments or if it is simply involved in tRNA turnover.

Recent work from our group and others suggests that C3PO, a multimeric complex of TRANSLIN (TSN) and the nuclease TRANSLIN-ASSOCIATED PROTEIN X (TSNAX), apart from its previously known roles in DNA-damage response and enhancement of RISC activity, possesses also a tRNA processing activity. In agreement with these observations, mutations in the components of C3PO phenocopy mutations in other tRNA processing enzymes by resulting in neurological and behavioral phenotypes⁵⁴. Also, lack of C3PO activity is associated with accumulation of tRFs¹⁰. Thus, it is possible that C3PO might be a regulator of gene expression that could bridge tRNA processing with PTGR. In this aim, I plan to apply the annotated list of tRNAs from aim 1 and the experimental framework for the study of tRNA-tRBP interactions from aim 2 to the study of C3PO, a regulator of gene expression that could possibly provide a new bridge between tRNA processing and PTGR.

C3PO notes. TRANSLIN (TSN) was initially described as a protein that would bind DNA breakpoint junctions at conserved sequences. Initial studies identified that TSN was specifically localized in the nucleus of lymphoid cell lines, suggesting that it might play a role in DNA repair, replication or recombination (Aoki et al. 1994, 1995).

TSN contains five potential protein kinase C phosphorylation sites, three tyrosine kinase phosphorylation sites (Aoki et al. 1995). TSN was detected in the cytosolic fraction of nonlymphoid and lymphoid cells, and in the nuclear ex-

tract of lymphoid cells. It was also detected with a slightly faster mobility in non-hematopoietic cells (e.g. HeLa). In conclusion, the nuclear relocalization of TSN is “selectively controlled in lymphoid lineage cells”.

In this study, TSN was identified as a ssDNA binding protein that binds conserved sequences at the breakpoint junctions in lymphoid cells that have undergone DNA translocations. A proposed role for TSN binding of these DNA loci was “DNA-unwinding, which makes these regions more susceptible to nuclease cleavage.”

“Since DNase I or S1 nuclease hypersensitive sites were observed near these breakpoints, and potential Z-DNA elements have been shown to be important in homologous recombination, it is conceivable that the above DNA structures influence chromatin structure and render a localized region of DNA more accessible to recombinase.”

“These results raise the intriguing possibility that a mechanism of active nuclear transport for TSN exists and that TSN might be associated with lymphoid specific processes, such as Ig/TCR rearrangement.”

Regarding the cytosolic function of TSN in non-lymphoid cells “one possibility is that, analogous to RecA protein, TSN may be required for repair of DNA damage caused by radiation or chemicals.”

The high expression pattern of TSN in testes, a tissue that is known to undergo extensive DNA rearrangements provides further evidence that the protein may be involved in DNA damage control.

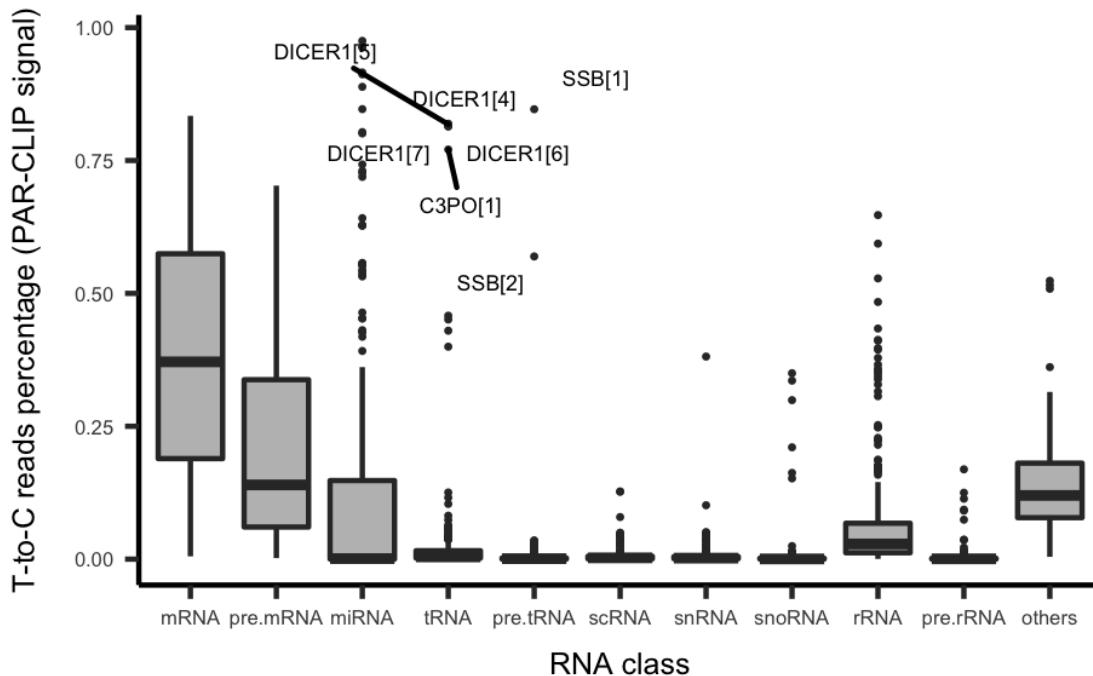


Figure 4.1: parclips. placeholder

4.2 From manny

C3PO is an RNA/DNA-binding protein complex of TRANSLIN (TSN) and TRANSLIN-ASSOCIATED FACTOR X (TSNAX), which was originally reported to co-localize with breakpoint junctions of chromosomal translocations. *Tsn*-/- mice were created with the hypothesis that its absence would affect genome stability. Interestingly, mice lacking *Tsn* exhibited learning and memory defects indicating that it is necessary for brain development or neurological maintenance. C3PO was recently reported to enhance RNAi activity in Drosophila and humans, wherein TSNAX possessed a novel and highly conserved ribonucleolytic activity. In collaboration with Dinshaw Patel's group (MSKCC), I biochemically characterized its nuclease activity and we solved the structure of Drosophila C3PO. Concurrently, the Liu laboratory (UTSW) crystallized human C3PO. Despite these recent efforts,

neither the endogenous targets nor the mechanism of C3PO assembly and nuclease activity are well understood. Given the positions of the nuclease catalytic sites, it is not clear if C3PO dynamically assembles onto its substrates, or if RNA molecules are fed inside the structure.

I have biochemical data showing that human and Drosophila C3PO possess a length-dependent ribonuclease activity (Fig. 4). C3PO nuclease activity is much reduced for RNA substrates >80 nt. The physiological importance of the length-dependent activity is not known, though one clue may come from a recent publication on a newly reported function of C3PO regarding maturation of tRNAs, which are typically between 70 and 90 nt in length. Independently, Andrew Fire and colleagues recently described small RNAs derived from tRNAs as a potential class of non-coding RNA regulators of RNA-induced silencing complex (RISC) activity. Given its purported roles in RNAi and tRNA maturation, it is tempting to hypothesize that C3PO fulfills aspects of both independently described mechanisms. Consistent with this working hypothesis, I find high crosslink evidence within the D-arm of tRNAs from a TSN PAR-CLIP (Fig. 5). These results indicate that TSN associates with specific tRNA hairpin loops containing an invariant sequence, a sequence I also find in the mRNAs that TSN binds. I plan to determine and characterize the endogenous RNA substrates of TSN and C3PO. Since TSN can form multimeric complexes without TSNA χ it is plausible that specific RNAs are binding targets of only TSN, as opposed to the C3PO complex. As the relationship between TSN-only vs. C3PO complexes is not well characterized, I have created stable cell lines that express TSN, TSNA χ , or the entire C3PO complex in an effort to determine whether there are target overlaps between the two types of TSN complexes. I will biochemically validate the RNA-protein interactions identified, prioritize these targets by determining TSN or C3PO enrichment with

RNAs and develop assays that, depending on the RNA category, functionally define the role of TSN and/or TSNA_X. I will initially focus on assays that investigate RNA stability, RISC-dependence, and tRNA involvement, since these areas reflect the current understanding in the field. In future studies, I plan to characterize C3PO-containing mRNP and tRNP complexes in order to better understand its role among other PTGR components and processes.

4.3 TSN binds tRNAs

Our laboratory has identified a limited set of the RNA targets of human C3PO by performing PAR-CLIP analysis of its components. Expression of TSNA_X alone did not yield any crosslinked RNAs – presumably due to its nuclease activity. However, PAR-CLIP with TSN did yield crosslinked RNAs. Preliminary analysis of this dataset shows that TSN exhibits high crosslinking efficiency to tRNAs, as well as mRNAs. In particular its consensus targets include the conserved UGGU motif of the dihydrouracil (DHU) stem-loop of tRNAs. A similar motif is also observed in the most prominent mRNA targets of TSN (Figure 5).

4.4 C3PO possesses a length- and structure-dependent endonucleolytic activity

4.5 Biochemical characterization of C3PO's tRNA processing activity

I plan to repeat the PAR-CLIP analysis of TSN, using the tools and expertise obtained from aims 1 and 2 in order to validate its interaction with tRNAs. At the same time, in order to gain insight into the substrates of the full complex, I plan to perform PAR-CLIP studies after co-expression of TSN with a catalytically inactive form of TSNA_X, due to a single amino acid mutation at its active site⁴⁰. Based on structural studies, a catalytically inactive TSNA_X is expected to form a similar stoichiometric complex with TSN as its wild-type form and bind its natural RNA targets, without being able to cleave them, thereby allowing their identification by PAR-CLIP. The targets identified by PAR-CLIP will be validated, in part, by *in vitro* nuclease assays so as to determine the details of the catalytic mechanism of C3PO on its substrates.

4.6 Functional validation of C3PO's targets

The impact of C3PO's activity on its targets will be examined in over-expression and knockdown experiments. I plan to perform HydroRNAseq and RIP-Seq studies to investigate the impact of C3PO expression levels on the stability of its RNA targets, and to further dissect its biological function. Lack of C3PO activity has been linked with elevated levels of mature tRNAs and pre-tRNA fragments¹⁰. At

the same time, C3PO activity is suggested to promote activity of RISC leading to enhanced RNA silencing³⁸. If the former is true, I expect to observe an accumulation of pre-tRFs and mature tRNAs in knockdown versus control or over-expression. If the latter holds, then I expect a change in the stability of the mRNA targets of C3PO. Finally, if time permits, I intend to study the effect of C3PO's activity on the protein level, by performing western blot and mass-spectrometry studies of the proteins whose mRNAs will be validated targets.

4.7 C3PO summary report 2014

Performing PAR-CLIP on C3PO confirmed its role as a bona fide tRNA binding protein. The evidence for this is twofold: a) tRNAs collect the largest number of sequenced reads, 31%, with the 2nd most represented class, mRNAs, collecting only 9% (table 1), b) tRNAs represent the largest percentage of PAR-CLIP clusters both within the total number of clusters, as well as within the top 100 clusters (ranked by read abundance) (figures 1, 2, table 2). Of note, my PAR-CLIP results seem to bring into question previous reports that implicated C3PO in enhancing RISC activity and promoting miRNA-mediated gene silencing, since only 4% of the total clusters map to miRNA and 3' UTRs, respectively (table 2). However, 5' UTRs represent 22% of all clusters, a result that comes as a surprise, as C3PO has not been reported to interact with 5' UTRs before. I have also determined the tRNA binding motif of C3PO (figure 3), using motif prediction algorithms (GIMSA and MEME). Currently I am carrying out RIP-seq experiments in order to validate and rank the targets of C3PO.

In order to investigate the biochemical function of C3PO, I am trying to identify other interacting proteins. A preliminary co-immunoprecipitation analysis yielded

no clear candidates, which suggests that either C3PO functions in no close association with any other protein or that the conditions of immunoprecipitation were not appropriate. At the moment, I am repeating the experiment at different salt concentrations, as well as in the presence of a cell-permeable crosslinking agent (DSP). I hope that in this way I will be able to stabilize the possible interactions of C3PO with other proteins, which I will then identify by using the mass-spectrometry core facility. Moreover, I have already carried out *in vitro* endonuclease cleavage assays, using recombinant C3PO. These experiments confirmed that C3PO cleaves *in vitro* transcribed tRNAs in a length- and structure-dependent manner. I will carry out similar studies using tRNA targets predicted from PAR-CLIP and confirmed by RIP-seq.

Since C3PO has been previously implicated in a plethora of biological processes, I am interested in elucidating the processes in which it partakes. For this purpose, I am carrying out a gene ontology analysis of the PAR-CLIP targets. Also, I have performed siRNA knockdown experiments against the two components of the C3PO complex, TSN and TSNA. I have validated commercially available antibodies, designed siRNAs, and have successfully knocked down C3PO in HEK293 cells by 3- to 5-fold. As the next step, I will perform two series of RNA-sequencing experiments in the context of C3PO knockdown: mRNA-seq (poly A selection) and HydroRNAseq, to determine the effect of C3PO expression on the mRNA and tRNA population, respectively.

Finally, a series of reports implicate tRNA metabolism in stress responses. Specifically, it has been suggested that upon cellular damage, tRNA endonucleolytic cleavage leads to translational arrest via the accumulation of tRNA halves. I subjected HEK293 cells to treatment with sodium arsenite (an inducer of oxidative stress), isolated RNA, and then performed small RNA sequencing that confirmed

the accumulation of stable 5' tRNA fragments. As a control, the miRNA or snRNA population remained largely unchanged. I confirmed these results by northern blots. Stress-induced endonucleolytic cleavage of tRNAs has been reported to be a function of the nuclease angiogenin. Nevertheless, angiogenin is not appreciably expressed in our cell culture system, neither in normal growth conditions nor upon oxidative stress. Therefore, it is intriguing to assume that C3PO might play a previously uncharacterized role in the metabolism of tRNAs during stress. For this purpose, I have carried out a PAR-CLIP experiment under conditions of oxidative stress and I am currently awaiting the sequencing results.

4.8 C3PO summary from annual report 2015

C3PO is a multimeric complex of the RNA binding protein TRANSLIN and the RNA endonuclease TRAX. A plethora of functions have been assigned to this complex, such as a role in the repair of DNA breaks, enhancement of RNAi activity, and tRNA processing. By performing PAR-CLIP I had previously identified C3PO as a tRNA binding protein, and showed that it does not bind to miRNAs or 3'UTRs of mRNAs, arguing against a role in RNAi. Except for tRNAs, the other main target of C3PO was 5' UTRs, something unusual for mRNA binding proteins apart from translation initiation factors.

In order to investigate the effect of C3PO in mRNA and tRNA stability, I performed mRNA and tRNAseq upon knockdown as well as induction of C3PO. To my surprise no significant effect was observed in either class of RNAs. These results can be interpreted in two ways. First, since C3PO is an enzymatic complex, perhaps partial loss of function conferred by siRNA knockdown is not sufficient to yield an observable phenotype, at least under the tested experimental conditions. Sec-

ond, C3PO has an effect on translation rather than on mRNA or tRNA stability. I am testing the former hypothesis, by performing sequencing on RNA isolated from Translin and Trax knockout flies, and the latter by performing polysome analysis and mRNA reporter translation assays. Finally, in order to elucidate the molecular mechanism of action by C3PO, I have engaged in a collaboration with the lab of Dinshaw Patel at MSK, trying to obtain a crystal structure of C3PO with a minimal or full-length RNA target identified in my PAR-CLIP analysis.

4.9 C3PO summary from annual report 2016

Despite extensive efforts in collaboration with the Patel lab (MSK), we have not been able to obtain a crystal structure of C3PO (complex of TRANSLIN and TRAX) bound to its RNA targets, which I have previously identified by PAR-CLIP. I have performed extensive electrophoretic mobility shift assays that have specified optimal substrates (e.g. full length tRNA, single stranded GGU repeats of various lengths, short RNA stemloops containing C3PO's putative binding motif TGGW - W= A or T). Using these substrates, the Patel group obtained well-diffracted crystals of TRANSLIN crystallized in presence of either single stranded RNA sequence 5'-(UG)3U(UG) or the tRNA dihydrouridine stem loop sequence (5'-GUAUAGUGGUUAGUAC) that belong to two different space groups. The structures were solved at 2.2 Å and 2.74 Å resolutions, respectively, and revealed minor differences in the arrangement of octameric TRANSLIN, but no bound RNA substrate in the expanded hollow interior of the closed-barrel structures. They have also obtained small crystals of the truncated wild-type C3PO in the RNA-free state, which are currently being optimized to attain diffraction quality.

In parallel, I have been characterizing the *in vivo* effects of C3PO, by loss-of-

function studies. Unexpectedly, I observed that siRNA knockdowns of TRANSLIN led to increased translation of mRNA targets identified by PAR-CLIP. At the same time, more than 30 ribosomal proteins were upregulated upon TRANSLIN knock-down. Therefore, there is evidence that C3PO has a role in regulating translational efficiency. I have confirmed by RNAseq that C3PO is ubiquitously expressed at moderate to high levels. Therefore, there are reasons to think that C3PO might be involved in fundamental and global regulation of translation, either by acting directly at the translational process or indirectly, by modulating tRNA levels. In fact, the binding properties of C3PO towards specific 5' UTRs of some mRNAs reflect those of known translational factors, inasmuch as they are usually single, short binding sites with GC-content significantly higher than randomly sampled, size-matched sequences from the 5' UTR background context¹⁹.

Since TRANSLIN knockdowns did not have a global pronounced effect on tRNA levels, I reasoned that the depletion achieved by the siRNA methodology (70% reduction in protein level) was not impactful enough to result in an observable effect on global tRNA levels. Therefore, I am obtaining CRIPSR knockouts²⁰ of TRANSLIN and TRAX, which I will start characterizing after the submission of my tRNA manuscript.

Although previous attempts to co-immunoprecipitate (co-IP) interactors of C3PO had been fruitless, I have optimized the IP conditions, managing to observe stoichiometric interactors. By western blot analysis of the interactors I identified a component of the RNase P complex, which renders further validation in the involvement of C3PO in tRNA processing. Having established appropriate IP conditions, I am scaling up and preparing for proteomic analysis of the C3PO interactors. Finally, I am setting up stable isotope labeling by amino acids in cell culture (SILAC²¹) experiments for C3PO wild-type versus C3PO knockout cell lines in

order to confirm C3PO's involvement in the regulation of translation.

Finally, recent publications have provided evidence for abundant stable tRNA fragments with various functions, but to date no factor has been identified as responsible for the biogenesis of these fragments²². To examine whether C3PO is involved I have prepared small RNA sequencing datasets upon gain- and loss-function of C3PO, by enriching for 5'-end phosphate containing small RNAs in the appropriate size range of 19-35 nts, such as tRNA fragments and miRNAs.

Chapter 5

Things that didn't work

As I had proposed last year, I am ultimately interested in applying the expertise and data obtained from tRNA sequencing towards understanding the interdependence (if any) between tRNA and mRNA abundance on a codon-by-codon level, as well as determine whether the balance between the two affects translational rates and mRNA stability. First, I retrieved all annotated isoforms of human mRNA transcripts, and obtained computationally an expression-weighted frequency table for all codons, by multiplying the incidence of every codon per every transcript with the abundance of each transcript. After accounting for wobble effects, I examined the relationship between codon and respective anticodon abundance. Interestingly, even though for the majority of codons, an increase in their abundance is mirrored by an increase in their cognate anticodon, there are outlier codons for which there is a discrepancy, with the codon being unusually more or less abundant than its anticodon. This led to the idea that the codon-anticodon abundance imbalance might have an observable effect on the translational efficiency of mRNA harboring such codon-anticodon pairs. To investigate this possibility I focused on ribosome profiling, a technique that determines the translational efficiency of mRNAs into

proteins by quantifying the number of ribosome-protected mRNA fragments by RNAseq following isolation of actively translating ribosomes¹². Recently, a very deep ribosome profiling dataset in HEK293 was published¹³. I started performing analysis using this dataset. At a first step, I created scripts that allow for general quality control of the data. With these I confirmed the read length distribution as well as the characteristic phasing of the 5' ends of ribosome profiling reads with respect to the first translation frame^{12,14}. At the same time, I was able to quantify ribosome protected fragments per every single codon. This now allows me to correlate translational efficiency with the relationship between codon and anticodon abundance. Finally, I am investigating whether there is a connection between the translational efficiency of an mRNA and its stability, since it has been previously postulated that stalled ribosomes on inefficiently translated messages can trigger mRNA degradation resulting in decreased mRNA abundance¹⁵⁻¹⁸.

Finally, recognizing that I can obtain tRNA expression and abundance profiles in a more reliable manner than previously published ones, such as tRNA arrays, I decided to approach long-standing fundamental questions about the nature of the genetic code. Particularly, I was interested in combining mRNA abundance data, and translational efficiency data obtained by ribosome profiling, with tRNA abundance data in order to understand translational dynamics better. The hypothesis behind this work was that codon biases observed in specific contexts could be directed by the availability of tRNAs.

For this reason, I turned my attention towards ribosome profiling. Since our lab is still in the process of establishing this methodology, I focused on already published data about our cell culture system. I developed a series of scripts that allow me to perform tailored analysis of the data. Although this analysis is far from complete, there have been some important observations already. First, as

previously shown by other groups, I showed (using our own RNAseq data) that although highly expressed mRNAs are also highly occupied by ribosomes, there are indeed cases where transcripts are efficiently transcribed but inefficiently translated and vice versa. This led to the hypothesis that specific tRNA availability may be the bottleneck in the translation of some transcripts. Using my tRNA abundance data, and after accounting for wobble effects, I computed an anticodon scoring index, which is essentially the weighted sum of the relative frequency of all anticodons required to decode a given open reading frame. Traditionally, it has been postulated that efficiently translated messages harbor common codons, reflecting codon optimization. Surprisingly, my preliminary data indicate that the most highly represented genes in ribosome profiling (top 10%) rank in the bottom 10% in anticodon score index. At the same time, these genes are highly transcribed, as evidenced by mRNAseq data. If these results are true (and not an experimental artifact, which I will validate by performing further replicates), then this would argue towards a lack of selective pressure for optimized codons in highly translated genes. One could then argue that increasing transcription rates for these genes leads to high protein expression levels, obviating the need for codon optimization at the DNA level.



Distinguishing Core and Holoenzyme Mechanisms of Transcription Termination by RNA Polymerase III

Aneeshkumar G. Arimbasseri, Richard J. Maraia

Intramural Research Program, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, USA

Transcription termination by RNA polymerase (Pol) III serves multiple purposes; it delimits interference with downstream genes, forms 3' oligo(U) binding sites for the posttranscriptional processing factor, La protein, and resets the polymerase complex for reinitiation. Although an interplay of several Pol III subunits is known to collectively control these activities, how they affect molecular function of the active center during termination is incompletely understood. We have approached this using immobilized Pol III-nucleic acid scaffolds to examine the two major components of termination, transcription pausing and RNA release. This allowed us to distinguish two mechanisms of termination by isolated *Saccharomyces cerevisiae* Pol III. A core mechanism can operate in the absence of C53/37 and C11 subunits but requires synthesis of 8 or more 3' U nucleotides, apparently reflecting inherent sensitivity to an oligo(rU-dA) hybrid that is the termination signal proper. The holoenzyme mechanism requires fewer U nucleotides but uses C53/37 and C11 to slow elongation and prevent terminator arrest. N-terminal truncation of C53 or point mutations that disable the cleavage activity of C11 impair their antiarrest activities. The data are consistent with a model in which C53, C37, and C11 activities are functionally integrated with the active center of Pol III during termination.

Transcription from bacteria to humans is carried out by evolutionarily related multisubunit RNA polymerases. Eubacteria and archaea contain a single RNA polymerase (Pol) comprised of 5 and 12 subunits, respectively, while eukaryotes contain Pol I, II, and III, with 14, 12, and 17 subunits, respectively, that transcribe different classes of genes (1, 2).

Termination of transcription helps delimit gene boundaries, enables recycling of polymerases that are costly to produce, and is a point of gene regulation (3–5). Transcription termination is a two-stage process that begins with recognition of a termination signal followed by pausing and dissociation of an otherwise highly stable and processive elongation complex (EC) (5–10).

Escherichia coli RNA polymerase (RNAP) uses two mechanisms of termination, factor dependent and factor independent; the latter is also known as intrinsic termination (10). The factor-dependent mechanism requires Rho, an RNA binding protein that recognizes C-rich sequence in the nascent transcript still attached to elongating RNAP (11). Intrinsic termination requires nothing more than the 5-subunit RNAP and a specific bipartite termination signal comprised of an oligo(dA-rU) hybrid and an adjacent upstream hairpin structure in the nascent RNA (10). Elegant investigations have uncovered the mechanisms at work within the bacterial RNAP active center during termination (10).

Pol I, II, and III use different signals in DNA to initiate termination processes that presumably converge on a similar mode of EC dissociation, with release of the DNA and RNA by the active centers. Less information is available on the later stage of termination than on the initial signaling phase for the eukaryotic polymerases (5, 6).

Termination by Pol II can occur by two mechanisms that use distinct termination signals and unique sets of factors. Termination by Pol I and Pol II resembles the factor-dependent mechanism for RNAP, involving RNA binding factors that associate with the nascent transcript attached to the elongating polymerase, followed by subsequent termination downstream (6, 12). In contrast, Pol III termination occurs within the termination signal, an oligo(dT) tract in the DNA (13, 14). The actual signal is oligo(dA)

on the template strand (see below). This is sufficient to halt Pol III and cause release of a nascent RNA containing a 3'-terminal oligo(U) copy of the terminator.

Human cell RNA content and doubling time suggest that Pol III is capable of producing ~100 transcripts/gene/min (2×10^7 ribosomes/200 5S genes/cell/16 h), reflective of efficient termination and reinitiation and comparable to the rate of 2 to 4 transcripts/s in yeast (15, 16, 17). Indeed, termination and reinitiation by Pol III appear to be mechanistically linked (18). Understanding the mechanisms that underlie this efficiency is important in its own right and also because Pol III is activated in many cancer cell types (19–22). Pol III contains a stable heterodimer, C53/37, whose heterodimerization domains bear homology to TFIIF α/β (2, 23). While these domains reside on a surface of Pol III near the cleft around incoming DNA (23–26), other parts of C53/37 apparently reach into the active center, where they appear to be involved in elongation and termination (26–28). C53/37 is also required for reinitiation along with C11 (18), a two-domain subunit similar to Rpb9 and TFIIS (29). The TFIIS-like domain of C11 is responsible for the intrinsic transcript 3' cleavage by Pol III that occurs during elongation pausing, proofreading, and termination (28, 30–33). How the activities of C53/37 and C11 contribute to termination remains enigmatic, especially since pausing appears largely to be due to C53/37 and the RNA 3' cleavage activity of C11 is not required for reinitiation (18).

C53/37 has been shown to slow Pol III elongation on the SUP4

Received 26 December 2012 Returned for modification 17 January 2013

Accepted 31 January 2013

Published ahead of print 11 February 2013

Address correspondence to Richard J. Maraia, maraiar@mail.nih.gov.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/MCB.01733-12>.

Copyright © 2013, American Society for Microbiology. All Rights Reserved.

[doi:10.1128/MCB.01733-12](https://doi.org/10.1128/MCB.01733-12)

Downloaded from <http://mcb.asm.org/> on April 16, 2016 by Rockefeller University

Arimbasseri and Maraia

tRNA gene, including at the 3' oligo(dT) tract, which increases dwell time and therefore termination (18). In particular, although C11 was implicated in termination 15 years ago (29) and later data indicated termination-related functions (32, 33), the extent of its involvement and of its RNA 3' cleavage activity in the termination process has remained unclear.

In this study, we examined termination by isolated Pol III and the roles of C53/37 and C11 in this process. This allowed us to distinguish two mechanisms by which Pol III can terminate. The first is a property of Pol III lacking C53/37 and C11 (Pol III Δ) that requires tracts of ≥ 8 T nucleotides for optimal function, which we refer to as a core mechanism of termination. The holoenzyme mechanism allows termination at ≤ 6 T nucleotides but requires C53/37 and C11 for optimal function. Further, the holoenzyme mechanism involves two components, slowing of elongation and prevention of terminator arrest, that collectively require C11, C53, and C37. Optimal slowing through oligo(dA) requires C37 and the dimerization domain of C53 in addition to C11. Antiarrest requires C37, the N-terminal domain of C53, and the acidic side chains of C11 needed for RNA 3' cleavage. A model of Pol III termination argues that from the position on the Pol III jaw, the dimerization domains of C53/37 sense the termination signal as it approaches the active center and clamp down, like a disc brake, to slow elongation (24). More recently published data suggest that parts of C53, C37, and C11 extend into the catalytic center (26–28). The present data support a revised model in which C53, C37, and C11 also work together in the Pol III active center.

MATERIALS AND METHODS

Templates. Single-stranded DNAs were purchased from IDT or Eurofins/MWG/Operon and purified by polyacrylamide gel electrophoresis (PAGE). For tailed templates, both strands were annealed, followed by another round of PAGE purification. RNA primers were obtained from IDT. A list of oligonucleotides used is provided in Table S1 in the supplemental material.

Pol III purification. Yeast Pol III was purified from a *Saccharomyces cerevisiae* strain carrying an N-His-FLAG-tagged RPC128 gene, and Pol III Δ was purified from a strain carrying N-His-FLAG C128 in which RPC11 was deleted and complemented with *Schizosaccharomyces pombe* C11 (27). In short, 80 g of cells was lysed by a bead beater in cold lysis buffer (40 mM Na-HEPES, pH 7.8, 5% glycerol, 10 mM 2-mercaptoethanol, 0.5 M NaCl, 7 mM MgCl₂, and protease inhibitors). Lysate was cleared by centrifugation in a Sorvall SLA-1500 rotor at 10,000 rpm for 30 min and then subjected to ultracentrifugation at 100,000 relative centrifugal force for 1 h at 4°C. The top layer was recovered, avoiding the murky layer above the pellet. The S100 was subjected to affinity purification using nickel-nitrilotriacetic acid (Ni-NTA) resin (Qiagen). Pol III was eluted with 20 mM Na-HEPES, pH 8, 20% glycerol, 500 mM NaCl, 10 mM 2-mercaptoethanol, 7 mM MgCl₂, 100 mM imidazole, and complete protease inhibitor (Roche). Activity was assayed on tailed templates, and active fractions were pooled and stored at -70°C in small aliquots.

Cloning. RPC11 was cloned into pGEX-4T-1. To make the D91A.E92A mutant, pGEX-4T-1-C11 was amplified using primers harboring the mutation and Phusion high-fidelity DNA polymerase (NEB), followed by DpnI digestion to digest parent plasmid before transforming DH5 α ultracompetent cells. RPC53 was cloned into a pET28a vector harboring a tobacco etch virus (TEV) site following a 6×His tag. To make C53-Nt Δ , RPC53, corresponding to amino acids 280 to 422, was PCR amplified and cloned into TEV-pET28a. N-terminally FLAG-tagged C37 was cloned into pET21b lacking the 6His sequence.

Pol III subunit expression and purification. pGEX-4T-1 C11 and pGEX-4T-1 D91A.E92A C11 plasmids were expressed in Rosetta (DE3) pLysS cells with 100 μ M isopropyl- β -D-thiogalactopyranoside (IPTG) for

3 h at 37°C. Cell lysate was prepared in lysis buffer (50 mM K-HEPES, pH 7.8, 500 mM NaCl, 5% glycerol, 10 mM 2-mercaptoethanol, 1%Triton X-100, and protease inhibitors). Cleared lysate was incubated with glutathione-Sepharose (GE Healthcare) equilibrated with lysis buffer, washed with lysis buffer, and then washed with wash buffer (WB; 20 mM K-HEPES, pH 7.8, 200 mM NaCl, 10% glycerol, 10 mM 2-mercaptoethanol). Protein was eluted by overnight digestion with thrombin at room temperature in WB. Thrombin was inactivated by 2 mM benzamidine-HCl, and the purified proteins were stored frozen at -70°C.

To purify C53/37 and C53-Nt Δ /C37 complexes, pET28-nH6TEVC53 or pET28-nH6TEVC53-Nt Δ plasmid was cotransformed along with pET21-nFLAGC37. Transformants were grown in LB media containing ampicillin (100 μ g/ml) and kanamycin (50 μ g/ml) and induced with 0.5 mM IPTG. Cleared lysate in lysis buffer (50 mM Na-HEPES, pH 8, 200 mM NaCl, 5% glycerol, 10 mM 2-mercaptoethanol, and protease inhibitors) was passed through Ni-NTA resin (Qiagen), and the complex was eluted with 20 mM K-HEPES, pH 8, 200 mM NaCl, 10 mM 2-mercaptoethanol, 10% glycerol 300 mM imidazole. Peak fractions were pooled, 40 μ l of Turbo TEV protease (Eaton Biosciences Inc.) was added, and the mix was dialyzed against elution buffer lacking imidazole. Samples were then passed through Ni-NTA to remove tagged fragments and undigested proteins and passed through an SP Sepharose (GE Healthcare) column equilibrated with dialysis buffer. C53/37 was eluted at 300 mM NaCl. The C53-Nt Δ /C37 complex was not retained by the column and was instead recovered from the flowthrough fraction. Peak fractions of wild-type (WT) complex and flowthrough fraction of C53-Nt Δ /C37 were pooled, and buffer (20 mM Tris-Cl, pH 8, 200 mM NaCl, 10 mM 2-mercaptoethanol, 10% glycerol, and protease inhibitors) was exchanged using a PD10 column (GE Healthcare). Finally, the C53/37 and C53-Nt Δ /C37 complexes were passed through a Q-Sepharose (GE Healthcare) column and eluted with 300 mM NaCl.

EC assembly and transcription was carried out as described previously (37). The 6×His-tagged Pol III or Pol III Δ was immobilized on Ni-NTA resin (Qiagen) and washed with EC buffer lacking Mg²⁺ (ECB; 20 mM Na-HEPES, pH 8, 3 mM β -mercaptoethanol, 5% glycerol, 0.1 mg/ml bovine serum albumin [BSA], 100 mM NaCl). Template DNA strand (70 pmol) and 5'-end-labeled (³²P) RNA primer (70 pmol) were annealed and incubated with immobilized WT Pol III (designated Pol III-WT) or Pol III Δ for 10 min at room temperature. Nontemplate DNA was added (140 pmol) and incubated for 10 min, followed by 3 washes, each with 30 column volumes (CV) of ECB, followed by 30 CV of ECB with 1 M NaCl and a final wash with 30 CV of ECB.

For all transcription reactions except those depicted in Fig. 3 and 4C, the ECs were carried out in ECB with 7 mM Mg²⁺ and 0.5 mM nucleoside triphosphates (NTPs) for 10 min at 25°C; transcription reactions depicted in Fig. 3 and 4C were done for 30 min. Three micromolar NTP was used for experiments shown in lanes 5 and 6 of Fig. 3 and lanes 5 to 8 and 12 to 16 of Fig. 4C.

For cleavage reactions, assembled ECs were treated with ECB containing 7 mM Mg²⁺. For transcription, the ECs were treated with ECB with 7 mM Mg²⁺ and 0.5 mM NTPs for 10 min unless otherwise noted.

Tailed template transcription reactions were in ECB with 7 mM Mg²⁺, 0.5 mM NTPs, and 20 ng template for 30 min before stopping with SDS plus EDTA. Where applicable, Pol III Δ was preincubated with proteins for 20 min before addition of template DNA.

Samples were precipitated, dissolved in formamide with dye, resolved on 15% sequencing gels, dried, exposed to a phosphorimager screen, and visualized using Typhoon (GE Healthcare). Image files were analyzed and profiles made using Multigauge software (Fuji). Profiles were exported to Excel and converted to line charts. Profiles of different lanes of the same gel were normalized to the bands of maximum intensity.

RESULTS

We examined termination by purified Pol III in the absence of transcription factors. We made use of the ability of isolating WT

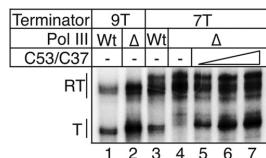


FIG 1 Two mechanisms of Pol III termination. Tailed template transcription was done with WT Pol III or Pol III Δ and templates with a 9T or 7T terminator. Reactions were in the presence of 0.5 mM (each) ATP, GTP, and UTP with 0.025 mM [α -³²P]CTP. T, terminated products; RT, read-through transcripts. Increasing amounts of purified recombinant C53/37 was included in lanes 5 to 7.

Pol III and Pol III Δ , which lacks C53/37 and C11, in parallel from *S. cerevisiae* (18). We initially used tailed templates that differ only in the length of their terminators, oligo(dT) tracts of 9 T nucleotides (9T) or 7T. Transcripts shown in Fig. 1 reflect the recognition of the oligo(dT) terminator (T) or read-through (RT). Pol III termination efficiency, as reflected by ratios of T and RT bands, can be complicated by multiple factors when using tailed templates in experiments such as that in Fig. 1 (34). The band corresponding to T might reflect pausing or termination, because tailed templates can produce extended RNA-DNA hybrids that fail to release and therefore exhibit low apparent termination efficiency (34). Nonetheless, it was unexpected that Pol III Δ would produce a strong band at the 9T terminator comparable to that of Pol III-WT (Fig. 1, lanes 1 and 2). This suggested a C53/37-independent mode of termination at 9T. Similar to previous findings, Pol III Δ failed to terminate at a 7T terminator and instead read through, while Pol III-WT terminated at 7T (lanes 3 and 4) (18, 29). Addition of C53/37 restored termination at 7T (lanes 4 to 7), as expected (18, 27). Also as expected, our Pol III Δ lacks intrinsic transcript cleavage activity, because our experiments showed cleavage products for Pol III-WT but not Pol III Δ (Fig. 2B, lanes 5 to 8, and C, lanes 3, 4, 9, and 10). Because the RNA in the absence of template was not cleaved by Pol III-WT under otherwise identical conditions (not shown), the CP bands in Fig. 2B and C reflect the intrinsic transcript cleavage activity of Pol III-WT. We show later that intrinsic cleavage activity can be restored to Pol III Δ by recombinant C11.

The data suggested that Pol III can terminate by two mechanisms, one of which we attribute to a core mechanism (i.e., by Pol III Δ) manifested at 9T tracts, while the holoenzyme mechanism operates in the presence of C53/37 at shorter T tracts. Note that we do not refer to Pol III Δ as a core enzyme. This is because it contains the C82/C34/C31 subcomplex, and the “core” terminology has been used to refer to Pol III lacking both C82/C34/C31 and C53/37 (35, 36). Because C82/C34/C31 is an initiation complex that has not been implicated in the termination mechanism, we consider it appropriate to refer to the termination properties observed for Pol III Δ as a core mechanism of termination.

Pol III Δ exhibits transcript release deficiency. As noted above, appearance of an RNA band whose length corresponds to the oligo(dT) tract, as does the T band in Fig. 1, reflects terminator recognition, i.e., pausing, but does not demonstrate termination. We therefore examined the transcript release properties of Pol III Δ and Pol III-WT using Pol III-immobilized ECs (37). This should not only avoid the potential for extended RNA-DNA hybrids that can occur with tailed templates but also allows ready

separation of released from nonreleased transcripts and their single-nucleotide resolution by electrophoresis.

For EC assembly, 80-nucleotide (nt) DNA and 10-nt 5'-radio-labeled RNA sequences were used (Fig. 2A). ECs that differ in the position or length of the oligo(dT) relative to the RNA primer were examined (Fig. 2A). In template 1, the terminator is 12 nt from the RNA primer, while in template 2 the terminator immediately follows the RNA primer. Templates 1 and 2 have the same 9T terminator (indicated as A9 on the template strand in Fig. 2A) and downstream sequences, and Pol III ECs form on both with comparable efficiency. Transcription products were separated into released and bound fractions (Fig. 2B and C).

For both templates, the RNAs produced by Pol III-WT and Pol III Δ differ in length and release efficiency (Fig. 2B and C). First, the released RNAs differ in length distributions, with Pol III Δ transcripts being longer. With template 1 the most abundant transcripts differed in length by 1 nt, with a peak at 8 U nucleotides for Pol III-WT and 9 U nucleotides for Pol III Δ (Fig. 2B). Pol III Δ specifically lacked transcripts in the shorter size ranges in the released fractions that were present with Pol III-WT. This is revealed in Fig. 2D, which shows band quantitation profiles for lanes 1 and 3 of Fig. 2B. A similar disparity in released transcripts was observed on template 2 (Fig. 2C, lanes 1 and 7).

Close examination revealed that the patterns of RNAs from templates 1 and 2 differ further. The peak of released transcripts is 1 nt shorter with template 2 than with template 1 for both Pol III Δ and Pol III-WT (Fig. 2B and C). We considered that this might reflect differences in elongation on the two templates. Addition of the first few nucleotides by Pol III occurs at a lower rate than that after several nucleotides have been synthesized (38). Accordingly, Pol III would be in slow-stepping mode upon encountering the template 2 terminator but in a faster mode upon encountering the template 1 terminator. Slow- and fast-stepping modes are in accord with a model for differences in termination by Pol III Δ and Pol III-WT (18). Slower elongation was accompanied by increased Pol III termination (29).

The ability of Pol III Δ to read through the first 7T tract of the SUP4 gene terminator was attributable to the fast-stepping elongation mode, whereas addition of C53/37 appeared to induce a slow-stepping mode during termination (18). We propose that the fast mode of Pol III Δ accounts for the increased length of released RNAs relative to Pol III-WT observed for the 9T terminators of templates 1 and 2. Accordingly, we consider the shorter-range transcripts to be reflective of a slow-stepping Pol III, whereas fast stepping allows longer extension into the 9T tract. We show below that slowing elongation by lowering NTP concentrations results in shorter RNAs at the terminator.

The notable paucity of transcripts with 4 to 7 U nucleotides released from Pol III Δ appears to be due to preferential retention of these shorter RNAs in the Pol III Δ -bound fraction relative to the Pol III-WT-bound fraction (Fig. 2B, compare lanes 2 and 4). Preferential retention of shorter transcripts by Pol III Δ was more prominent with template 2 (Fig. 2C). The distribution patterns of released and bound transcripts for Pol III-WT were similar. In contrast, Pol III Δ -bound transcripts were skewed toward shorter RNAs relative to released transcripts (Fig. 2C). The data indicate that in addition to 3' lengthening of RNAs, Pol III Δ also exhibits transcript release deficiency at the terminator that manifests by quantitative and qualitative differences in bound versus released transcripts compared to Pol III-WT.

Downloaded from <http://mcb.asm.org/> on April 16, 2016 by Rockefeller University

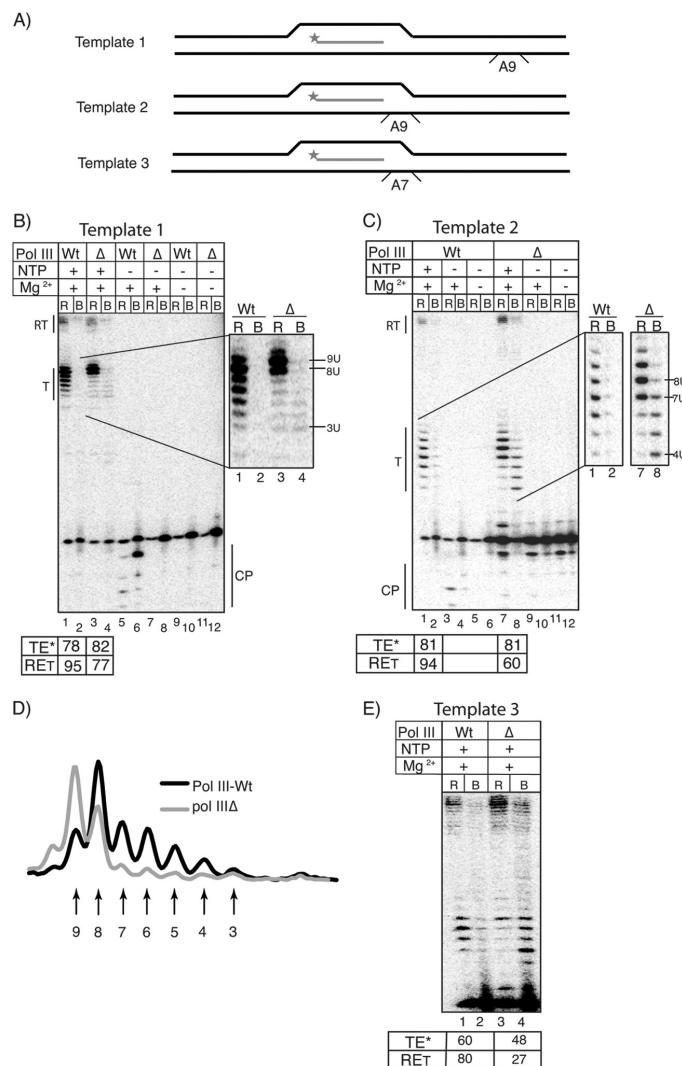


FIG 2 Pol IIIΔ exhibits two termination deficiencies, release and arrest. (A) Schematic of templates used for elongation complex (EC) assembly. Templates 1 and 2 differ only in the position of the terminator (A9 on the template strand) relative to the RNA primer (gray line; an asterisk indicates 5' ³²P). Template 1 has a terminator 12 nt from the RNA 3' end, while template 2 terminator abuts the RNA primer. Template 3 differs from 2 in that its terminator is 7T (A7 on the template); the same RNA was used for the three templates. (B) Transcription (lanes 5 to 8) activities of ECs with Pol III-WT and Pol IIIΔ on template 1. R, released; B, bound; T, terminated; RT, read-through; CP, cleavage products. Lanes 1 to 4 were enlarged; transcript bands reflecting the 3' rU length are indicated on the right. Termination efficiency (TE) and release efficiency at the terminator (RET) are listed below the lanes. (C) Same as panel B but using template 2. Enlargements of lanes 1, 2, 7, and 8 are on the right. (D) Quantitative lane tracing profiles of the T (termination) region of lanes 1 and 3 of panel B. Lengths of 3' rU residues are indicated under the corresponding peaks. (E) Same as panel B but using template 3. TE was determined by the formula [(terminator-released RNA + terminator-bound RNA)/(terminator-released RNA + terminator-bound RNA + read-through RNA bound + read-through RNA released)] × 100. RET is defined by the formula [terminator-released RNA/(terminator-released RNA + terminator-bound RNA)] × 100.

Downloaded from <http://mcb.asm.org/> on April 16, 2016 by Rockefeller University

We also examined an EC with a 7T terminator (template 3) (Fig. 2A and E). Pol III-WT terminated with efficient release of transcripts with 4 to 7 U nucleotides, retaining only a small amount in the bound fraction (Fig. 2E, lanes 1 and 2). In contrast,

only about half of the 7U-terminated RNAs were released by Pol IIIΔ, and there was more read-through as well (Fig. 2E, lanes 3 and 4), consistent with prior reports (18, 26, 29). More dramatically, a larger amount of Pol IIIΔ products with fewer than 7 U nucleo-

tides were associated with the 7T terminator complex in the bound fraction (Fig. 2E, lane 4). The data indicate that termination by Pol III Δ requires a minimum of 7 U nucleotides but is closer to optimal with longer tracts of U nucleotides, whereas Pol III-WT can terminate after synthesis of 6 U nucleotides.

To quantify the termination deficiencies of Pol III Δ we used two parameters, termination efficiency (TE), which is the percentage of total transcripts corresponding to the T band, and release efficiency at the terminator (RET). To calculate the TE of a standard transcription reaction in which the RNA primer is end labeled, one quantifies the bands corresponding to T and RT and applies them to the equation $TE = (T/RT + T) \times 100$. Consistent with Fig. 1, the TE for Pol III Δ and Pol III-WT were comparable at $\sim 80\%$ on the 9T terminators (Fig. 2B and C). In contrast to its behavior at the 9T terminator, Pol III Δ showed a significant decrease in TE relative to Pol III-WT on the 7T terminator (Fig. 2E, lanes 1 to 4). Pol III Δ and Pol III-WT differ more significantly in RET. While the TEs of Pol III Δ and Pol III-WT were comparable at 9T, RET for Pol III Δ was relatively lower (Fig. 2B and C). Pol III Δ showed even lower RET at the 7T terminator than Pol III-WT (Fig. 2E).

Pol III Δ -retained transcripts reflect terminator-arrested complexes. The Pol III Δ complexes and their unreleased RNAs were further characterized. Experiments using single NTPs or combinations thereof showed that the appearance of the terminator-associated Pol III Δ -bound RNAs required UTP but was not due to misincorporation. As expected, the isolated terminator-bound Pol III Δ transcripts could not be chased by fresh NTPs (see Fig. S2, lanes 5 and 6, in the supplemental material). Moreover, the ECs containing Pol III Δ -bound transcripts were quite stable, as they could withstand extensive washing with 1 M NaCl (see Fig. S1 in the supplemental material). The cumulative findings indicate that these terminator-associated Pol III Δ complexes can neither extend nor release their RNAs and are therefore in a state of arrest.

Two components of holoenzyme Pol III termination are elongation slowing and prevention of arrest. We hypothesized that if the terminator-associated arrest observed most prominently on template 2 was due to slow elongation, then lowering NTP concentration might also lead to increased arrest on template 1. Lowering NTP concentrations slows elongation (39, 40) and also corrects the terminator read-through defect of Pol III Δ in standard *in vitro* transcription reactions (29).

Lowering the concentration of the NTPs from 500 to 3 μ M had two effects on Pol III Δ termination on template 1 (Fig. 3), it altered the pattern of transcript length distribution and release efficiency. The increase in paused and arrested transcripts at 3 μ M NTPs (indicated by P in Fig. 3A) is evidence that the elongation rate was slowed by lowering the NTP concentration. At 3 μ M NTPs, Pol III Δ released terminated transcripts of shorter lengths than at 500 μ M in a pattern similar to that for Pol III-WT at 500 μ M NTPs (Fig. 3A and B, compare lanes 1, 3, and 5). Thus, with regard to the length pattern of transcripts released at the terminator, lowering the NTP concentration led to shorter RNAs and accordingly could compensate for the lack of C11/53/37. The data indicate that Pol III Δ advanced less far into the 9T tract when in slow mode due to lower concentrations of NTPs. This suggests that the elongation rate is a determinant of the 3' length of terminated RNAs. However, lowering the NTP concentration did not reverse the release deficiency of Pol III Δ but instead led to increased amounts of transcripts arrested at the terminator (Fig. 3A, compare lanes 4 and 6). The significant conclusion here is that

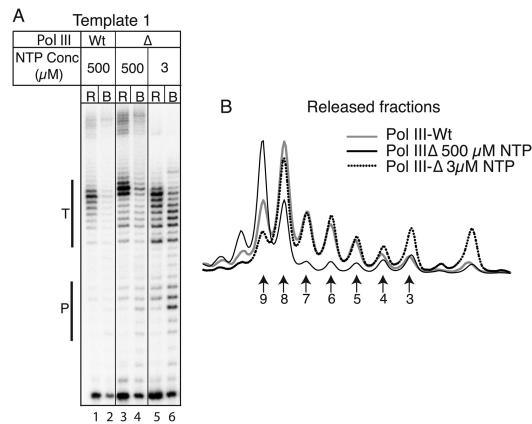


FIG 3 Slowing elongation is insufficient to correct release deficiency of Pol III Δ . (A) EC transcription reaction of WT and Pol III Δ at different NTP concentrations. EC was assembled on template 1, and transcription was done with the NTP concentrations indicated above the lanes (in μ M). Transcription reactions were for 30 min. R and B indicate released and bound fractions, respectively. P, paused/arrested transcripts; T, terminator. (B) Superimposed quantitative lane tracing profiles of lanes 1, 3, and 5 from panel A. The arrows with numbers reflect the number of 3' U nucleotides of the nascent RNA for each peak.

slowing Pol III Δ is insufficient to explain the termination differences of Pol III-WT and Pol III Δ (18). The data suggest that the mechanism by which C53/37 promotes termination involves two components, reduction of elongation rate, as previously proposed, and a second activity, prevention of terminator arrest.

We suspected that Pol III Δ terminator arrest involves backtracking (41), because RNAP tends to backtrack upon encountering a weak RNA-DNA hybrid, such as oligo(rU-dA) (40, 42–45). Consistent with this is the more prominent arrest on template 2 compared to template 1 (Fig. 2) and more arrest on template 1 with low NTPs (Fig. 3), because a slowly stepping polymerase is more prone to backtracking than a fast polymerase at a weak hybrid (40). Also consistent with this is that oligo(dA) on the template strand induces termination by Pol III, while oligo(dT) on the nontemplate strand does not (see Fig. S3 in the supplemental material). Nonetheless, the more unlikely possibility that the arrested complexes are forward translocated (46) or are in another state remains open.

C11 can rescue minimally backtracked Pol III Δ complexes. We wanted to know if C11 could rescue backtracked Pol III Δ ECs. We assembled an artificially backtracked complex with an RNA primer that has 10 nt complementary to the template plus a 2-nt mismatch at the 3' end (Fig. 4A). Biochemical and crystallographic evidence indicate that such artificially assembled ECs are similar to naturally formed backtracked complexes (37, 47–49). Transcriptional extension of the RNA primer from a backtracked Pol III Δ EC should occur only if C11 could promote cleavage of the RNA 3' mismatch, and this would represent evidence of functional C11 activity. A second way to monitor C11 activity is to detect RNA cleavage products upon addition of Mg²⁺ in the absence of NTPs. As expected, Pol III Δ alone or in the presence of NTPs and Mg²⁺ showed no significant transcription elongation (Fig. 4B, lanes 1 to 6). Likewise, addition of Mg²⁺ did not alter the

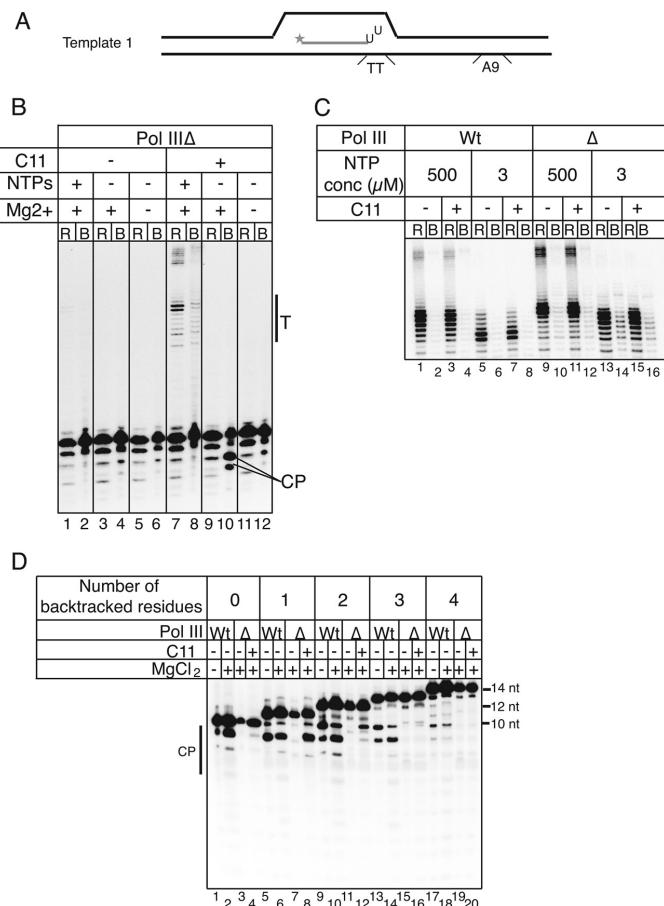


FIG 4 Cleavage-active C11 cannot prevent terminator arrest by Pol III Δ . (A) Schematic of template and RNA used for assembly of a 2-nt-backtracked complex. The 3' end of the 12-nt RNA has a 2-nt rU-dT mismatch. (B) Transcription and cleavage by backtracked ECs incubated with either buffer alone (lanes 1 to 6) or with recombinant C11 (lanes 7 to 12) for 20 min prior to addition of Mg²⁺ and/or NTPs. Lanes 1, 2, 7, and 8 show products of transcription reactions; lanes 3, 4, 9, and 10 show cleavage reactions; and lanes 6, 7, 11, and 12 show starting material after incubation with and without C11 as indicated above the lanes. R and B represent released and bound, respectively. Cleavage products (CP) are indicated with black lines. (C) Transcription reactions in the presence of C11 at different NTP concentrations as indicated; the same amount of C11 as that used for panel B was used. (D) Cleavage assay for Pol III-WT and Pol III Δ ECs with various numbers of backtracked residues. ECs were assembled with RNAs having various numbers of mismatches, and after washes, complexes were treated with Mg²⁺ and/or C11 as indicated above each lane.

RNA pattern, indicating no cleavage activity, also as expected for Pol III Δ (Fig. 4B, lanes 3 and 4). In contrast, addition of recombinant C11 to Pol III Δ led to transcription elongation in the presence of NTPs and Mg²⁺ (Fig. 4B, lanes 7 and 8), as well as production of RNA cleavage products with Mg²⁺ only (lanes 9 and 10). Incubation of recombinant C11 alone with the RNA in the absence of Pol III and template revealed no RNA nuclease activity (not shown). Together, the results indicate activity of C11 in the intrinsic cleavage reaction and its function in rescuing backtracked Pol III Δ . However, lane 8 and other data revealed that C11 alone could not prevent Pol III Δ arrest at the terminator. In support of this, Pol III Δ arrested at the terminator of template 2 was isolated, subjected to C11, and again separated into released and bound fractions, but it remained largely refractory, as most tran-

scripts persisted in the bound fraction even if fresh NTPs were included (see Fig. S2 in the supplemental material).

Failure of C11 to prevent terminator arrest in standard reactions with 0.5 mM NTPs prompted us to ask if reducing the elongation rate by lowering concentrations of NTPs would bypass a need for C53/37 and allow C11 to prevent arrest. Addition of C11 to Pol III-WT or Pol III Δ with 0.5 mM or 3 μM NTPs had little, if any, effect on terminator arrest (Fig. 4C). The results suggest that a combination of slow elongation plus C11 does not reconstitute the holoenzyme termination mechanism and that C53/37 is also required, consistent with an antiarrest function in addition to elongation rate reduction.

Pol III Δ +C11 does not wholly reconstitute the cleavage properties of Pol III-WT. Resistance of Pol III Δ terminator-ar-

rested complexes to C11 raised the possibility that they are extensively backtracked, i.e., more so than the 2-nt-backtracked complex that was shown to be a C11 substrate in Fig. 4B. This was considered, because while TFIIS-mediated cleavage of a 2-nt-backtracked complex occurs proficiently, more extensively backtracked complexes are poorer substrates (50). Also, bacterial RNAP uses two intrinsic transcript cleavage factors, GreA and GreB, which differ in length of RNA cleavage products and anti-arrest activity (51). We assembled backtracked Pol III Δ complexes with mismatches of 0, 1, 2, 3, or 4 nt at the 3' end of the RNA, in addition to their 10 nt of complementarity, and subjected them to C11-mediated cleavage (Fig. 4D). In addition to showing that extensions of >2 nt were progressively poorer substrates for C11, the results additionally revealed that Pol III Δ +C11 did not completely reconstitute the cleavage properties of Pol III-WT. First, Pol III-WT readily cleaved the RNA primer with no mismatch, while Pol III Δ +C11 was deficient (Fig. 4D, lanes 1 to 4). This deficiency was not due to a limited quantity of C11, since we had titrated the samples to an amount that yielded maximum activity on this and other substrates. Second, despite identical parallel preparations and dialyzing of Pol III Δ and Pol III-WT, Pol III Δ +C11 reproducibly required addition of Mg²⁺ for cleavage activity while Pol III-WT did not (Fig. 4B, lanes 9 to 12, and D, lanes 5 to 12). Pol III-WT efficiently cleaved the 0-, 1-, and 2-nt-backtracked complexes with progressively less activity on the 3- and 4-nt-backtracked complexes (Fig. 4D) and no activity on 5-nt-backtracked complexes (data not shown). While Pol III Δ +C11 showed robust Mg²⁺-dependent cleavage activity on the 1-nt-backtracked complex, comparable to that of Pol III-WT, its activity on the 2-nt backtracked complex was reproducibly less than that of Pol III-WT (Fig. 4D, lanes 9 to 12). Most of the deficiency of Pol III Δ +C11 on the 2-nt-backtracked complex was caused by the failure to produce the 9-mer RNA cleavage product, whereas Pol III-WT produces this as a major product (Fig. 4D, lanes 10 and 12). Again, this was not due to limitation of C11, reaction time, or ability to detect cleavage products. Pol III Δ +C11 showed no activity on \geq 3-nt-backtracked complexes (Fig. 4D, lanes 13 to 20).

The cumulative data indicate that Pol III Δ +C11 does not completely reconstitute the cleavage properties of Pol III-WT and therefore suggest that C53/37 influences C11 activity within Pol III. This is not unexpected, since parts of C53 and C37 have been independently localized at or near the catalytic center of Pol III (26, 27) and because multiple point mutations in this region of C37 cause terminator read-through (28). Since the \geq 3-nt-backtracked complexes are very poor substrates of Pol III Δ +C11, the data in Fig. 4B and C raise the possibility that the Pol III Δ terminator-arrested complexes are refractory to C11 because they are more extensively backtracked than 2 nt. Although the known weakness of the oligo(rU-dA) hybrid and its role in transcript release suggest that these arrested complexes are extensively backtracked rather than forward translocated, our attempts to address this using conventional methods have been unsuccessful.

C53/37 and C11 collectively reconstitute a holoenzyme mode of Pol III termination. The data suggested that Pol III terminates transcription by preventing arrest while slowing elongation of an oligo(rU-dA) hybrid. We next asked if addition of C53/37 to Pol III Δ is sufficient for both activities and if C11 also contributes. For this analysis, we first focus on the elongation rate component of the holoenzyme mechanism as reflected by lengths of released RNAs. Addition of C53/37 to Pol III Δ altered the length distribu-

tion of transcripts released from the template 1 terminator (Fig. 5A). The most abundantly released transcript was 9 nt for Pol III Δ and 8 nt for Pol III-WT (Fig. 5A, lanes 2 and 3, and B); this is also illustrated in Fig. 5B, in which the relative band intensities of each lane are quantitatively compared. However, addition of C53/37 alone did not reconstitute the relative band intensities seen for the transcripts released by Pol III-WT (compare lanes 1 and 3). Only after addition of C11 along with C53/37 was the pattern of relative band intensities further shifted toward the shorter end of the range, more similar to that of Pol III-WT (Fig. 5A; compare lanes 1, 2, and 4). This is better revealed in Fig. 5B, in which relative band intensities of each lane of released RNAs are compared. The cumulative data show that C11 can affect the length of terminated transcripts in the presence of C53/37 but not in its absence (Fig. 4B and C).

The effects of C53/37 and C11 on the lengths of released transcripts were also apparent on template 2 (Fig. 5C). While C53/37 shifted peak transcript length toward shortening, the shift was greater when C11 was also present (Fig. 5C, lanes 5 and 7). Again, the distribution pattern of terminator-released transcripts was most similar to that of Pol III-WT only when both C53/37 and C11 were added to Pol III Δ (Fig. 5C, lanes 1, 3, 5, and 7). This is also seen in Fig. 5D, which reflects only the lanes corresponding to released RNA.

Focusing next on the bound fractions shown in Fig. 5C, addition of C53/37 to Pol III Δ altered the pattern of arrested transcripts but did not rescue the release deficiency completely (compare lanes 2, 4, and 6). When C11 was added along with C53/37 (Fig. 5C, lane 8), the bound fraction showed better restoration to the Pol III-WT pattern than C53/37 alone (Fig. 5, even-numbered lanes). Figure 5E reflects on the lanes of Fig. 5C corresponding to the terminator-bound RNAs. The results suggest that C53/37 and C11 work together to fully reconstitute the holoenzyme mechanism of Pol III termination. We note that while the results are subtle for each single-protein addition, they are additive for C53/37 and C11 and are, with regard to effects on elongation slowing and antiarrest, cumulatively significant overall. The specificity and significance of these data showed that while C53/37 and C11 could prevent terminator arrest, other experiments revealed that they could not rescue previously arrested complexes (see Fig. S2 in the supplemental material).

Intact cleavage activity of C11 is required to prevent terminator arrest. The C-terminal domain of C11 is highly homologous to the zinc ribbon motif of TFIIS, including two acidic side chains at the tip of its hairpin loop that are responsible for intrinsic transcript cleavage (32, 52). We examined effects of C11 double-tip mutations, D91A · E92A, on Pol III Δ (Fig. 6). C11-D91A · E92A is deficient for transcript cleavage, as expected (Fig. 6A, lanes 6 to 8). Add-back experiments reproduced the result that C11 plus C53/37 could prevent arrest as well as shift the length of released transcripts to one more similar to that of Pol III-WT (Fig. 6B, lanes 1 to 8). Length distributions of released transcripts are shown in Fig. 6C. Notably, while C11-D91A · E92A plus C53/C37 corrected the lengths of released transcripts (Fig. 6B, lane 9), it was deficient in preventing arrest, as it produced a pattern more similar to that of Pol III Δ +C53/37 with no added C11 (Fig. 6B, compare lanes 6, 8, and 10). The distribution profiles for the bound RNAs are illustrated in Fig. 6D.

The N-terminal domain of C53 functions in termination. C53 and C37 associate with each other and with Pol III via their dimerization domains. We made a C53 truncation mutant that

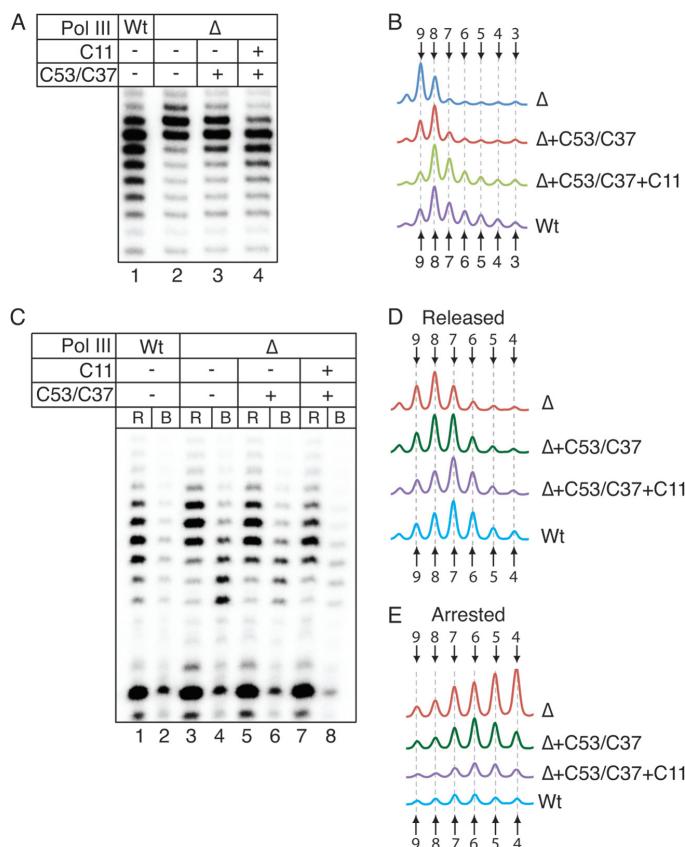


FIG 5 C53/37 and C11 constitute the holoenzyme mechanism of Pol III termination. (A) The released fractions only after transcription by Pol III-WT and Pol III Δ on template 1 in the presence of C53/37 or C53/37 and C11 as indicated above the lanes. (B) Quantitative lane tracing profiles of fractions shown in panel A as indicated by the reaction components at the right. (C) Transcription reactions done with template 2 in the presence of C53/37 and C53/37/C11 as indicated. (D) Quantitative lane tracing profiles of released fractions (lanes 1, 3, 5, and 7) shown in panel C. (E) Quantitative lane tracing profiles of bound fractions (lanes 2, 4, 6, and 8) shown in panel C. The number of 3' U residues each peak corresponds to is indicated above and below the profiles.

Downloaded from <http://mcb.asm.org/> on April 16, 2016 by Rockefeller University

contains the dimerization domain but lacks the 280 N-terminal amino acids. As expected, this forms a stable dimer with C37 in *E. coli* that could also be affinity purified, as was the full-length heterodimer. Addition of C53-Nt Δ /C37 along with WT C11 rescues the released transcript length defect (Fig. 6B, lane 13) but fails to prevent arrest (Fig. 6B, lane 14). The data suggest that the C53 N-terminal domain works with C11 to prevent terminator arrest and also helps explain why C11 alone was insufficient to prevent arrest even at a low NTP concentration (Fig. 4).

DISCUSSION

Subunit-dependent and core mechanisms of Pol III termination. There are two mechanisms by which *S. cerevisiae* Pol III can terminate transcription. One is C53/37 and C11 independent, and the other is dependent on C53/37 and C11; they are referred to as the core and holoenzyme mechanisms, respectively. The core mechanism requires a longer oligo(dT) tract than does the holoenzyme. *E. coli* RNA polymerase and Pol II can each use two

types of termination signals to initiate different termination mechanisms. This work has exposed two pathways by which Pol III can terminate at a single type of signal, oligo(rU-dA). Although the holoenzyme mechanism would operate at *S. cerevisiae* tRNA genes with ≤ 7 T nucleotides and even at those with longer T tracts, whether the core mechanism functions at all at longer T tracts *in vivo* remains unknown.

The core mechanism of transcription termination by Pol III likely is driven by the uniquely low thermal stability of the oligo(rU-dA) hybrid that forms within the Pol III active center (42). This is consistent with our data that show that oligo(dA) on the template strand induces termination by Pol III while oligo(dT) on the template or nontemplate strand does not (see Fig. S3 in the supplemental material). We propose that the Pol III core mechanism is highly sensitive to an RNA-DNA hybrid of 8 or more rU-dA base pairs, which would appear to autonomously destabilize the Pol III EC.

C53/37 and C11 contribute to the Pol III active site during termination. Addition of C53/37 alone significantly improved,

Downloaded from <http://mcb.asm.org/> on April 16, 2016 by Rockefeller University

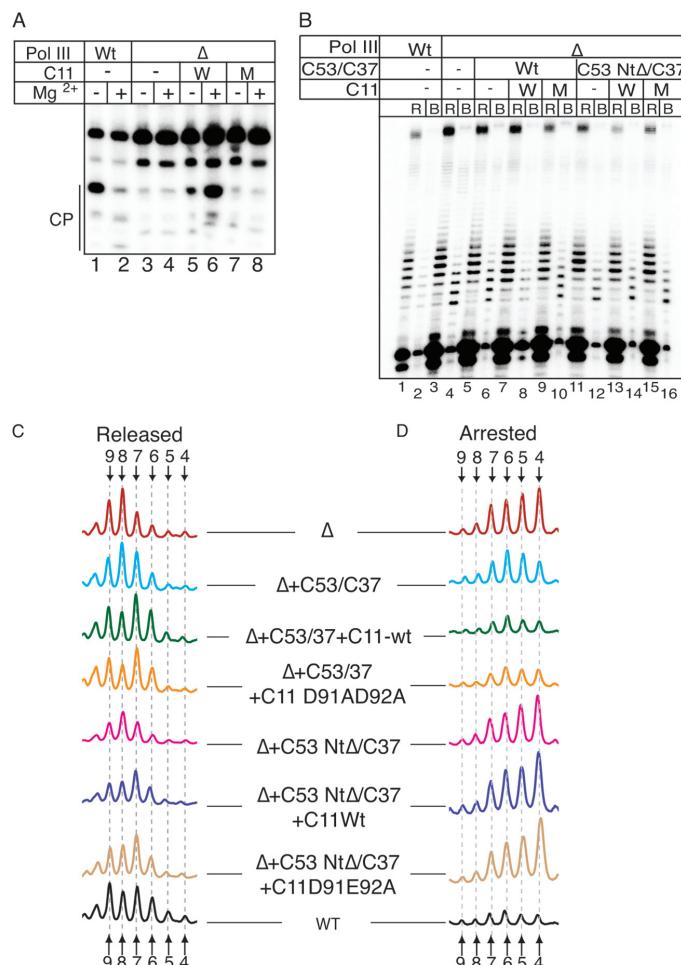


FIG 6 C11 cleavage activity and the C53 N-terminal region are required for antiarrest activity but not released transcript length reduction. (A) Cleavage assay for wild-type C11 and D91A · E92A mutant C11. Two-nucleotide-backtracked complexes were assembled as described for Fig. 4B, followed by incubation with WT (W; lanes 5 and 6) or mutant C11 (M; lanes 7 and 8). Reactions were initiated for Pol IIIA samples by addition of MgCl₂ (lanes 6 and 8). (B) Transcription reactions on template 2 ECs for Pol III-WT or Pol IIIΔ with or without addition of WT or mutant C11 and C53/37; W indicates wild-type C11, and M indicates C11-D91A · E92A. C53-NtΔ/C37 represents the truncation mutant of C53 that lacks the first 280 amino acids. Proteins were incubated with elongation complexes for 20 min prior to addition of NTPs (0.5 mM). (C) Quantitative lane tracing profiles of the lanes corresponding to released samples (odd-numbered lanes). (D) Quantitative lane tracing profiles of the lanes corresponding to bound (arrested) transcripts (even-numbered lanes).

but did not fully correct, the two termination deficiencies of Pol IIIΔ, while C11 alone could correct neither. While C11 could mediate intrinsic transcript cleavage by Pol IIIA in the absence of C53/37, the data suggested that it does so with different characteristics than Pol III-WT. Moreover, the ability of C11 to functionally contribute to both components of the holoenzyme mechanism of termination depends on C53/37. Our data show that the termination properties of Pol IIIΔ+C53/37 are significantly different from those of Pol III-WT and Pol IIIΔ+C53/37+C11. The data suggest that in the presence of C53/37, C11 further affects Pol III termination, including its contribution to antiarrest at the termini.

These and other recent data suggest a Pol III catalytic center that is more complex than envisioned based on the models provided by *E. coli* RNA polymerase, in which only the two largest subunits contribute to the active site. This view emerges from the recent findings that parts of both C53 and C37 lay close to, if not within, the active site of Pol III (26, 27). The C-terminal region of C37 that is localized to the active site of *S. cerevisiae* Pol III was found to be a hot spot for mutations in *S. pombe* C37 that cause terminator read-through (28). Also, an intricate relationship between transcript cleavage and termination activities established for RPC2 (31) was recently extended to C37 and C11, with each

Arimbasseri and Maraia

involved in both activities (28, 33). A model that arises is that parts of C53, C53, and C11 intricately and functionally contribute to the active site of Pol III along with the two largest subunits during termination.

The dimerization domains of C53 and C37 hold these two subunits together and also attach them to the lobe surface of Pol III (23, 25). The interdependency of C53 and C37 complicates our ability to examine the role of one without the other. We circumvented this by using a deletion construct of C53 comprised of its dimerization domain only, which was shown to work with C37 to slow elongation but not prevent arrest.

C11 contributes to two components of a Pol III holoenzyme termination mechanism. Elongation slowing, as manifested by shorter transcripts, confirms that C11 contributes to this component of termination. However, this activity was insensitive to C11 mutations that ablate transcript cleavage. This is reminiscent of the finding that while the C11 polypeptide is required for facilitated recycling, its cleavage activity is not (18). C11 also contributes to antiarrest at the terminator, although this activity was sensitive to the mutations that ablate transcript cleavage.

C53/37-dependent C11 activity appears to slow elongation through oligo(dA) more so than C53/37 alone, as manifested by the shorter RNAs. This could occur by one and/or both of two mechanisms. In addition to already-slowed elongation induced by C53/37, engagement of C11 might modify the catalytic center in a manner that further decreases forward movement, increasing dwell time and consequent release of shorter fragments. C11-mediated cleavage during termination might also decrease the length of the oligo(rU-dA) hybrid, leading to EC destabilization. Pol III cleaves terminal U residues cotranscriptionally (45, 53), suggesting that a termination complex is a substrate for C53/37-dependent cleavage by C11. Since the ability of C11 to optimally prevent terminator arrest requires its cleavage activity, we suppose that RNA 3' nibbling enhanced by C53/37 contributes to the antiarrest activity of C11. This is consistent with C11-mediated cleavage of 3' oligo(U) at termination (32). C37 point mutants that cause a significant fraction of Pol III to read through terminators nonetheless release terminated transcripts whose 3' oligo(U) tracts are shorter than those released in a C11 cleavage-deficient double mutant (28).

Holoenzyme mechanism of Pol III termination. Pol IIIΔ displayed two termination defects, transcript release deficiency for short 3' oligo(U)-containing RNAs and arrest at the terminator. This led to the conclusion that the holoenzyme mechanism consists of two components, slowing of elongation in the oligo(dA) tract and prevention of arrest in the oligo(dA) tract. Lowering NTP concentrations was used to slow elongation, and this led to better release of shorter transcripts but was insufficient to abate terminator arrest, indicating that the slow elongation and antiarrest components could be distinguished. Arrest could be neither rescued nor prevented by C11 under conditions of low NTP concentration. The two components of the holoenzyme mechanism could also be distinguished by mutating C53 and C11. The C11 protein, but not its cleavage activity, is required for optimal slowing, as reflected by the reduced length of released RNAs. In contrast, the N-terminal domain of C53 as well as the cleavage activity of C11 function in antiarrest activity.

A proposed essential function of C11 cleavage activity. Data reported here help explain why the cleavage activity of C11 is essential for yeast. Point mutations in the TFIIS-homologous loop

tip of C11 that ablate its transcript cleavage activity are lethal in *S. cerevisiae* and *S. pombe* (29, 32), yet the same mutations have no effect on Pol III-facilitated recycling (18). The only known functions that depend on cleavage-active C11 are proofreading and RNA 3' oligo(U) length determination (30, 32). Because the nucleotide selectivity of Pol III at 5.6×10^3 (30) predicts only 1 misincorporation per ~ 50 tRNAs, the inability to correct mistakes via C11 cleavage activity seems an unlikely cause of lethality. Likewise, inability to trim 3' oligo(U) tracts of nascent transcripts (32) also seems an unlikely cause of lethality. We propose that the inability to prevent Pol III from arresting, including at terminators, accounts for its essential function.

A need for preventing arrest during termination was not anticipated. Although pausing is prerequisite to termination, it increases the propensity to arrest. We suspect that preventing arrest is linked to RNA release and possibly resetting Pol III for reinitiation.

ACKNOWLEDGMENTS

We thank G. A. Kassavetis (UCSD) for discussion, consultation, and subunit plasmids; E. P. Geiduschek, N. Komissarova (NCI), and J. Zhang (NHLBI) for discussion; K. Potrykus and The Friday Seminar members for comments; and M. Blum for media.

This work was supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH. R.J.M. is a member of the U.S. Public Health Service Commissioned Corps.

REFERENCES

- Roeder RG, Rutter WJ. 1969. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* 224:234–237.
- Carter R, Drouin G. 2010. The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors. *Mol. Biol. Evol.* 27:1035–1043.
- Santangelo TJ, Artsimovitch I. 2011. Termination and antitermination: RNA polymerase runs a stop sign. *Nat. Rev. Microbiol.* 9:319–329.
- Padmanabhan K, Robles MS, Westerling T, Weitz CJ. 2012. Feedback regulation of transcriptional termination by the mammalian circadian clock PERIOD complex. *Science* 337:599–602.
- Kuehner JN, Pearson EL, Moore C. 2011. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat. Rev. Mol. Cell Biol.* 12:283–294.
- Richard P, Manley JL. 2009. Transcription termination by nuclear RNA polymerases. *Genes Dev.* 23:1247–1269.
- Mischio HE, Proudfoot NJ. 2012. Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochim. Biophys. Acta* 1829:174–185.
- Arimbasseri AG, Rijal K, Maraia RJ. 2012. Transcription termination by the eukaryotic RNA polymerase III. *Biochim. Biophys. Acta* doi:[10.1016/j.bbagen.2012.10.006](https://doi.org/10.1016/j.bbagen.2012.10.006). [Epub ahead of print.]
- Arndt KM, Kane CM. 2003. Running with RNA polymerase: eukaryotic transcript elongation. *Trends Genet.* 19:543–550.
- Peters JM, Vangeloff AD, Landick R. 2011. Bacterial transcription terminators: the RNA 3'-end chronicles. *J. Mol. Biol.* 412:793–813.
- Epshtain V, Dutta D, Wade J, Nudler E. 2010. An allosteric mechanism of Rho-dependent transcription termination. *Nature* 463:245–249.
- Kawauchi J, Mischio H, Braglia P, Rondon A, Proudfoot NJ. 2008. Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes Dev.* 22:1082–1092.
- Cozzarelli NR, Gerrard SP, Schlessel M, Brown DD, Bogenhagen DF. 1983. Purified RNA polymerase III accurately and efficiently terminates transcription of 5S RNA genes. *Cell* 34:829–835.
- Bogenhagen DF, Brown DD. 1981. Nucleotide sequences in Xenopus 5S DNA required for transcription termination. *Cell* 24:261–270.
- Cabart P, Lee J, Willis IM. 2008. Facilitated recycling protects human RNA polymerase III from repression by Maf1 in vitro. *J. Biol. Chem.* 283:36108–36117.

16. Stults DM, Killen MW, Pierce HH, Pierce AJ. 2008. Genomic architecture and inheritance of human ribosomal RNA gene clusters. *Genome Res.* 18:13–18.
17. Moir RD, Willis IM. 2012. Regulation of Pol III transcription by nutrient and stress signaling pathways. *Biochim. Biophys. Acta* doi:[10.1016/j.bbagen.2012.11.001](https://doi.org/10.1016/j.bbagen.2012.11.001). [Epub ahead of print.]
18. Landrieux E, Aliz N, Ducrot C, Acker J, Riva M, Carles C. 2006. A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. *EMBO J.* 25:118–128.
19. White RJ. 2005. RNA polymerases I and III, growth control and cancer. *Nat. Rev. Mol. Cell Biol.* 6:69–78.
20. Geiduschek EP, Kassavetis GA. 2006. Transcription: adjusting to adversity by regulating RNA polymerase. *Curr. Biol.* 16:R849–R851.
21. Willis IM, Moir RD. 2007. Integration of nutritional and stress signaling pathways by Maf1. *Trends Biochem. Sci.* 32:51–53.
22. Marshall L, Goodfellow SJ, White RJ. 2007. Diminished activity of RNA polymerase III selectively disrupts tissues with the most actively dividing cells. *PLoS Biol.* 5:e286. doi:[10.1371/journal.pbio.0050286](https://doi.org/10.1371/journal.pbio.0050286).
23. Vannini A, Ringel R, Kusser AG, Berninghausen O, Kassavetis GA, Cramer P. 2010. Molecular basis of RNA polymerase III transcription repression by Maf1. *Cell* 143:59–70.
24. Fernandez-Tornero C, Bottcher B, Riva M, Carles C, Steuerwald U, Ruigrok RW, Sentenac A, Muller CW, Schoehn G. 2007. Insights into transcription initiation and termination from the electron microscopy structure of yeast RNA polymerase III. *Mol. Cell* 25:813–823.
25. Fernandez-Tornero C, Bottcher B, Rashid UJ, Steuerwald U, Florchinger B, Devos DP, Lindner D, Muller CW. 2010. Conformational flexibility of RNA polymerase III during transcriptional elongation. *EMBO J.* 29:3762–3772.
26. Wu CC, Lin YC, Chen HT. 2011. The TFIIF-like Rpc37/53 dimer lies at the center of a protein network to connect TFIIC, Bdp1, and the RNA polymerase III active center. *Mol. Cell.* 31:2715–2728.
27. Kassavetis GA, Prakash P, Shim E. 2010. The C53/C37 subcomplex of RNA polymerase III lies near the active site and participates in promoter opening. *J. Biol. Chem.* 285:2695–2706.
28. Rijal K, Maraia RJ. 2013. RNA polymerase III mutants in TFIIFα-like C37 cause terminator readthrough with no decrease in transcription output. *Nucleic Acids Res.* 41:139–155.
29. Chedin S, Riva M, Schultz P, Sentenac A, Carles C. 1998. The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. *Genes Dev.* 12:3857–3871.
30. Alic N, Ayoub N, Landrieux E, Favry E, Baudouin-Cornu P, Riva M, Carles C. 2007. Selectivity and proofreading both contribute significantly to the fidelity of RNA polymerase III transcription. *Proc. Natl. Acad. Sci. U. S. A.* 104:10400–10405.
31. Bobkova EV, Habib N, Alexander G, Hall BD. 1999. Mutational analysis of the hydrolytic activity of yeast RNA polymerase III. *J. Biol. Chem.* 274:21342–21348.
32. Huang Y, Intine RV, Mozlin A, Hasson S, Maraia RJ. 2005. Mutations in the RNA polymerase III subunit Rpc11p that decrease RNA 3' cleavage activity increase 3'-terminal oligo(U) length and La-dependent tRNA processing. *Mol. Cell. Biol.* 25:621–636.
33. Iben JR, Mazeika JK, Hasson S, Rijal K, Arimbasseri AG, Russo AN, Maraia RJ. 2011. Point mutations in the Rpb9-homologous domain of Rpc11 that impair transcription termination by RNA polymerase III. *Nucleic Acids Res.* 39:6100–6113.
34. Campbell FE, Jr, Setzer DR. 1992. Transcription termination by RNA polymerase III: uncoupling of polymerase release from termination signal recognition. *Mol. Cell. Biol.* 12:2260–2272.
35. Lorenzen K, Vannini A, Cramer P, Heck AJ. 2007. Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. *Structure* 15:1237–1245.
36. Lane LA, Fernandez-Tornero C, Zhou M, Morgner N, Ptchelkina D, Steuerwald U, Politis A, Lindner D, Gvozdenovic J, Gavin AC, Muller CW, Robinson CV. 2011. Mass spectrometry reveals stable modules in holo and apo RNA polymerases I and III. *Structure* 19:90–100.
37. Komissarova N, Kireeva ML, Becker J, Sidorenkov I, Kashlev M. 2003. Engineering of elongation complexes of bacterial and yeast RNA polymerases. *Methods Enzymol.* 370:233–251.
38. Matsuzaki H, Kassavetis GA, Geiduschek EP. 1994. Analysis of RNA chain elongation and termination by *Saccharomyces cerevisiae* RNA polymerase III. *J. Mol. Biol.* 235:1173–1192.
39. Bintu L, Kopaczynska M, Hodges C, Lubkowska L, Kashlev M, Bustamante C. 2011. The elongation rate of RNA polymerase determines the fate of transcribed nucleosomes. *Nat. Struct. Mol. Biol.* 18:1394–1399.
40. Bar-Nahum G, Epshteyn V, Ruckenstein AE, Rafikov R, Mustaev A, Nudler E. 2005. A ratchet mechanism of transcription elongation and its control. *Cell* 120:183–193.
41. Komissarova N, Kashlev M. 1997. RNA polymerase switches between inactivated and activated states by translocating back and forth along the DNA and the RNA. *J. Biol. Chem.* 272:15329–15338.
42. Martin FH, Tinoco I, Jr. 1980. DNA-RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. *Nucleic Acids Res.* 8:2295–2299.
43. Nudler E, Mustaev A, Lukhtanov E, Goldfarb A. 1997. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* 89:33–41.
44. Nudler E. 2012. RNA polymerase backtracking in gene regulation and genome instability. *Cell* 149:1438–1445.
45. Bobkova EV, Hall BD. 1997. Substrate specificity of the RNase activity of yeast RNA polymerase III. *J. Biol. Chem.* 272:22832–22839.
46. Santangelo TJ, Roberts JW. 2004. Forward translocation is the natural pathway of RNA release at an intrinsic terminator. *Mol. Cell* 14:117–126.
47. Chen ZA, Jawhari A, Fischer L, Buchen C, Tahir S, Kamenski T, Rasmussen M, Lariviere L, Bukowski-Wills JC, Nilges M, Cramer P, Rappaport J. 2010. Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.* 29:717–726.
48. Cheung AC, Cramer P. 2011. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* 471:249–253.
49. Sydow JF, Brueckner F, Cheung AC, Damsma GE, Dengl S, Lehmann E, Vassylyev D, Cramer P. 2009. Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol. Cell* 34:710–721.
50. Wang D, Bushnell DA, Huang X, Westover KD, Levitt M, Kornberg RD. 2009. Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science* 324:1203–1206.
51. Borukhov S, Sagitov V, Goldfarb A. 1993. Transcript cleavage factors from *E. coli*. *Cell* 72:459–466.
52. Kettenberger H, Armache KJ, Cramer P. 2003. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* 114:347–357.
53. Whitehall SK, Bardeleben C, Kassavetis GA. 1994. Hydrolytic cleavage of nascent RNA in RNA polymerase III ternary transcription complexes. *J. Biol. Chem.* 269:2299–2306.

Downloaded from <http://mcb.asm.org/> on April 16, 2016 by Rockefeller University

References

1. Crick, F. H. C. *From DNA to protein On degenerate templates and the adapter hypothesis: a note for the RNA Tie Club* 1955.
2. Woese, C. *The Genetic Code. The Molecular basis for Genetic Expression* 1st ed. (Harper, 1967).
3. Soll, D. & RajBhandary, U. *tRNA: Structure, Biosynthesis and Function* 1st ed. (ASM press, 1995).
4. Cooper, T. A., Wan, L. & Dreyfuss, G. RNA and Disease. *Cell* **136**, 777–793 (Feb. 2009).
5. Park, S. G., Schimmel, P. & Kim, S. Aminoacyl tRNA synthetases and their connections to disease. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 11043–11049 (Aug. 2008).
6. Griffiths, E. J. in, 249–267 (Springer Netherlands, Dordrecht, Dec. 2011).
7. McFarland, R., Elson, J. L., Taylor, R. W., Howell, N. & Turnbull, D. M. Assigning pathogenicity to mitochondrial tRNA mutations: when ‘definitely maybe’ is not good enough. *Trends in Genetics* **20**, 591–596 (Dec. 2004).
8. Dana, A. & Tuller, T. Determinants of Translation Elongation Speed and Ribosomal Profiling Biases in Mouse Embryonic Stem Cells. *PLoS Computational Biology* **8**, e1002755–11 (Nov. 2012).

9. Dana, A. & Tuller, T. Mean of the typical decoding rates: a new translation efficiency index based on the analysis of ribosome profiling data. *G3 (Bethesda, Md.)* **5**, 73–80 (Dec. 2014).
10. Mahlab, S., Tuller, T. & Linial, M. Conservation of the relative tRNA composition in healthy and cancerous tissues. *RNA* **18**, 640–652 (Mar. 2012).
11. Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics* **12**, 32–42 (Jan. 2011).
12. Tuller, T. *et al.* An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell* **141**, 344–354 (Apr. 2010).
13. Weinberg, D. E. *et al.* Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. *Cell Reports* **14**, 1787–1799 (Feb. 2016).
14. Hasler, D. *et al.* The Lupus Autoantigen La Prevents Mis-channeling of tRNA Fragments into the Human MicroRNA Pathway. *Molecular Cell* **63**, 110–124 (July 2016).
15. Ivanov, P., Emara, M. M., Villen, J., Gygi, S. P. & Anderson, P. Angiogenin-Induced tRNA Fragments Inhibit Translation Initiation. *Molecular Cell* **43**, 613–623 (Aug. 2011).
16. Lee, Y. S., Shibata, Y., Malhotra, A. & Dutta, A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes & Development* **23**, 2639–2649 (Nov. 2009).
17. Haussecker, D. *et al.* Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* **16**, 673–695 (Mar. 2010).

18. Babiarz, J. E., Ruby, J. G., Wang, Y., Bartel, D. P. & Blelloch, R. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes & Development* **22**, 2773–2785 (Oct. 2008).
19. Iben, J. R. & Maraia, R. J. tRNA gene copy number variation in humans. *Gene* **536**, 376–384 (Feb. 2014).
20. Pechmann, S. & Frydman, J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nature Publishing Group* **20**, 237–243 (Dec. 2012).
21. Gingold, H. *et al.* A Dual Programfor Translation Regulationin Cellular Proliferation and Differentiation. *Cell* **158**, 1281–1292 (Sept. 2014).
22. Kutter, C. *et al.* Pol III binding in six mammals shows conservation among amino acid isotypes despite divergence among tRNA genes. *Nature Genetics* **43**, 948–955 (Aug. 2011).
23. Moqtaderi, Z. *et al.* Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. *Nature Structural & Molecular Biology* **17**, 635–640 (Apr. 2010).
24. Oler, A. J. *et al.* nsmb.1801. *Nature Structural & Molecular Biology* **17**, 620–628 (Apr. 2010).
25. Dittmar, K. A., Mobley, E. M., Radek, A. J. & Pan, T. Exploring the Regulation of tRNA Distribution on the Genomic Scale. *Journal of Molecular Biology* **337**, 31–47 (Mar. 2004).
26. Goodarzi, H. *et al.* Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. *Cell* **165**, 1416–1427 (June 2016).

27. 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 (Sept. 2015).
28. Zheng, G. *et al.* Efficient and quantitative high-throughput tRNA sequencing. *Nature Methods*, 1–5 (July 2015).
29. Trotta, C. R. & Abelson, J. *The RNA World* Second edition (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 561–584 (Cold Spring Harbor Laboratory Press, 1999).
30. Paushkin, S. V., Patel, M., Furia, B. S., Peltz, S. W. & Trotta, C. R. Identification of a human endonuclease complex reveals a link between tRNA splicing and pre-mRNA 3' end formation. *Cell* **117**, 311–321 (Apr. 2004).
31. Weitzer, S. & Martinez, J. The human RNA kinase hClp1 is active on 3' transfer RNA exons and short interfering RNAs. *Nature* **447**, 222–226 (May 2007).
32. Popow, J. *et al.* HSPC117 is the essential subunit of a human tRNA splicing ligase complex. *Science* **331**, 760–764 (Feb. 2011).
33. Popow, J., Jurkin, J., Schleiffer, A. & Martinez, J. Analysis of orthologous groups reveals archease and DDX1 as tRNA splicing factors. *Nature* **511**, 104–107 (June 2014).
34. Phizicky, E. M. & Hopper, A. K. tRNA biology charges to the front. *Genes & Development* **24**, 1832–1860 (Sept. 2010).
35. Hopper, A. K., Pai, D. A. & Engelke, D. R. Cellular dynamics of tRNAs and their genes. *FEBS Letters* **584**, 310–317 (Jan. 2010).

36. Hopper, A. K. Transfer RNA post-transcriptional processing, turnover, and subcellular dynamics in the yeast *Saccharomyces cerevisiae*. *Genetics* **194**, 43–67 (May 2013).
37. Dittmar, K. A., Goodenbour, J. M. & Pan, T. Tissue-Specific Differences in Human Transfer RNA Expression. *PLoS Genetics* **2**, e221 (2006).
38. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25**, 955–964 (Mar. 1997).
39. Chan, P. P. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Research* **37**, D93–D97 (Jan. 2009).
40. Gu, W. tRNAHis maturation: An essential yeast protein catalyzes addition of a guanine nucleotide to the 5' end of tRNAHis. *Genes & Development* **17**, 2889–2901 (Dec. 2003).
41. Jackman, J. E. & Alfonzo, J. D. Transfer RNA modifications: nature's combinatorial chemistry playground. *Wiley Interdisciplinary Reviews: RNA* **4**, 35–48 (Nov. 2012).
42. Lee, J.-H., Ang, J. K. & Xiao, X. Analysis and design of RNA sequencing experiments for identifying RNA editing and other single-nucleotide variants. *RNA* **19**, 725–732 (June 2013).
43. Gustilo, E. M., Vendeix, F. A. & Agris, P. F. tRNA's modifications bring order to gene expression. *Current Opinion in Microbiology* **11**, 134–140 (Apr. 2008).
44. Hafner, M. *et al.* Barcoded cDNA library preparation for small RNA profiling by next-generation sequencing. *Methods* **58**, 164–170 (Oct. 2012).

45. Maraia, R. J. & Lamichhane, T. N. 3' processing of eukaryotic precursor tRNAs. *Wiley Interdisciplinary Reviews: RNA* **2**, 362–375 (Nov. 2010).
46. Stefano, J. E. Purified lupus antigen La recognizes an oligouridylate stretch common to the 3' termini of RNA polymerase III transcripts. *Cell* **36**, 145–154 (Jan. 1984).
47. Bayfield, M. A. & Maraia, R. J. Precursor-product discrimination by La protein during tRNA metabolism. *Nature Structural & Molecular Biology* **16**, 430–437 (Mar. 2009).
48. Arimbasseri, A. G., Kassavetis, G. A. & Maraia, R. J. Transcription. Comment on "Mechanism of eukaryotic RNA polymerase III transcription termination". *Science* **345**, 524–524 (Aug. 2014).
49. Nielsen, S., Yuzenkova, Y. & Zenkin, N. Mechanism of eukaryotic RNA polymerase III transcription termination. *Science* **340**, 1577–1580 (June 2013).
50. Teplova, M. *et al.* Structural Basis for Recognition and Sequestration of UU-UOH 3' Temini of Nascent RNA Polymerase III Transcripts by La, a Rheumatic Disease Autoantigen. *Molecular Cell* **21**, 75–85 (Jan. 2006).
51. Machnicka, M. A. *et al.* MODOMICS: a database of RNA modification pathways—2013 update. *Nucleic Acids Research* **41**, D262–7 (Jan. 2013).
52. Torres, A. G. *et al.* Inosine modifications in human tRNAs are incorporated at the precursor tRNA level. *Nucleic Acids Research*, 1–13 (Apr. 2015).
53. Namavar, Y., Barth, P. G., Poll-The, B. T. & Baas, F. Classification, diagnosis and potential mechanisms in Pontocerebellar Hypoplasia. *Orphanet Journal of Rare Diseases* **6**, 50 (July 2011).

54. Baltz, A. G. *et al.* The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Molecular Cell* **46**, 674–690 (June 2012).
55. Maraia, R. J. & Bayfield, M. A. The La Protein-RNA Complex Surfaces. *Molecular Cell* **21**, 149–152 (Jan. 2006).
56. Arimbasseri, A. G. & Maraia, R. J. Mechanism of Transcription Termination by RNA Polymerase III Utilizes a Non-template Strand Sequence-Specific Signal Element. *Molecular Cell* **58**, 1124–1132 (June 2015).
57. Arimbasseri, A. G. & Maraia, R. J. Distinguishing Core and Holoenzyme Mechanisms of Transcription Termination by RNA Polymerase III. *Molecular and Cellular Biology* **33**, 1571–1581 (Mar. 2013).
58. Iben, J. R. & Maraia, R. J. tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote. *RNA* **18**, 1358–1372 (July 2012).
59. Spitzer, J., Landthaler, M. & Tuschl, T. *Rapid Creation of Stable Mammalian Cell Lines for Regulated Expression of Proteins Using the Gateway® Recombination Cloning Technology and Flp-In T-REx® Lines* 1st ed. (Elsevier Inc., 2013).
60. Garzia, A., Meyer, C., Morozov, P., Sajek, M. & Tuschl, T. Optimization of PAR-CLIP for transcriptome-wide identification of binding sites of RNA-binding proteins. *Methods*, 1–17 (Oct. 2016).