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

A Thesis Presented to the Faculty of

The Rockefeller Uniersity

in Partial Fullfillment of the Requirements for

the degree of Doctor of Philosophy

by

Tasos Gogakos

June 2017

Abstract

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

Tasos Gogakos, Ph.D.

The Rockefeller University 2017

The participation of transfer RNAs (tRNAs) in test2 (tEsT) 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 sequencing read mapping efforts. To address these issues, we developed hydro-tRNAseq, a method based on partial alkaline RNA hydrolysis that generates fragments amenable for sequencing. To identify transcribed tRNA genes, we further complemented this approach with Photoactivatable Crosslinking and Immunoprecipitation (PAR-CLIP) of SSB/La, a conserved protein involved in pre-tRNA processing. Our results show that approximately half of all predicted tRNA genes are transcribed in human cells. We also report predominant nucleotide modification sites, their order of introduction, and identify tRNA leader, trailer and intron sequences. By using complementary sequencing-based methodologies we present a human tRNA atlas, and determine expression levels of mature and processing intermediates of tRNAs in human cells.

Στους γονείς και τον αδερφό μου

Acknowledgments

First, I would like to thank my

contents

Table of Contents

List of Figures

List of Tables

Glossary

List of Abbreviations

ncRNA noncoding RNA.

tEsT test2.

tRNA transfer RNA.

Chapter 1

Introduction

1.1 tRNA biology

tRNAs have been among the earliest studied noncoding RNA (ncRNA) non-coding RNA molecules [1, 2], established as the adaptor molecules responsible for translating the genetic code during protein synthesis cite Crick tie club letter. Yet, in recent years tRNAs received new attention in the context of codon-resolved translational control [3–8], 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) [9–11]. 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 [7, 12, 13]. 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 [**Gingold:2014iz**]. Similarly, experimental tRNA gene annotation in the past had to focus on RNA polymerase III (POLR3) ChIP-seq [**Kutter:2011ff**] [**Moqtaderi:2010hc**] [**Oler:2010fb**] or hybridization-based ap-

proaches [**Dittmar:2004fb**] [**Goodarzi:2016gd**]. 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 [**Cozen:2015ds**] [**Zheng:2015dw**]. However, an extensive annotation of human genes and transcripts was foregone because the focus was either on mature tRNAs only [**Zheng:2015dw**] or on tRNA fragments not inclusive of full-length precursor tRNA (pre-tRNA) transcripts [**Cozen:2015ds**]. Thus, to-date an experimentally validated list of curated mature and pre-tRNA sequences and annotating tRNA genes in human is still missing.

We have combined complementary high-throughput techniques for obtaining the sequence composition and abundance of tRNAs in human embryonic kidney cells (HEK293). We developed hydro-tRNAseq, a modified small RNA sequencing protocol based on partial alkaline hydrolysis of input RNA, in order to identify and quantify tRNAs, and provided evidence for the validity of this approach when determining the accumulation of disease-associated tRNA intron fragments caused by mutations in the tRNA splicing machinery [**Karaca:2014em**]. Here we extend this approach by applying it to tRNA-enriched size fragments with the aim to annotate and curate all tRNAs. Since tRNA processing, such as precursor trimming and intron removal, is a fast process[**Foretek:2016ea**], we also aimed to enrich specifically for pre-tRNAs in order to identify and annotate the corresponding unique tRNA gene template. Thus, we performed PAR-CLIP on SSB, a con-

served and ubiquitous protein involved in 3' tRNA processing [**Bayfield:2009cx**] [**Bayfield:2010cs**] [**Stefano:1984wp**].

1.1.1 tRNA biogenesis

tRNA genes are transcribed by RNA polymerase III (POLR3) that uses promoters internal to the DNA sequence of the tRNA gene (tDNA). In humans, a minority of tRNA transcripts (see section XXX) harbor introns. A dedicated tRNA splicing complex composed fo core and accessory proteins carries out tRNA splicing cite references. Precursor tRNA (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, and ELAC2, the human homolog of tRNase Z trims the 3' trailer. Next, the universally conserved 3' terminal CCA tail is added by TRNT1, the tRNA nucleotidyll transferase 1 (TRNT1), and acts as the acceptor of the amino acid. tRNAs are further modified by chemical nucleotide modifications (see section XXX), exported from the nucleus to the cytoplasm where they can undergo further modifcations, 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. 1.1.

tRNA genes are transcribed by RNA polymerase III (POLR3) that uses promoters internal to the DNA sequence of the tRNA gene (tDNA). In humans, a minority of tRNA transcripts (see section XXX) harbor introns. A dedicated tRNA splicing complex composed fo core and accessory proteins carries out tRNA splicing cite references. Precursor tRNA (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, and ELAC2, the human homolog of tRNase Z trims the 3' trailer. Next, the universally conserved 3' terminal CCA tail is added by TRNT1, the tRNA nucleotidyll transferase 1 (TRNT1), and acts as the acceptor of the amino acid. tRNAs are further modified by chemical nucleotide modifications (see section XXX), exported from the nucleus to the cytoplasm where they can undergo further modifcations, are aminoacylated with their cognate amino

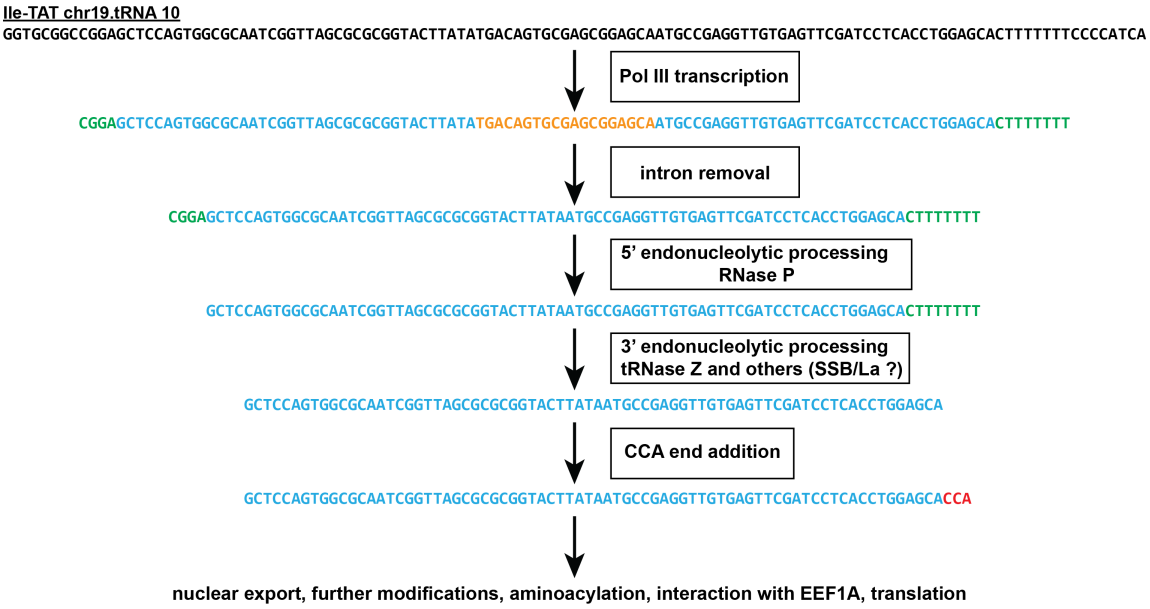

Figure 1.1: Overview of tRNA biogenesis and processing. thththt

Chapter 2

Results

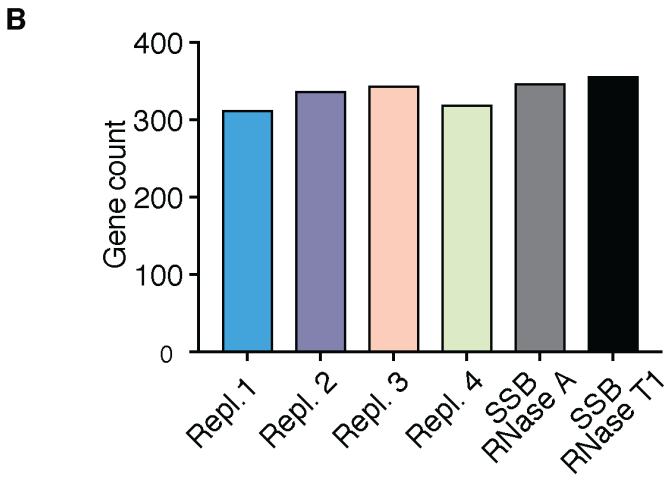

Figure 2.1: Venn Bars

[2, 14]

Chapter 3

woohooo

this is a new chapter

this is pager 2

References

1. Woese, C. *The Genetic Code. The Molecular basis for Genetic Expression* 1st ed. (Harper, 1967).

2. Soll, D. & RajBhandary, U. *tRNA: Structure, Biosynthesis and Function* 1st ed. (ASM press, 1995).

3. 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).

4. 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).

5. Mahlab, S., Tuller, T. & Linial, M. Conservation of the relative tRNA composition in healthy and cancerous tissues. *RNA* **18,** 640–652 (Mar. 2012).

6. 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).

7. Tuller, T. *et al.* An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell* **141,** 344–354 (Apr. 2010).

8. Weinberg, D. E. *et al.* Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. *CellReports* **14,** 1787–1799 (Feb. 2016).

9. 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).

10. 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).

11. 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).

12. Iben, J. R. & Maraia, R. J. tRNA gene copy number variation in humans. *Gene* **536,** 376–384 (Feb. 2014).

13. Pechmann, S. & Frydman, J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nature Publishing Group* **20,** 237–243 (Dec. 2012).

14. Arimbasseri, A. G. & Maraia, R. J. RNA Polymerase III Advances: Structural and tRNA Functional Views. *Trends in Biochemical Sciences,* 1–14 (Apr. 2016).