Title

Characterization of tRNAs, Associated Fragments, and Genomic Loci in Primate Neural Development and Beyond

Author

Bagi, Alex Laszlo

Publication Date

2024

UNIVERSITY OF CALIFORNIA

SANTA CRUZ

CHARACTERIZATION OF TRNAS, ASSOCIATED FRAGMENTS, AND GENOMIC LOCI IN PRIMATE NEURAL DEVELOPMENT AND BEYOND

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOINFORMATICS AND BIO-ENGINEERING

by

Alex Laszlo Bagi

June 2024

The Dissertation of Alex Laszlo Bagi
is approved:

Professor Todd Lowe, Chair

Professor Sofie Salama, Chair

Professor Richard Edward Green

Professor Russ Corbett-Detig

Professor Rohinton Kamakaka

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Characterization of tRNAs, Associated Fragments, and Genomic Loci in Primate Neural Development and Beyond

by

Alex Laszlo Bagi

This dissertation investigates the multifaceted roles of transfer RNAs (tRNAs) in gene expression regulation, extending beyond their traditional roles in protein translation. Chapter I explores tRNAs in human brain development, utilizing cerebral cortical organoids and specialized tRNA sequencing to reveal dynamic expression patterns of tRNAs and tRNA-derived small RNAs (tDRs) in the early human cerebral cortex. Notably, it identifies a variety of upregulated tDRs from diverse isodecoders, with sequence-specific conservation among neural-specific groups, suggesting a pivotal role in neural development. Chapter II examines the impact of complete telomere-to-telomere (T2T) assemblies of great ape (and human) genomes on tRNA gene discovery. This has led to the identification of nearly 100 new human cytosolic-tRNA gene loci, particularly in regions of chromosome 1 associated with neural gene regulation and implicated in the neural expansion of humans and primates. The chapter also discusses the dynamic nature of tRNA loci copy number variation across tandem repeats and their roles on adjacent neural genes. Chapter III introduces two novel bioinformatics tools, tRNAgraph, and tRNAmap, designed to address the challenges of analyzing tRNA sequencing data across multiple species and experimental conditions. tRNAgraph offers multi-variate analysis, automated clustering, and classification of tRNAseq data, while tRNAmap aligns the 'tRNAome' of eukaryotic species for cross-species comparison. These tools enable advanced analysis of the increasingly common tRNA sequencing data, contributing significantly to our understanding of tRNAs in neural development and evolution. This dissertation's exploration into the complex roles of tRNAs and the innovative tools developed for their analysis not only advances our

understanding of neural development and evolution but also paves the way for future genomic research to uncover the intricate mechanisms of gene regulation.

I extend deep gratitude to my advisors, Dr. Todd Lowe and Dr. Sofie Salama, for their unwavering support and invaluable time throughout my academic journey. Their mentorship provided me with a remarkable opportunity to delve into the intricacies of tRNAs and organoid cell culture while also honing my computational and data science skills. I am immensely grateful for the chance to explore a diverse range of scientific questions, which has significantly enriched my understanding of the human-primate lineage and tRNA biology.

I am equally thankful to my doctoral committee for their insightful input and guidance during my PhD studies and to all the collaborators who have contributed over the years. A special acknowledgment goes to my lab managers Dr. Jonathan Howard and Kristof Tigyi, for their expert advice on experimental methods, life, and career throughout my tenure at UCSC.

My passion for research was ignited at a tender age by my grandmothers, Ann Garland and Beverly Bigtree. Their encouragement was pivotal and inspirational, motivating me to pursue a career in science and strive to make a positive impact in the world.

Lastly, I owe a debt of gratitude to my loving and supportive husband, Nolan, and my corgi, Jace. To my family—my mom, dad, brother, and sister—who have been my constant source of conversation and support through both good times and bad. To my close friends, who have kept me motivated and cheered me on to the finish line—I am eternally grateful. The journey to this point has been unforgettable, and it is the people in my life who have made it so.

Part I

Chapters

Chapter 1

Specialized Neural tRNA Expression in Brain Organoids

1.1 Background

Transfer RNAs are the largest, most complex small non-coding RNA family, and are universal to all living systems. Historically, the majority of tRNA biochemistry, processing, and functional studies in eukaryotes have utilized the yeast *Saccharomyces cerevisiae* (Phizicky and Hopper 2010). Yet, the tRNA gene diversity is relatively limited in this single-celled organism, containing only 55 unique mature tRNA transcripts encoded redundantly by 275 genes (Chan and Todd M. Lowe 2016). By contrast, most mammalian genomes have at least 400-600 tRNA genes (tDNAs), with a surprisingly large number of unique sequences due to an expansion of isodecoder diversity (tRNAs that have the same anticodon but different sequences); in human, there are 252 unique tRNA gene sequences (five times as many as yeast), even after discounting likely pseudogenes (Chan and Todd M. Lowe 2016). It is unclear why the complexity, diversity, and number of unique tRNAs increase in higher eukaryotes relative to single-celled eukaryotes. However, increased tissue diversity, regulatory programs, and unique developmental needs are likely driving factors.

A growing body of evidence suggests that tRNAs play regulatory roles that go beyond the scope of codon translation. Some of these extra-translational roles include global regulation of RNA silencing, immune system dysregulation leading to disease, and cancer progression regulation (Haussecker et al. 2010; Nie et al. 2019; Su et al. 2019; Goodarzi, Nguyen, et al. 2016). In addition to mature tRNAs, tRNA-derived small RNAs (tDRNAs or tDRs) have also been demonstrated to have roles in regulation and disease (Goodarzi, Liu, et al. 2015; Dou, Y. Wang, and Lu 2019; Soares and Santos 2017; Avcilar-Kucukgoze and Kashina 2020; Morisaki et al. 2021). Some specific angiogenin-induced tDRs also have the ability to form into G-quadruplexes, which have neuroprotective roles in motor neurons, affecting the pathogenesis of diseases such as amyotrophic lateral sclerosis (Yang 2014; Ivanov et al. 2014). Further, the bioavailability of specific amino acids such as Arginine, Glycine, and Selenocysteine have been known to play roles in neurological function, disorders, and development, which can be related to the expression of specific tRNAs and their associated tDRs (Pitts et al. 2014; Ullah et al. 2020; Mader and Czorlich 2021).

Spatial- and time-dependent expression of protein-coding genes is an essential function for cell patterning and differentiation of neural tissues in mammalian neurodevelopment. Recent studies have found several instances where tDRs may play roles in neurological diseases (reviewed in Qin et al. 2020). A landmark study found that a specific but previously unremarkable Arg-UCU tRNA (encoded by one of five different Arg-TCT genes) is abundant almost exclusively in the brain (Ishimura et al. 2014). This study in mice showed that severe neurodegeneration could be caused by a single point mutation in the brain-specific tRNA gene when combined with a mutation in a ribosome release factor, showing that individual tRNA genes can have tissue-specific importance. In addition, it has been shown that loss of the mature tRNA Arg-TCT-4-1 (*n-Tr20* in mouse), leads to changes in seizure susceptibility and mTORC1 suppression leading to altered neurotransmission (Kapur et al. 2020). Except for this one specific tRNA in mouse, no known cytosolic tRNAs have been shown to play a specialized role(s) in neuro-development or mature brain function.

We have employed AlkB-facilitated RNA methylation sequencing (ARM-seq) (Cozen et al. 2015) to enable comprehensive analysis of tRNAs and their derived small RNAs. This specilized tRNA sequencing method use *Escherichia coli* dealkylating enzyme AlkB treatments to remove common tRNA modifications (primarily m^1A, m^1G, m^3C) that can cause reverse transcription to terminate before reaching the end of a tRNA or tDR. This and other similar methods, such as YAMAT-seq (Shigematsu et al. 2017), hydro-seq (Cheng et al. 2019), and QuantM-tRNA-seq (Pinkard et al. 2020), yield a much more complete picture of tDRs than standard small RNA sequencing methods. Furthermore with specific tRNA analysis pipelines, tDRs can be sorted into more complex categories than the common 5' and 3' halves and fragments distinction.

We applied ARM-seq to cerebral cortex brain organoids, an established model for early events in fetal brain development (Eiraku et al. 2008; Camp et al. 2015; Kelava and Lancaster 2016; Kadoshima et al. 2013), to profile tRNAs and tDRs at a series of time points from pluripotent cells through the generation of excitatory projection neurons of the developing cortical plate (Pollen et al. 2019; Fiddes, Pollen, et al. 2019; Fiddes, Lodewijk, et al. 2018). Studying tRNA dynamics in this unique system has allowed us to identify previously unknown brain-correlated tRNAs and tDRs expressed during human cortical neurogenesis. We found that, in addition to an increase in the previously studied Arg-TCT-4 tRNA, a broader collection of tRNAs and tDRs changed as well. To explore these dynamics further, dimensionality reduction, clustering, and classification were utilized to recognize processing commonalities among tDRs. Of the classes we identified, one contains Arg-UCU tDRs as well as other highly expressed tDRs, with apparent sequence conservation at base pair 30:40 in the anticodon stem of the tRNA and conservation of other unique sequence motifs.

1.2 Results

1.2.1 Dynamic patterns of small RNA expression during cerebral cortex organoid differentiation

Human induced pluripotent stem cells (GM12878-c305 iPSCs) were grown into cerebral cortical organoids as a model to study tRNA changes in early human embryogenesis. A standard time course of culturing conditions was used to achieve neural induction and organoid growth (Figure 1.1A). Cells were aggregated into embryoid bodies and differentiated into cortical tissues over a period of 10 weeks. Samples were taken for characterization and sequencing at the developmental time points of day 0 (D0 - stem cells), day 14 (D14 - neural epithelium and radial glia neural progenitor cells), day 35 (D35 - emergence of deep layer neurons) and day 70 (D70 - continued generation of cortical projection neurons, emergence of outer radial glia) based on the emergence of key cell types in the course of dorsal cortex development in previous studies (Pollen et al. 2019).

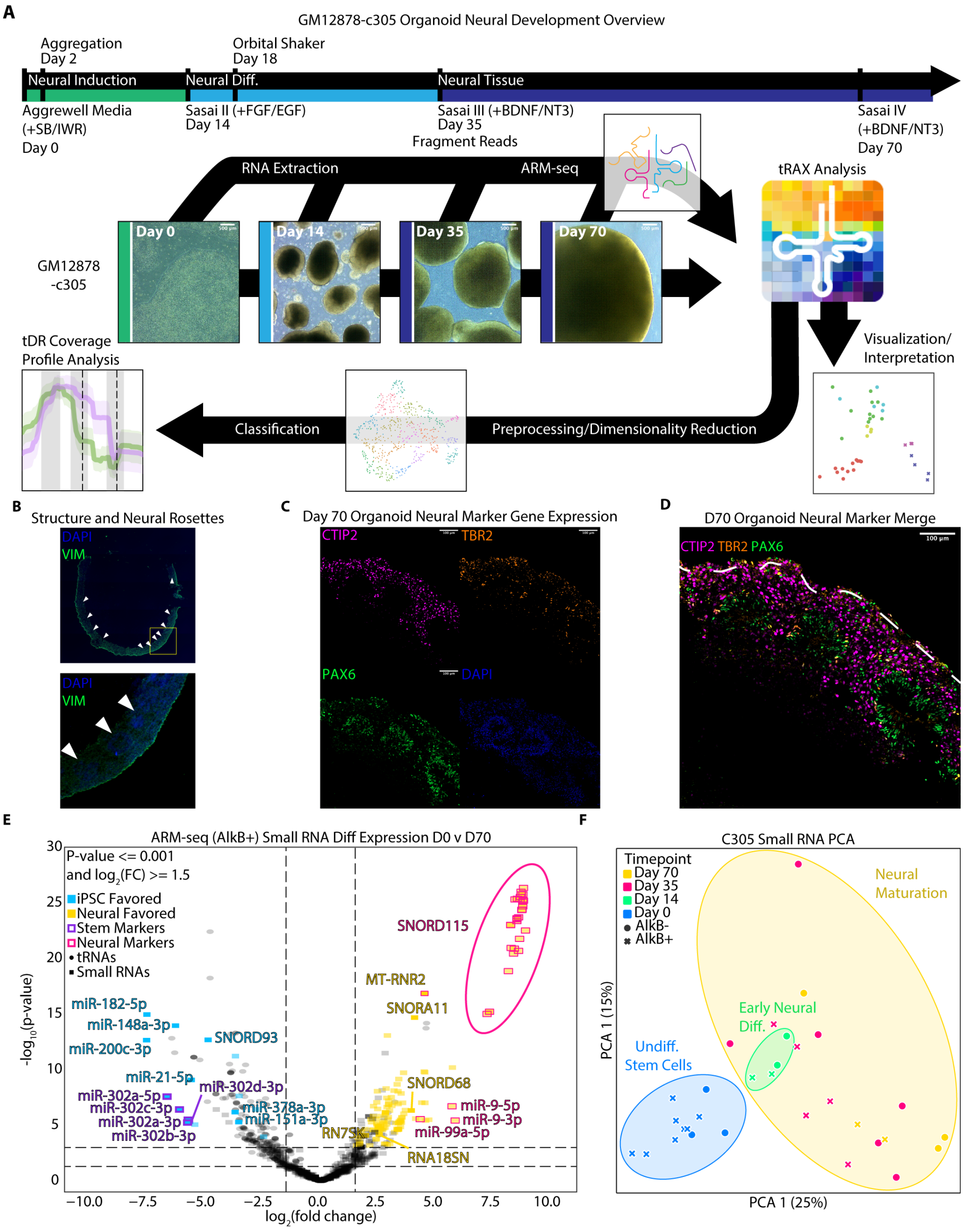

A
GM12878-c305 Organoid Neural Development Overview
Aggregation
Day 2
Orbital Shaker
Day 18
Neural Induction
Neural Diff.
Neural Tissue
Aggrewell Media
(+SB/IWR)
Day 0
Sasai II (+FGF/EGF)
Day 14
Sasai III (+BDNF/NT3)
Day 35
Sasai IV
(+BDNF/NT3)
Day 70
Fragment Reads
RNA Extraction
ARM-seq
tRAX Analysis
GM12878
-c305
Day 0
Day 14
Day 35
Day 70
tDR Coverage
Profile Analysis
Visualization/
Interpretation
Classification
Preprocessing/Dimensionality Reduction
B
Structure and Neural Rosettes
DAPI
VIM
DAPI
VIM
C
Day 70 Organoid Neural Marker Gene Expression
CTIP2
TBR2
PAX6
DAPI
D
D70 Organoid Neural Marker Merge
CTIP2 TBR2 PAX6
100 µm
E
ARM-seq (AlkB+) Small RNA Diff Expression D0 v D70
P-value <= 0.001
and log2(FC) >= 1.5
iPSC Favored
Neural Favored
Stem Markers
Neural Markers
tRNAs
Small RNAs
miR-182-5p
miR-148a-3p
miR-200c-3p
SNORD93
miR-21-5p
miR-302d-3p
miR-302a-5p
miR-302c-3p
miR-302a-3p
miR-302b-3p
miR-378a-3p
miR-151a-3p
RN7SK
SNORD115
MT-RNR2
SNORA11
SNORD68
miR-9-5p
miR-9-3p
miR-99a-5p
RNA18SN
-log10(p-value)
log2(fold change)
30
25
20
15
10
5
0
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
F
C305 Small RNA PCA
Timepoint
Day 70
Day 35
Day 14
Day 0
AlkB-
AlkB+
Neural
Maturation
Early Neural
Diff.
Undiff.
Stem Cells
PCA 1 (15%)
PCA 1 (25%)

Figure 1.1

Figure 1.1: Human cerebral cortical organoids show neural marker gene expression and small RNA expression diversity compared to stem cells

(A) *Graphical abstract of the project depicting media and small molecule drug changes across organoids developmental time-course. Representations of organoids and relative size at crucial time points are shown with 500μm scale bars. RNA collection, tRNA sequencing, analysis pipeline and classification are also depicted.* ***(B)*** *D70 organoid stains using vimentin and dapi are used showing cytoarchitecture of the organoid. Emphasis placed with white arrows showing locations or neural rosettes (early endo tube like formation) found throughout the perimeter of the organoid stains.* ***(C)*** *Antibody stains PAX6 a radial glial marker, CTIP2 a deep layer neuron marker, and TBR2 an intermediate neural progenitor marker, show neural marker gene expression with PAX6 staining the inside of neural rosettes, TBR2 surrounding the rosettes and CTIP2 encompassing the exterior of the organoid.* ***(D)*** *Neural marker stains were merged with a dotted white line for the exterior perimeter of the organoid.* ***(E)*** *Volcano plot showing D0 vs. D70 log$_2$ fold-changes in abundance of small RNAs against -log$_1 0$(p-value) with tRNAs shown for scale. Targets of particular interest are labeled with horizontal dotted lines at p-value=0.05 and p-value=0.001 and vertical dotted lines at 1.5 fold difference. Small RNAs highly expressed in either stem or neural cells across multiple conditions (D0 vs. D14, D0 vs. D35, or D0 vs. D70) were highlighted in blue or yellow respectively. Small RNAs known as stem or neural markers were highlighted in purple or magenta as well.* ***(F)*** *PCA plot of all ARM-seq samples for all measured small RNAs across organoid time points with PC1 seeming to represent organoid maturation.*

D70 organoids expressed expected neural cell-type markers using immunofluorescence staining (IF-staining) when compared against D0 stem cells (Figure 1.1). We observe circular structures known as neural rosettes (early neural tube-like formations) throughout the perimeter of the organoid that stain with antibodies for PAX6 (a radial glia neural stem cell marker), and vimentin (a radial glial fiber marker) showing the presence of radial glia neural stem cells (Figure 1.1B-D). The PAX6+ neural rosettes were surrounded by TBR2+ intermediate neural progenitors and CTIP+ neurons beyond the neural rosettes and intermediate progenitors (Figure 1.1D). In general, the presence of different neural subtypes increased over time as the organoids developed and increased in size.

To establish the patterns of tRNA and tDR expression in neural development, we performed differential expression analysis using tRNA Analysis of eXpression (tRAX) (Holmes et al. 2022) on read data across the differentiation time course (D0, D14, D35, D70). We validated that the small RNA sequencing reflected the neural differentiation with small RNA markers known to have specific expression in neurons or stem cells (Figure 1.1E). Known stem-associated mir302 family miRNAs were more highly expressed in D0 stem cells, whereas neural-associated miR-9 and miR-LET7 (miR-99a) family miRNAs were more highly expressed in the D70 organoids (Coolen, Katz, and Bally-Cuif 2013; Cimadamore et al. 2013; Zhao et al. 2010; Fairchild et al. 2019) (Table S1.1). MT-RNR2, a mitochondrial ribosomal RNA found to play a neuroprotective role, is also upregulated in organoids relative to stem cells (Hashimoto, Niikura, Tajima, et al. 2001; Hashimoto, Niikura, Ito, et al. 2001; Ying et al. 2004). Many SNORD115 RNAs, known to be induced by neuronal differentiation (Bratkovič et al. 2018), were also highly expressed in neural cell time points, confirming expected small RNA transcriptional changes during the organoid time-course experiment. Principal Components Analysis (PCA) was performed on small RNA read counts across all time points (Figure 1.1F), and showed that biological replicates and adjacent time points cluster closest to one another, as expected.

1.2.2 Broad changes in tRNA Isotypes do not capture tDR-Specific Changes

The relative expression of the major classes of small RNAs was assessed across the developmental time course, with ARM-seq picking up many small RNAs in addition to tRNAs, and tRNA reads increasing with AlkB treatment (Figure 1.2A). The diversity of tRNA isotype groups was analyzed in the AlkB positive conditions to look for changes across time points, as it had increased tRNA reads and diversity (Figure S1.1). Ala, Asn, iMet, Phe, SeC, and Ser had greater than two fold-change increases between D0 and D70 (Figure 1.2B). These same groups also appeared to have the highest expression differences between D0 and D35, with all but Ser having a greater than two-fold increase in the neural condition. In parallel, more than 40% of Glu, His, Lys, Thr, and Tyr diminished in D0 vs. D70, with His and Tyr also decreasing in the D0 vs. D35 condition.

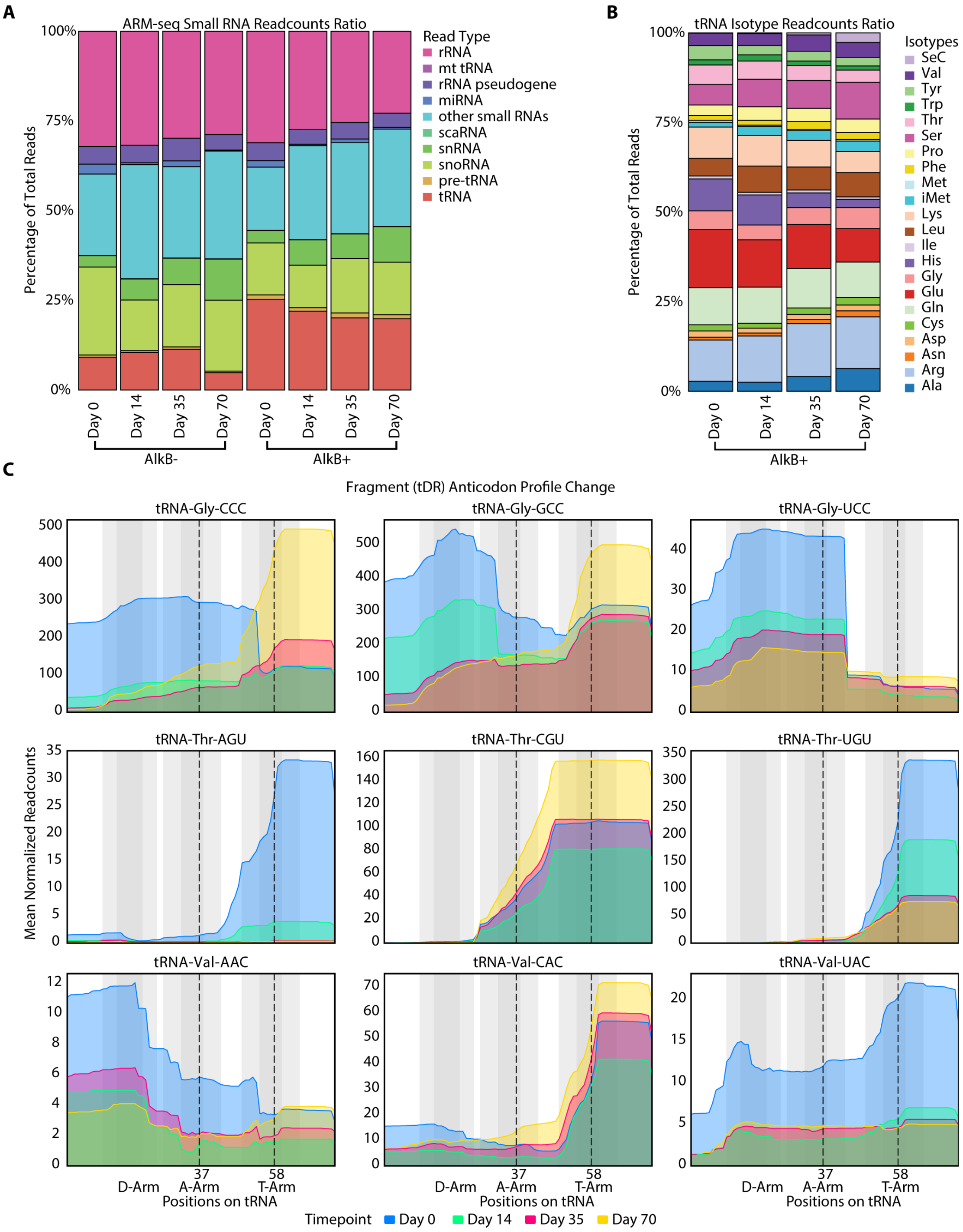

A
ARM-seq Small RNA Readcounts Ratio
100%
75%
50%
25%
0%
Percentage of Total Reads
Read Type
rRNA
mt tRNA
rRNA pseudogene
miRNA
other small RNAs
scaRNA
snRNA
snoRNA
pre-tRNA
tRNA
Day 0
Day 14
Day 35
Day 70
Day 0
Day 14
Day 35
Day 70
AlkB-
AlkB+
B
tRNA Isotype Readcounts Ratio
100%
75%
50%
25%
0%
Percentage of Total Reads
Isotypes
SeC
Val
Tyr
Trp
Thr
Ser
Pro
Phe
Met
iMet
Lys
Leu
Ile
His
Gly
Glu
Gln
Cys
Asp
Asn
Arg
Ala
Day 0
Day 14
Day 35
Day 70
AlkB+
C
Fragment (tDR) Anticodon Profile Change
tRNA-Gly-CCC
tRNA-Gly-GCC
tRNA-Gly-UCC
tRNA-Thr-AGU
tRNA-Thr-CGU
tRNA-Thr-UGU
tRNA-Val-AAC
tRNA-Val-CAC
tRNA-Val-UAC
0
100
200
300
400
500
0
10
20
30
40
0
5
10
15
20
25
30
35
0
20
40
60
80
100
120
140
160
0
50
100
150
200
250
300
350
0
2
4
6
8
10
12
0
10
20
30
40
50
60
70
0
5
10
15
20
Mean Normalized Readcounts
37
58
D-Arm
A-Arm
T-Arm
Positions on tRNA
Timepoint
Day 0
Day 14
Day 35
Day 70

Figure 1.2

Figure 1.2: tRNA and tDR changes across early human cortical neurogenesis

(A) *Human small RNA reads-counts distributions showing relative ratio of small RNAs across organoid time-course, sequencing type and alkB treatment.* ***(B)*** *Human tRNA isotype reads-counts distributions showing relative ratio of tRNA isotypes across organoid time-course, sequencing type and alkB treatment.* ***(C)*** *Coverage profiles for the Gly, Thr, and Val isotype groups showing max expression across AlkB+ samples per position within each isotype across organoid time points. D-loop, anticodon-loop, and T-loop shown in gray, with lines added at specific coverage breakpoints in each plot.*

Relative fragment profiles and their expression as a function of time seem to be dynamic. Gly, Thr, and Val tRNA isotypes showed changes in expression and diversity in the tDRs between isodecoders as well as between developmental time points (Figure 1.2C). In Gly a 5' fragment is evident in D0 samples, and this fragment diminishes over time, although at different rates in the three isodecoders. However, in Gly-CCC and to a lesser extent in Gly-GCC, a 3' fragment increases in expression as the cells differentiate. In the Thr isotype, a 3' fragment shape appears in all cases, however, both Thr-AGU and Thr-UGU isodecoders decrease in expression and share a decrease in coverage at position 58. In contrast, Thr-CGU isodecoders increase in expression over time and have coverage beyond position 58, suggesting that modifications at this position may play a role in the processing and abundance of this fragment. Val tRNA isotypes have the widest variety of tDR coverage profiles with a distinct pattern for each time point and isodecoder. In addition to Gly, Thr, and Val, other isotypes also had distinct relative fragment profiles among their isodecoders, but the changes in the relative expression of their codon pools was less dynamic (Figure S1.2A).

1.2.3 Expression patterns of neural tDRs are unique and change over time

Individual tDRs were chosen to be further assessed rather than isodecoder groups or full-length reads (Due to their abundance in ARM-seq) in order to find neural specific fragments and thier dynamic expression patterns (Table S1.2). tDR reads were divided into five main subgroups consisting of whole tRNA reads (within 10 bases of the start and of the full tRNA sequence), 5' fragments (within 10 bases of the start position but do not reach end), 3' fragments (within 10 bases of end position, but do not reach start), other fragments (internal fragments), and pre-tRNAs (pre-partial tRNAs and pre-tRNAs). The expression of D0 vs. D70 tDRs was compared for all tDRs and all pre-tRNA type reads (Figure 1.3A). Between D0 and D70 samples, 27 tDRs were found to have a significant (p-value <= 0.05) and $\log_2$(fold-change) greater than 1.5 increase in expression in the cerebral cortex organoids relative to undifferentiated stem cells (Supp. Table 2). The differences between D0 and D14 samples were far less significant, likely since neural induction has just started and the space between time points is much closer. As such, tDR changes after neural differentiation were mostly examined between D0/D35 and D0/D70 time points. Among the top rank-expressed neural tDRs, several amino families appeared multiple times: Ala (4), Arg (3), Gly (2), and Leu (3).

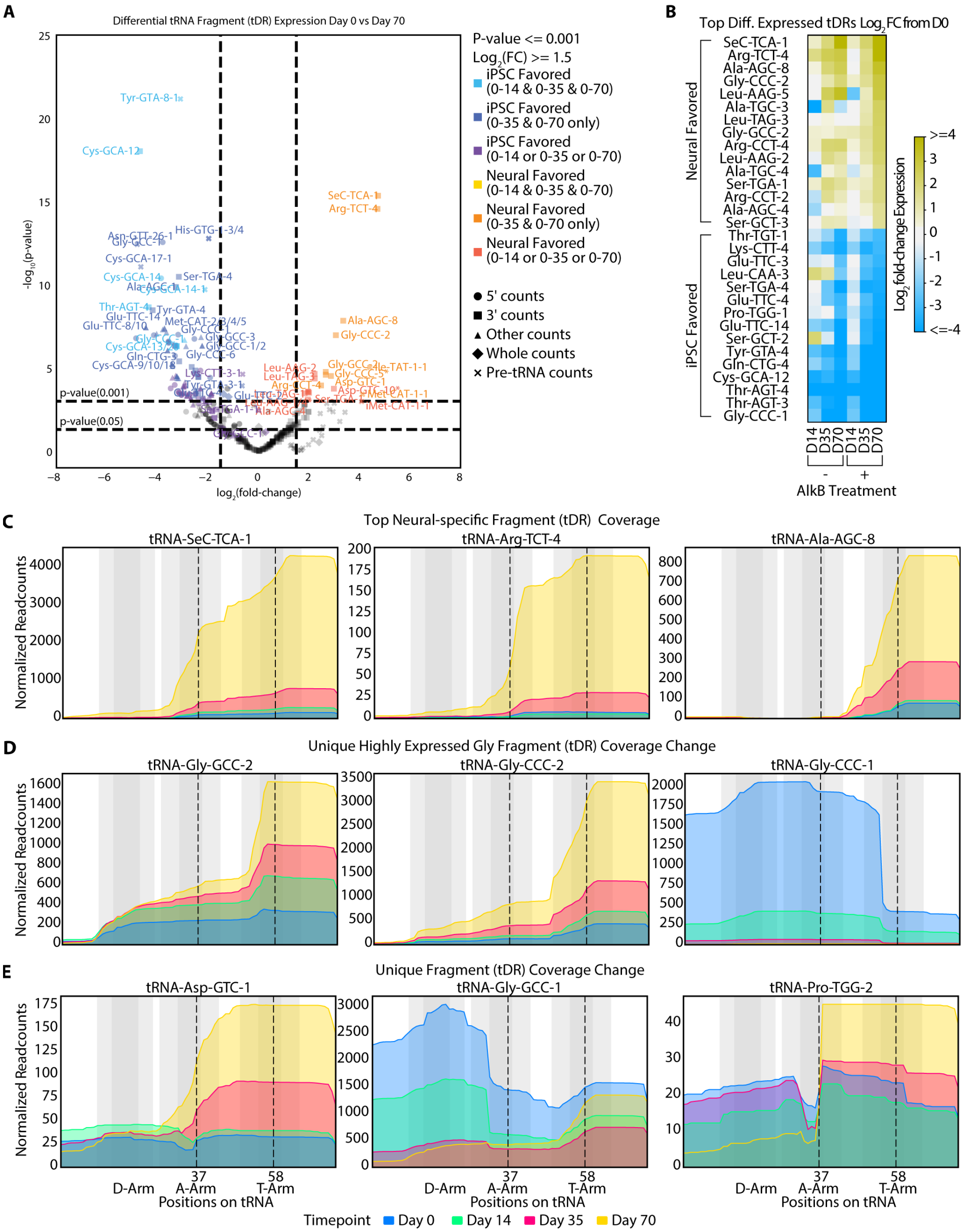

A
Differential tRNA Fragment (tDR) Expression Day 0 vs Day 70
-log10(p-value)
log2(fold-change)
25
20
15
10
5
0
−8
−6
−4
−2
0
2
4
6
8
Tyr-GTA-8-1
Cys-GCA-12
SeC-TCA-1
Arg-TCT-4
His-GTG-1-3/4
Asn-GTT-26-1
Gly-GCC-1
Cys-GCA-17-1
Cys-GCA-14
Ser-TGA-4
Ala-AGC-1
Cys-GCA-14-1
Thr-AGT-4
Tyr-GTA-4
Glu-TTC-14
Met-CAT-2/3/4/5
Glu-TTC-8/10
Gly-CCC-1
Gly-GCC-3
Gly-CCC-1
Gly-GCC-1/2
Cys-GCA-13/2
Gly-CCC-6
Gln-CTG-3
Cys-GCA-9/10/18
Ala-AGC-8
Gly-CCC-2
Leu-AAG-2
Gly-GCC-2
Ile-TAT-1-1
Lys-CTT-3-1
Leu-TAG-3
Gly-CCC-5
Tyr-GTA-3-1
Arg-TCT-4
Asp-GTC-1
Glu-TTC-2
Asp-GTC-10
Met-CAT-1-1
Ser-TGA-1
iMet-CAT-1-1
Ala-AGC-4
Gly-GCC-1
p-value(0.001)
p-value(0.05)
P-value <= 0.001
Log2(FC) >= 1.5
iPSC Favored (0-14 & 0-35 & 0-70)
iPSC Favored (0-35 & 0-70 only)
iPSC Favored (0-14 or 0-35 or 0-70)
Neural Favored (0-14 & 0-35 & 0-70)
Neural Favored (0-35 & 0-70 only)
Neural Favored (0-14 or 0-35 or 0-70)
● 5' counts
■ 3' counts
▲ Other counts
◆ Whole counts
× Pre-tRNA counts
B
Top Diff. Expressed tDRs Log2 FC from D0
Neural Favored
SeC-TCA-1
Arg-TCT-4
Ala-AGC-8
Gly-CCC-2
Leu-AAG-5
Ala-TGC-3
Leu-TAG-3
Gly-GCC-2
Arg-CCT-4
Leu-AAG-2
Ala-TGC-4
Ser-TGA-1
Arg-CCT-2
Ala-AGC-4
Ser-GCT-3
iPSC Favored
Thr-TGT-1
Lys-CTT-4
Glu-TTC-3
Leu-CAA-3
Ser-TGA-4
Glu-TTC-4
Pro-TGG-1
Glu-TTC-14
Ser-GCT-2
Tyr-GTA-4
Gln-CTG-4
Cys-GCA-12
Thr-AGT-4
Thr-AGT-3
Gly-CCC-1
D14
D35
D70
D14
D35
D70
-
+
AlkB Treatment
Log2 fold-change Expression
>=4
3
2
1
0
-1
-2
-3
<=-4
C
Top Neural-specific Fragment (tDR) Coverage
tRNA-SeC-TCA-1
tRNA-Arg-TCT-4
tRNA-Ala-AGC-8
Normalized Readcounts
4000
3000
2000
1000
0
200
175
150
125
100
75
50
25
0
800
700
600
500
400
300
200
100
0
D
Unique Highly Expressed Gly Fragment (tDR) Coverage Change
tRNA-Gly-GCC-2
tRNA-Gly-CCC-2
tRNA-Gly-CCC-1
Normalized Readcounts
1600
1400
1200
1000
800
600
400
200
0
3500
3000
2500
2000
1500
1000
500
0
2000
1750
1500
1250
1000
750
500
250
0
E
Unique Fragment (tDR) Coverage Change
tRNA-Asp-GTC-1
tRNA-Gly-GCC-1
tRNA-Pro-TGG-2
Normalized Readcounts
175
150
125
100
75
50
25
0
3000
2500
2000
1500
1000
500
0
40
30
20
10
0
37
58
D-Arm
A-Arm
T-Arm
Positions on tRNA
Timepoint
Day 0
Day 14
Day 35
Day 70

Figure 1.3

Figure 1.3: Groups of tDRs change during neural differentiation

***(A)** Volcano plot showing D0 vs. D70 log$_2$ fold-changes in abundance of tDRs against -log$_1$0(p-value) with 3', 5' and internal fragment counts and whole length and pre-mature tRNA counts. Targets of particular interest are labeled with horizontal dotted lines at p-value=0.05 and p-value=0.001 and vertical dotted lines at 1.5 fold difference. tDRs highly expressed in either stem or neural cells across multiple conditions (D0 vs. D14, D0 vs. D35, or D0 vs. D70) were highlighted in shades of blue or yellow depending on how many conditions they appeared in. All samples were batch corrected and normalized using Deseq2 (Love, Huber, and Anders 2014) to account for differences in sample read counts. **(B)** Heatmap showing the top 15 neural and top 15 stem favored tDRs across all conditions. Names in orange have less than 80 uniquely mapping read counts. **(C-E)** Coverage profiles for the top 3 highly expressed tDRs over organoid neurogenesis **(C)**, highly expressed stem and neural Gly tDRs **(D)**, and unique tDRs with coverage changes **(E)**. D-loop, anticodon-loop, and T-loop shown in gray, with lines added at specific coverage breakpoints in each plot.*

Nearly all tDRs that increase after neural differentiation are 3' end-counts (3' fragments that terminate within 10 bp of the CCA tail), with several pre-mature-tRNA counts also being detected. The top 15 neural and top 15 stem-expressed tDRs between D0 and D70 were compared (Figure 1.3B), and the p-values and read counts were assessed to determine overall significance and expression (Figure S1.3A,B). While 5' end-counts (5' fragments that terminate within 10 bp of the tRNA start site) are less common when compared to 3' end-counts, they tend to appear in D0 samples and diminish entirely after neural differentiation in both D35 and D70 time points. With many 3' tDRs being expressed at D35 and D70, there is also great diversity in the shape of fragment coverage with drop-offs of coverage at different conserved tRNA positions. This suggests differential modifications of the mature-tRNA the tDR is derived from since mature tRNA modifications often act as signals for cleavage creating tDRs (Lyons, Fay, and Ivanov 2018).

The diversity of the top 3 most neural expressed tDR 3' fragments was explored in detail (Figure 1.3C). While coverage drop-off commonly occurred near position 58, other coverage changes differed between tRNA genes. tRNA-SeC-TCA-1 and tRNA-Arg-TCT-4 derived tDRs, for example, have a unique coverage drop through the anticodon loop, making these tDRs match more traditional 3' half designations. Additionally, while both tRNA-Arg-TCT-4 and tRNA-Arg-TCT-1 increase in expression over time, the former has severe drop off in coverage around position 40, whereas the later persists into the D-arm before dropping off (Figure S1.3C). tRNA-Ala-AGC-8-derived tDRs, however, had distinct drops in coverage between the anticodon and t-loops with residual coverage encompassing the remainder of the tRNAs canonical position. tRNA-Gly-GCC-2-, and tRNA-Gly-CCC-2-derived tDRs also had distinct drops in coverage and appeared in the top 10 most neural expressed tDRs in contrast to tRNA-Gly-CCC-1 (The most stem expressed tDR) (Figure 1.3D). Interestingly both neural specific Gly tDRs (tRNA-GCC-2/CCC-2) were 3' specific whereas tRNA-Gly-CCC-1 appeared to be highly 5' biased with dropoff before position 58. tRNA-Sec-TCA-1, tRNA-Arg-TCT-4, and tRNA-Ala-AGC-8 were validated as increasing in full length tRNA expression across neural differentiation via Northern blots (Figure S1.3D). Most tDRs that maintain the same fragment coverage over the timecourse with changes in relative abundance rather than the fragment that is expressed changing.

Additional tRNAs showed pronounced changes in the tDR fragments generated over this developmental time course. tRNA-Asp-GTC-1 derived tDRs look similar in shape and expression at D0 and D14 but gain a large increase in 3' fragment expression at D35 and D70 (Figure 1.3E). tRNA-Gly-GCC-1-derived tDRs have a complementary effect with the diminishment of the 5' end coverage as the organoids reach D35 and D70 time points. tRNA-Pro-TGG-2 tDRs exhibit a mix of both effects, with a decrease in 5' fragments paired with an increase in 3' fragments over time. Since excitatory neuron formation increases between D35 and D70, 3' fragments may be associated with these neurons. tRNA-Arg-TCT-2 and tRNA-Ser-GCT-5 tDRs form a unique group that has a preference for higher expression in D0 and D70 relative to D14 and D35, which may suggest lower expression of these 3' tDRs in neural progenitor cells (Figure S1.3E).

1.2.4 Multidimensional analysis of tDRs facilitates the discovery of complex relationships

To assess the full diversity of unique tRNA expression, clustering and classification of fragment profiles split across different time points was performed. This was done aggregating read coverage, read-starts/ends, deletions and misincorporation information (suggesting modified positions) and clustering the combined dataset using UMAP (Uniform Manifold Approximation and Projection) followed by classification via hdbscan (McInnes, Healy, and Astels 2017; McInnes, Healy, and Melville 2018). The resulting clusters were then manually annotated into neural and stem categories as described in the methods. This revealed a highly complex, diverse set of tRNAs, each comprising multiple isotypes, and neural-stem expression levels with no bias in time point observed (Figure 1.4A). Three major groups of HDBScan clusters were identified (consisting of 21 specific subgroups) labeled A (stem), B (neural), and C (neutral) as well as the hdbscan unannotated category (U) for tRNAs that were too hard to class into a single cluster. Spatially clusters tended to correlate with isotype and neural/stem favored tRNAs with no observable bias in time point. We next explored the relationship between isotype and cluster to see if genomic traits could also be correlated to tDR expression.

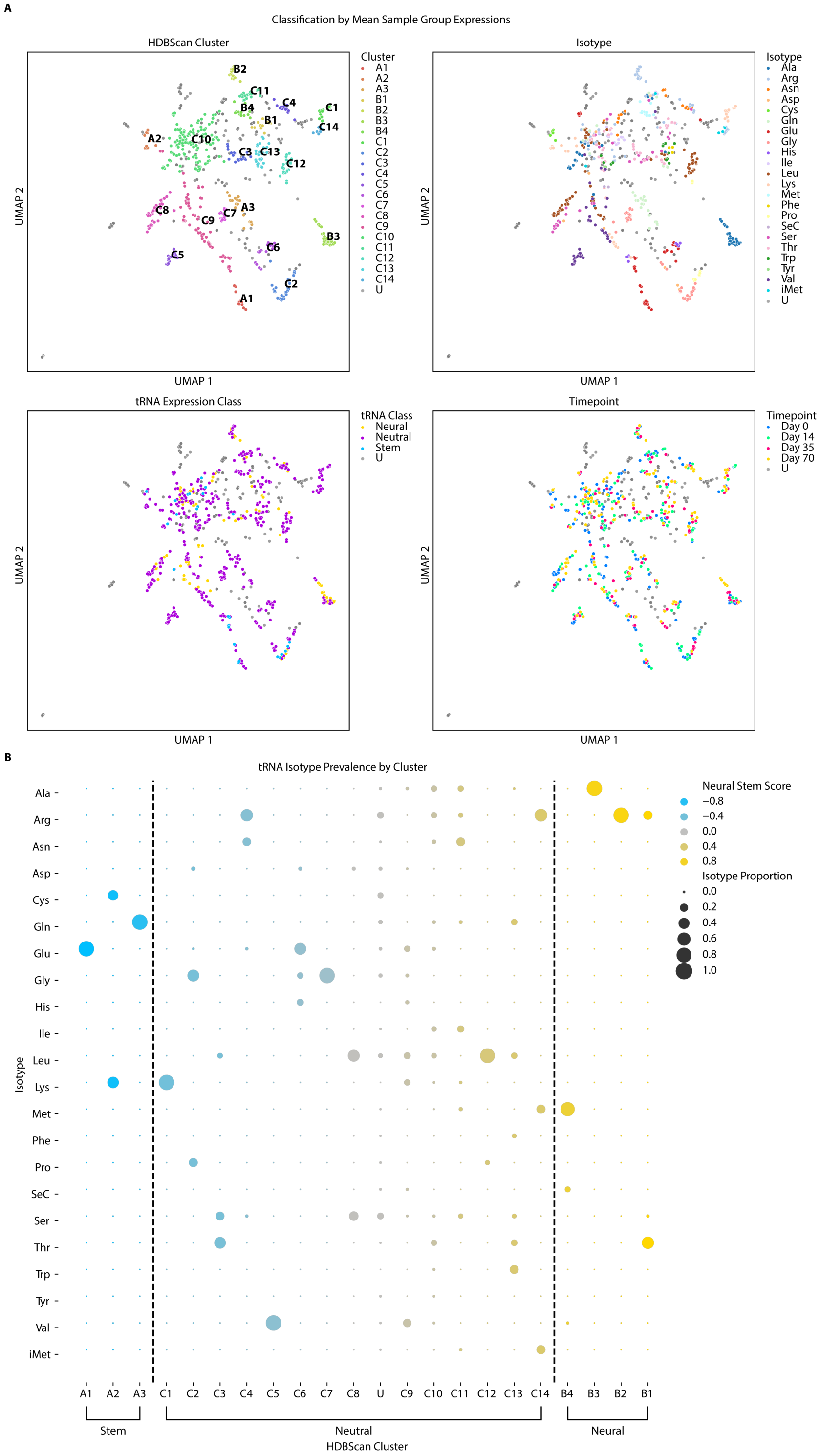

A
Classification by Mean Sample Group Expressions
HDBScan Cluster
B2
C11
B4
C4
C1
B1
C14
A2
C10
C3
C13
C12
C8
A3
C7
C9
B3
C6
C5
C2
A1
UMAP 1
UMAP 2
Cluster
A1
A2
A3
B1
B2
B3
B4
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
U
Isotype
Isotype
Ala
Arg
Asn
Asp
Cys
Gln
Glu
Gly
His
Ile
Leu
Lys
Met
Phe
Pro
SeC
Ser
Thr
Trp
Tyr
Val
iMet
U
tRNA Expression Class
tRNA Class
Neural
Neutral
Stem
U
Timepoint
Timepoint
Day 0
Day 14
Day 35
Day 70
U
B
tRNA Isotype Prevalence by Cluster
Isotype
Ala
Arg
Asn
Asp
Cys
Gln
Glu
Gly
His
Ile
Leu
Lys
Met
Phe
Pro
SeC
Ser
Thr
Trp
Tyr
Val
iMet
A1
A2
A3
C1
C2
C3
C4
C5
C6
C7
C8
U
C9
C10
C11
C12
C13
C14
B4
B3
B2
B1
Stem
Neutral
Neural
HDBScan Cluster
Neural Stem Score
−0.8
−0.4
0.0
0.4
0.8
Isotype Proportion
0.0
0.2
0.4
0.6
0.8
1.0

Figure 1.4

Figure 1.4: tRNA-seq allows for clustering of tDRs into functional groups

(A) *UMAP projections of tRNA sequencing profiles using unique coverage, 3' and 5' read ends (with consideration to the tRNA structure), deletions, and mismatched bases at position after hdbscan clustering and manual annotation into stem (A), neural (B), neutral (C), and unannotated (U) classes. Projections of isotype, neural/stem tRNAs, and time point masked by U class are also shown.* ***(B)*** *Dot plot representation of the hdbscan clusters with dot size correlating to proportion of isotype found in each cluster and the color based on the mean neural stem score (as defined in methods) with vertical lines at the top and bottom 25th percentiles defining the groups.*

The clusters were often dominated by specific tRNA isotypes with several clusters containing greater than 50% of specific isotypes: A1 (Glu), A2 (Gly), A3 (Gln), C1 (Lys), C2 (Asp), C3 (Leu), C5 (Val), C6 (His), C7 (Gly), C8 (Ser), C12 (Leu), C14 (Arg), B1 (Ala), B2 (Arg), and B3 (Ala) (Figure 1.4B). The clusters often have distinct isotype (Figure S1.4A), AlkB treatment (Figure S1.4B), and neural expression patterns (Figure S1.4D) defined by their sequencing characteristics making the uniquely identifiable. Cluster C10 and C11 are exceptions to this both expressing very diverse isotype compositions and seeming to correlate more closely with the unannotated (U) category. This can likely be explained by low mean read coverage in clusters C10 and C11 relative to other clusters making them hard to annotate (Table S1.3). AlkB conditions were generally diverse with a slight preference for AlkB+ conditions in clusters with higher neural expression. We also checked subtle genomic features of tRNAs such as tRNAscanSE covariance, HMM, and secondary structure scores (Infernal - Inference of RNA Alignments) against the clusters and found that scores tended to match within clusters (Figure S1.4C). These scores measure how much each individual tRNA sequence matches its canonical isotype which matches what we observed with individual isotypes defining certain clusters. This suggests that differential expression and tDR formation are influenced by these genomic features but most likely they are not the only driving factor (Chan, Lin, et al. 2019; Nawrocki and Sean R. Eddy 2013).

Looking further into the relationship between isotype and cluster, we looked at Arg specific tDRs. Disregarding U, C10, and C11 (due to aforementioned indiscernibility) Arg tDRs mainly are split across B1, B2, C4, and C14 clusters all residing in the same spatial area as one another (Figure S1.4D). Interestingly Arg-TCT-4 shows up in both B1 (neural) and C10 (neutral) clusters with D70 Arg-TCT-4 appearing in B1. This is most likely due to the extreme change in expression and coverage after neural differentiation with lower read coverage harder to discern. This indicates that other “neutral” tDRs appearing in C clusters may also be neurally associated but can be hard to discern with low read coverage. B1 clusters also contain Arg-TCG-1 and Arg-TCG-5 tDRs, with B2 clusters comprised almost entirely of Arg-CCT tDRs, all tRNAs that increase moderately over differentiation. Interestingly Arg-TCT-1 (which also increases across neural differentiation) appears in C10 (neutral) and U (unannotated) clusters with D70 appearing in C10, this mimics the change over the time course seen in Arg-TCT-4 albeit from U to C10 rather than C10 to B1. While both Arg-TCT-1 and Arg-TCT-4 increase in neural expression, the former has less of an expression increase than the later, and they express entirely different fragments. This revealed that Arg tDR expression as a function of neural expression is correlated to specific tRNA transcripts rather than specific isotypes and subtle changes to genomic sequence correlate with broad effects in the type and expression levels of tDRs.

1.2.5 Sequence characteristics of human neural tDR-generating tRNAs are conserved across multiple amino groups

To further understand what causes specific neural tDR shape and expression we looked at tRNA modifications and their enzymes and how they correlated with underlying characteristics of the parent tRNA. Modification data from Zhang et al. 2024 was taken and projected onto the clusters showing that specific modifications such as acp^{3}U20, m^{3}C20, and m^{1}I37 correlate strongly with just neural clusters (B), and acp^{3}U20a, m^{2,2}G26, m^{3}C32, and I34 correlate with neural and neutral (B and C) clusters (Figure 1.5A). Inosine (at position 37) has been associated with intellectual disabilities, and microcephaly and is added by ADAT3 in humans. m^{2,2}G (a modification found at position 26 in Arg-TCT-4) has been associated with intellectual disabilities and microcephaly via TRMT1 (Crécy-Lagard et al. 2019; Chujo and Tomizawa 2021). m^{1}A58 is a ubiquitous modification found in nearly all tRNAs and is consistently expressed across all groups. While not present in our modification data PUS3 and NSUN2 are known to play a roles in pseudouridylation and m^{5}C addition at positions 39 and 40, respectively, and are both tied to intellectual development (J. Chen and Patton 2000; Shaheen et al. 2016; Tuorto et al. 2012). We found that PUS3 and NSUN2 increased while ADAT3 decreased over a five week time course of human embryonic stem cell derived cortical organoids using data from Fiddes, Lodewijk, et al. 2018 (Figure S1.4E).

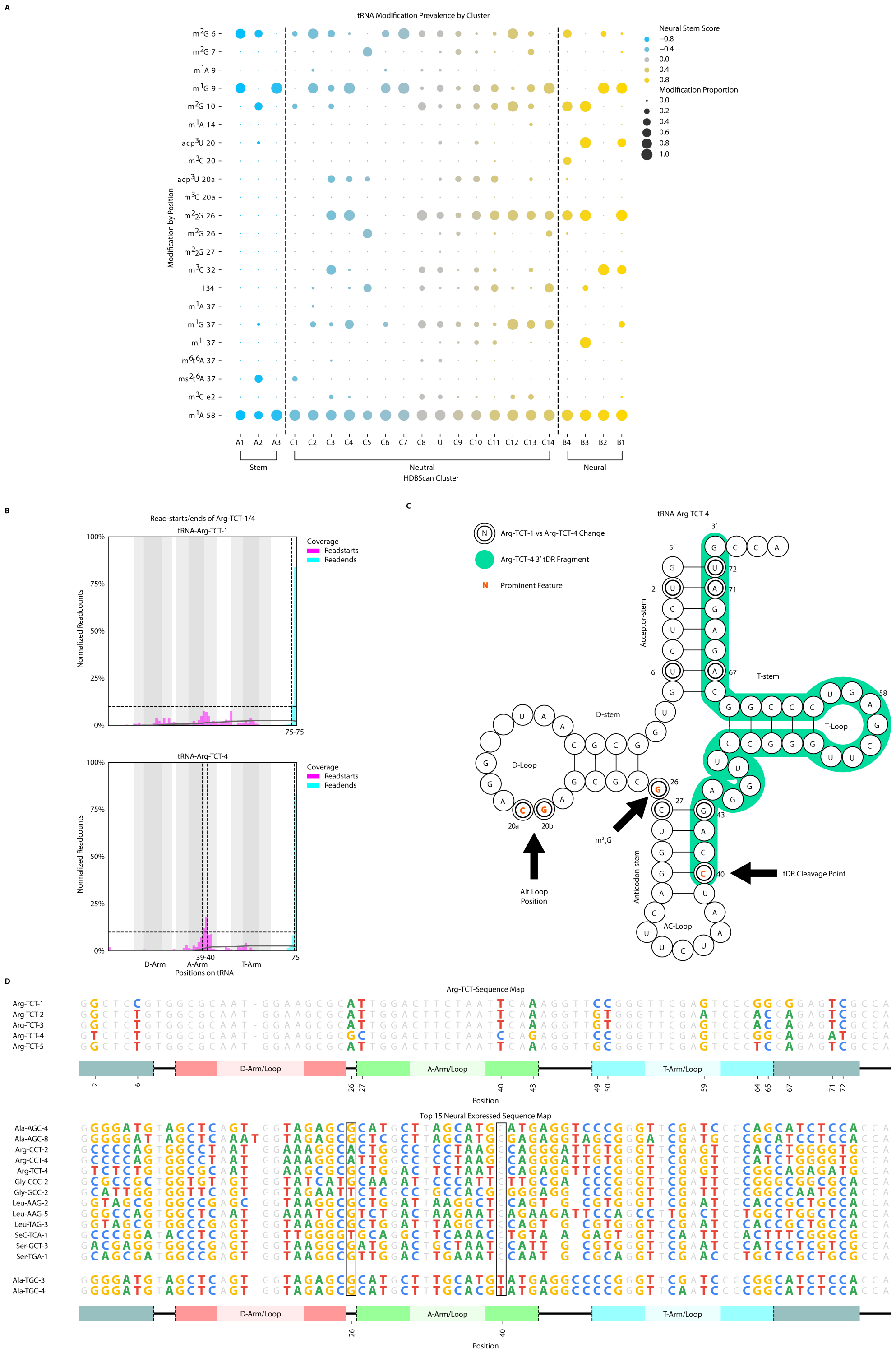

A
tRNA Modification Prevalence by Cluster
Modification by Position
m2G 6
m2G 7
m1A 9
m1G 9
m2G 10
m1A 14
acp3U 20
m3C 20
acp3U 20a
m3C 20a
m22G 26
m2G 26
m22G 27
m3C 32
I 34
m1A 37
m1G 37
m1I 37
m6t6A 37
ms2t6A 37
m3C e2
m1A 58
A1 A2 A3 C1 C2 C3 C4 C5 C6 C7 C8 U C9 C10 C11 C12 C13 C14 B4 B3 B2 B1
Stem
Neutral
Neural
HDBScan Cluster
Neural Stem Score
−0.8
−0.4
0.0
0.4
0.8
Modification Proportion
0.0
0.2
0.4
0.6
0.8
1.0
B
Read-starts/ends of Arg-TCT-1/4
tRNA-Arg-TCT-1
Coverage
Readstarts
Readends
Normalized Readcounts
100%
75%
50%
25%
0%
75-75
tRNA-Arg-TCT-4
39-40
75
D-Arm
A-Arm
T-Arm
Positions on tRNA
C
tRNA-Arg-TCT-4
Arg-TCT-1 vs Arg-TCT-4 Change
Arg-TCT-4 3' tDR Fragment
Prominent Feature
Acceptor-stem
D-stem
D-Loop
T-stem
T-Loop
Anticodon-stem
AC-Loop
Alt Loop Position
m22G
tDR Cleavage Point
5'
3'
20a
20b
26
27
43
40
58
67
71
72
2
6
D
Arg-TCT-Sequence Map
Arg-TCT-1
Arg-TCT-2
Arg-TCT-3
Arg-TCT-4
Arg-TCT-5
D-Arm/Loop
A-Arm/Loop
T-Arm/Loop
Position
2 6 26 27 40 43 49 50 59 64 65 67 71 72
Top 15 Neural Expressed Sequence Map
Ala-AGC-4
Ala-AGC-8
Arg-CCT-2
Arg-CCT-4
Arg-TCT-4
Gly-CCC-2
Gly-GCC-2
Leu-AAG-2
Leu-AAG-5
Leu-TAG-3
SeC-TCA-1
Ser-GCT-3
Ser-TGA-1
Ala-TGC-3
Ala-TGC-4
26
40
Position

Figure 1.5

Figure 1.5: Neural tDRs reveal sequence-specific characteristics

(A) *Dot plot representation of the hdbscan clusters with dot size correlating to proportion of modification at position found in each cluster and the color based on the mean neural stem score (as defined in methods) with vertical lines at the top and bottom 25th percentiles defining the groups.* ***(B)*** *The percentage of 3' and 5' read ends of Arg-TCT-1 and Arg-TCT-4 tDRs plotted against representative fragment profile curves. Positions with greater than 10% of relative normalized fragment readstarts/readends indicated with vertical dotted lines.* ***(C)*** *Schematic of Arg-TCT-4 with tertiary folding positions connected and Arg-TCT-1 divergence in sequence in bold.* ***(D)*** *Sequence map of Arg-TCT tRNAs and top 15 most neural expressed tDRs.*

Sequence alignments of the tRNAs in the neural clusters (B) revealed conservation of tRNA sequence at positions 26, 30, 32, and 40 (Figure S1.5A). This includes bases in close proximity to the cleavage at position 40. Finally, we wanted to see if sequence motifs (relative to background frequency) played a role in clusters as they have potential to affect tertiary structure formation. Motifs such as GGGG (1-4, 68-71), and CCCC (2-5, 60-63) and GCC were observed to be more common in neural (B) clusters (Figure S1.5B). The motif GCAT (26-29) also indicates a higher frequency of of G at position 26 which most likely corresponds to the higher frequency of m^{2,2}G26 found in the neural (B) clusters. ATG (28-30), GTA (39-41) and GCC (39-41 in neural, 40-42 in stem) indicated that the G-C pair between 30 and 40 occurs more frequently in neural groups but a G is more likely at position 40 in stem. With neutral (C) and unannotated (U) clusters comprising the majority of tRNAs found in the clustering, relative motif frequency was much closer to expected background.

Finally we also looked at the internal arrangement of the tDRs for both Arg-TCT-1 and Arg-TCT-4 by looking at the ratio of 3' read ends (in relation to the tRNA structure) to 5' read ends (Figure 1.5B). In Arg-TCT-4 read starts are grouped around position 39 and 40 indicating an association with having a C rather than U base changes at that position as found in other Arg-TCT isodecoders. Arg-TCT has five distinct genomic tRNA sequences that can exist for that isodecoder, with Arg-TCT-4 having a unique sequence relative to the others containing a 20b position G as well as 2T, 20aC, 26G, 27C, 40C, 43G, 71A, and 72T (Figure 1.5C). Further when we compared Arg-TCT-4 against the 15 top neural expressed tDRs we found conservation of C at position in 40 in all but Ala-TGC-3 and Ala-TGC-4 (Figure 1.5D). Interestingly Ala-TGC-3/4 both express more in D0 relative to D14 before increasing in D35 and D70 suggesting that they are active at different neural timepoints and might be part of a different neural expression pathway.

1.3 Discussion

tRNA sequencing data on human pluripotent stem cells and cerebral cortex organoids derived from them has yielded a wealth of previously unknown information. In particular, specific mature tRNAs, as well as tDRs derived from them, show striking changes in abundance during early human corticogenesis. We have identified many unique tRNAs/tDRs that change their expression profile and have categorized them. Further, the tDRs have been grouped together to assist in finding parallel traits shared between neural tDRs. In doing so we are able to combine both RNA sequencing features as well as genomic trait loci characteristics to provide a much better understanding of tRNA/tDRs associated with cortical neurogenesis.

tRNA-Arg-TCT-4 has been previously identified in mouse as a neural specific tRNA with defects leading to truncal ataxia and severe neurodegeneration when position 50 has a C to T mutation (Ishimura et al. 2014), here we show that the Arg-TCT-4 derived tDRs abundance increases across early human neurogenesis as well. This profile of brain organoids has also allowed for the discovery of new tDRs that are induced during neural differentiation, such as those from SeC-TCA-1 and Ala-AGC-8. Further, while most studies focus on single tRNAs/tDRs, using ARM-seq combined with cerebral brain organoids has allowed for the discovery of many tDRs with dynamic expression in early human embryonic development. These fragments each have unique breakpoints in their read coverage as well, indicative of a unique modification profile on a tRNA by tRNA basis. The top hits for neural specific tDRs seemed to share a coverage profile similar to one another, suggesting they may be generated by a common mechanism.

The profiles of these tDRs tended to fall into similar patterns that could be visually grouped together. For example, some of the top neural favored tDRs (Arg-TCT-4 and Ala-AGC-8) each have a 3' fragmentation pattern that is visually unique, reflecting a combination of associated sequence factors that allows for the fragments to be classified into distinct clusters that are proximal to one another (B1 and C13). This suggests that advanced sequence characteristics beyond unique read coverage drive the grouping of neural expressed tDRs. Pairing this information with known regulatory tDRs can be further used to understand the relationship between unique fragment profiles and tRNA regulation.

While this approach was used in the context of human cerebral cortex differentiation, identifying and classifying tDRs can be used in a broader context. Understanding the relationship of tDRs in a more expansive way than 3' and 5' derived can lead to a better understanding of tDR regulation overall. For example, the abundance of Arg-TCT-4 and Arg-TCT-1 both increase over neural development, however, Arg-TCT-4's associated 3' tDR spans from the anticodon to the end, whereas the one in Arg-TCT-1 starts in the D-loop instead. Arg-TCT-4 is also expressed at 1/10th the level of Arg-TCT-1, despite having a known regulatory role in the brain. Taken in conjunction, it is suggestive that the uniqueness of the Arg-TCT-4 fragment and not just the abundance of read coverage in mature neurons are what defines Arg-TCT-4 as a neural specific tDR regulator. As such tDRs roles in neurological development may go beyond just upregulation of expression across neural differentiation, and instead be a combination of genomic and RNA features (such as sequence motif and modifications) that play greater roles in regulation rather than translation.

1.4 Methods

1.4.1 Organoid methods

Human GM12878-c305 induced pluripotent stem cells (iPSCs) were generated from the GM12878 lymphoblastoid cell line (Coriell) using a proprietary episomal vector-based reprogramming method by Cellular Dynamics International (fujifilmcdi.com) and were confirmed to have a normal karyotype using the KaryoStat Karyotyping service (ThermoFisher). Undifferentiated cells were grown in feeder-free conditions on matrigel (Corning) with mTeSR Plus (Stemcell Technologies) or on vitronectin with Essential-8 Flex media (ThermoFisher). Cerebral cortex organoids were generated using a protocol adapted from Kadoshima et al. 2013. 10,000 cells per embryoid body were aggregated using AggreWell-800 plates in AggreWell media (Stemcell Technologies) supplemented with 10 uM Y-27632 rock inhibitor (Stemcell Technologies), and transferred to low attachment 6-well dishes (Corning) on day 2. These methods supplemented the respective media with 10 uM SB431542 (SB, Millipore), and 1 uM IWR-1 (Millipore) for the first 14 days of differentiation. The media was then changed to Sasai II media (DMEM/F12 + glutamax supplemented with N2) on day 14. At this point, organoids were supplemented with 10ng/mL beta fibroblast growth factor (bFGF) and 10ng/mL epidermal growth factor (EGF) to improve survival in Sasai II media. On day 18 the organoids were transferred to an in-incubator orbital shaker (100 rpm) where they remained for the duration of the experiment. After day 35, all cultures were grown in Sasai III media (Sasai II media supplemented with 50 mL of FBS (Hyclone)) supplemented with 10 ng/mL brain-derived neurotrophic factor (BDNF) and 10 ng/mL of neurotrophin-3 (NT-3) (Kindberg et al. 2014).

1.4.2 Immunofluorescence Microscopy

Organoids were collected and fixed in 4% Paraformaldehyde (PFA) (ThermoFisher), washed 3x with PBS, then incubated in 500μL of 30% Sucrose for 2-3 days at 4°C until they began to float. They were then embedded in square cryomolds with Tissue-Tek O.C.T. Compound (Sakura) and placed at -80°C for storage. They were then sectioned to 18 μm using a cryostat (Leica Biosystems) directly onto glass slides. After allowing samples to come to room temperature, 3 washes of 35 minutes in 1X PBS were performed. The sections were then incubated in a blocking solution of 10% BSA for 2 hours. The sections were then incubated in primary antibodies and blocking solution at 1:1000 dilution overnight at 4°C (Supp. Table 4). They were then washed 3 times for 30 minutes and incubated in secondary antibodies for 2 hours at room temperature. They were then washed 3 times for 30 minutes in PBS and sealed with ProLong™ Gold Antifade Mountant with DAPI (ThermoFisher Scientific # P36935). Microscopy was performed using a Zeiss Axio Imager and Zen Software suite, with processing of the images performed using Fiji (Schindelin et al. 2012).

1.4.3 RNA Isolation

Isolation of total RNA from cerebral cortical organoids and stem cells was performed using Direct-Zol RNA MiniPrep Kit (Zymo Research) with TRI Reagent (Molecular Research Center, Inc.). Since a single organoid would yield far less total RNA, approximately 3-8 organoids would be pooled in Trizol depending on organoid size for each sample replicate. The manufacturer's recommended volume of TRI Reagent was added to each sample ($\sim$1 mL). For RNA purification of stem cells, Trizol was directly added to cell culture plates on ice, scraped, and homogenized via pipetting. Organoids in Trizol were broken down via pipetting inside a 1mL Eppendorf tube on ice via a syringe. All total RNA was processed using a MirVana miRNA Isolation Kit (Life Technologies), according to the manufacturer's instructions, to select for RNA <200 nt. This was followed by an RNA Clean and Concentrate-25 (Zymo Research).

Samples were divided into plus-AlkB experimental treatment and minus-AlkB control as previously described in ARM-seq methods (Cozen et al. 2015). AlkB sample treatment was used to effectively increase RT processivity of hyper-modified tRNA/tDRs while also removing sequencing bias favoring hypo-modified tRNA/tDRs (Figure S1.1A). For example, the proportion of ARM-seq tRNA reads more than doubles (8.9% to 21.8%) when comparing untreated versus AlkB-treated samples. This was followed by a phenol-chloroform cleanup treatment. RNA samples were then used for library preparations.

1.4.4 ARM-seq library preparation

ARM-seq libraries were constructed as described previously (Cozen et al. 2015) utilizing the NEBNext Multiplex Small RNA Library Prep Set (New England Biolabs). Treated RNA (Minus- or Plus-AlkB treatment; 100ng total small RNA) was used as input into the library preparation, and ¼ reaction volume of all reagents were used with the NEBNext kit. PCR-amplified libraries were purified using a phenol-chloroform extraction cleanup and then size-selected (140-250 nts) on a 6% non-denaturing TBE-acrylamide gel to remove unwanted primer dimer products. Libraries were eluted from sliced gel pieces using Gel Elution Buffer (New England Biolabs) and precipitated using 0.3 M NaOAc, 80% Ethanol, and 1 μL of Linear Acrylamide (supplied in NEBNext Kit) at final concentration. Samples were left in -80°C freezer overnight to precipitate. Precipitated libraries were then pelleted, washed twice in 80% Ethanol, and resuspended in pure H2O. Libraries were then quantified using NanoDrop and Agilent DNA High Sensitivity kit.

1.4.5 RNA sequencing and differential expression analysis

Libraries prepared for ARM-seq were sequenced using Illumina MiSeq. 75-nt pair-ended reads were produced as FASTQ files that were analyzed using tRNA Analysis of eXpression (tRAX) (Holmes et al. 2022). Sequence adapters were trimmed, and pair-ended reads were merged using the tool trimadapters.py in tRAX. The reference database for tRAX was built with high-confidence tRNA predictions retrieved from the Genome tRNA Database (Chan and Todd M. Lowe 2016) and the sequences of human genome assembly GRCh38. Other gene annotations were obtained from Ensembl release 102. Biological replicates of organoids at each time point were grouped as sample replicates for tRAX inputs and different time points were marked as pairs for differential expression comparison for each sequencing type. For ARM-seq samples, default options of tRAX were used. To check for biases in organoid time points, total normalized reads (Figure S1.1A,B) were compared, showing no notable differences in abundance between time points. Additionally, we found the relative levels of tRNAs were consistent between sample groups, with no significant difference in total reads found between samples within AlkB treatment groups (Figure S1.6A,B). AlkB treatment significantly increases tRNA read counts relative to overall RNA reads (Figure S1.1B). ARM-seq also preferentially picks up more tRNA fragments and their read counts relative to full-length tRNAs (Figure S1.1C). In order to standardize tRNA coverage profiles, each series of reads was aligned to conserved tRNA positions (Sprinzl et al. 1996), dropping gap and extension positions. All visualizations were created using tRNAgraph (https://github.com/alba1735/tRNAgraph) and custom Python scripts.

1.4.6 tRNA Classification Analysis

Sequencing output data from tRAX was used for tDR classification. Deseq2 normalized read counts were combined into an AnnData object and were arranged by coverage-associated data across human GRCh38 tRNAs (Virshup et al. 2021; Love, Huber, and Anders 2014). These were comprised of unique coverage (normalized read count with reads that are uniquely mapped to feature at the specific position), 3’ and 5’ read ends (with consideration to the tRNA structure), deletions (number of reads that have a gap at the specific position), and mismatched bases per position (normalized read count that does not match the reference base at specific position). Unique reads were chosen over total reads as this provided more specific coverage changes in the clusters and better tRNA uniformity therein. Thus, each cluster portrays the coverage makeup of component tDRs and gives a fragment profile in addition to other underlying sequence features. The data was further preprocessed by removing read counts less than 20, regressing out the number of reads, and scaling and centering the data. Dimensionality reduction was performed using Uniform Manifold Approximation and Projection (UMAP) and clustering using hdbscan (McInnes, Healy, and Astels 2017; McInnes, Healy, and Melville 2018). Classifying tRNAs into neural, neutral, and stem categories was based on the $\log_2$ fold-changes between day 0 and day 70 time points split into the top and bottom 20th percentiles of all tRNAs, after removing reads with low coverage and p-value >0.05. $\log_2$ fold-change was normalized between -1 and 1 with each tRNA transcript and assigned a “neural stem score” based on its relative expression between the differentiated and undifferentiated timepoints. To determine if a cluster was neural or stem specific, the mean of the “neural stem scores” was taken for each cluster, and the top and bottom 25 percentiles correlated to neural (B) and stem (A) clusters, respectively. Clusters from neither group were labeled as neutral (C) and unannotated (U) results from hdbscan, their own group. Sequence logos were generated using Logomaker (Tareen and Kinney 2020) by using the mean normalized read counts per position and converting them into information scores (T. D. Schneider and Stephens 1990).

Chapter 2

Telomere-to-telomere Ape Assemblies Reveal Novel tRNAs and Expanded Evolutionary Relationships in The Hominid Lineage

2.1 Background

tRNAs are ubiquitous RNA molecules essential for all life forms, often only viewed in the context of their roles in translation. Despite this, recent tRNA research has focused on roles outside of translation, such as in mammalian gene regulatory function (Avcilar-Kucukgoze and Kashina 2020; G. Li et al. 2022) and disease (Chujo and Tomizawa 2021; Morisaki et al. 2021; Qin et al. 2020; Nie et al. 2019; Sun et al. 2018). Despite the importance of understanding these regulatory mechanisms, most eukaryote tRNA research focuses on the yeast Saccharomyces cerevisiae, an organism with only 55 unique mature tRNA sequences (encoded by approximately 275 gene loci) (Phizicky and Hopper 2010). In contrast to single-celled organisms, most mammalian genomes contain a surprisingly large number of unique sequences ($\sim$400-600) due to an expansion of isodecoder diversity (tRNAs with the same anticodon but different sequences). For instance, after discounting likely pseudogenes, the human GRCh38/hg38 assembly has 252 unique tRNA gene sequences, five times as many as yeast (Chan and Todd M. Lowe 2016). Despite the wide variety of tRNAs found in humans, our understanding of their relationship with other ape primates remains incomplete and is currently limited to *Pongo abelii* and *Pan troglodytes*.

To identify the tRNA genes in a new genome assembly, tRNAs are found using tRNAscan-SE and curated in the tRNA database gtRNAdb using a standardized nomenclature (Chan, Lin, et al. 2019; Chan and Todd M. Lowe 2016). This standard nomenclature makes tRNAs easier to understand and discuss in human-readable language, unlike NCBI's RefSeq automated naming systems, by giving information about amino acid (isotype), anticodon (isoacceptor), unique sequence (isodecoder), and genomic copy number. Additionally, this system divides tRNAs into what is known as the high-confidence set of tRNAs predicted to actively be transcribed as tRNAs and the total set, which includes likely pseudogenes and other tRNA-like genes. This is done by attributing scores to tRNA genes based on their match to the archetypical tRNA for a domain of life, their predicted secondary structure, and their match against all possible isotypes, then filtering based on these scores with specific cutoffs. After filtering, these tRNAs in the high confidence set are predicted to be the most likely involved in ribosomal translation, but additional tools such as tRNA Activity Predictor (tRAP) can further refine this list depending on the circumstances (Thornlow et al. 2019). Additionally, while the high-confidence set is a good guideline for translational activity, evidence for alternative regulatory roles means that the tRNAs found in the total set are also important. Furthermore, tRNAs serve roles in genomic architecture and can mark regions of relative instability and evolution as they often move via the translocation of tRNA clusters (Bermudez-Santana et al. 2010; Guimarães et al. 2021; Parisien, X. Wang, and Pan 2013).

In older human assemblies GRCh37/hg19, 416 tRNA genomic loci were found in the high confidence set, which was later increased to 429 with GRCh38/hg38 (Chan, Lin, et al. 2019). This number of tRNAs was relatively stable, and most of the changes were attributed to sequencing errors and a better reference assembly. Additionally, other tRNA-like predictions, such as possible pseudogenes, increased from 177 to 187, with many falling into a neural expansion region (NER) comprised of the 1q21 (1q21.1/2) and 1p11 regions of chromosome 1, a notoriously dynamic and highly repetitive human-primate expansion region caused by many recent duplication-deletion events derived from 1p11 onto 1q21 (Fiddes, Lodewijk, et al. 2018). Despite having a high copy number, many of these tRNA pseudogenes contain short interspersed nuclear elements (SINEs) and retain promoter elements, signifying evolutionary conservation for translational alternative roles (Palazzo and Lee 2015; Kutter et al. 2011). Some tRNAs appear to have tissue-specific transcription or are not transcribed despite having high sequence conservation (Kundaje et al. 2015). Recent tools have been made to predict the activity of tRNAs (such as tRAP). However, they rely on epigenetic information from large-scale histone Chromatin Immunoprecipitation data, often missing in the repetitive regions where many tRNA genes are located (Thornlow et al. 2019).

tRNAs are widely associated with their critical roles in translation, and so typically, tRNA gene copy number has been matched to codon use (Duret 2000; Novoa et al. 2012; Pechmann and Frydman 2013). Thus, the link between translation efficiency and codon mismatching is highly correlated with the tRNA gene pool and can be affected by various mechanisms, such as stress response and cell-cycle growth

(Begley et al. 2007; Berg and Kurland 1997; Letzring, Dean, and Grayhack 2010; Patil et al. 2012; Plotkin and Kudla 2011; Xu et al. 2013). Due to these constraints, high tRNA gene variability is not expected, and what variability is observed is often in tRNA-derived SINEs, not in essential “core” tRNAs. In the cases where high copy number variation (CNV) of tRNA genes have been studied, the analysis was constrained to closely related species with smaller tRNA gene sets relative to Human, showing alternative codon usage or wobble-dependent pairing occurs, which has various effects on translational stability (Higgs and Ran 2008; Iben and Maraia 2012).

A few studies have focused on tRNA gene CNV in Human, observing variable repeats in the 1q23.3 region (more than observed in hg19) and deletion of a tRNA gene on Chr 11 (Iben and Maraia 2014). Another study found this same region of tandem repeats of tRNA genes varying between 9 to 43 repeat units across mammals but focused on its function as a genomic boundary element (Darrow and Chadwick 2014). Due to the limited quality of hg19/hg38 genome assemblies in the vicinity of tandem repeats, however, this region has collapsed to five repeats in the UCSC Genome Browser, and these studies focused on boundary elements rather than specific tRNA sequence variation within this region. While predictions were made about the activity states of tRNAs in these repetitive regions, the lack of granular observation of the individual tRNAs in the tandem repeats and incomplete chromatin data leaves an open question over what is and isn’t transcriptionally active.

Although the updated hg38 human reference genomes enhanced our comprehension of certain specific genomic regions, it did not encompass all tRNA genomic information because numerous tRNAs were still located on unplaced scaffolds and hard-to-sequence centromeric regions (Chaisson et al. 2015; V. A. Schneider et al. 2017). This was recently solved with the release of CHM13v2.0/hs1 by the Telomere-to-Telomere (T2T) Consortium, which addressed the remaining unresolved regions of the genome representing a complete 3.055 billion-base pair gapless assembly (Nurk et al. 2022). Notably, the completed regions include segmental duplication regions that have expanded across the primate lineage, such as the NER and 1q23.3. The NER is also interesting in the context of T2T assembly as it harbors the NOTCH2-N-terminus-like (N2NL) family (A/B/C/R) and neuroblastoma breaking-point family (NBPF) genes, which are essential in human-primate brain evolution and has been notoriously difficult to assemble because of the many segmental duplications in this region (Fiddes, Pollen, et al. 2019).

The new human T2T genome (hs1) provides a fresh perspective on human tRNA genes, diverging significantly from earlier assemblies (hg19/hg38). Expanding upon the human telomere-to-telomere (T2T) project, recent genome assemblies have employed comprehensive end-to-end sequencing in ape primates, giving a greater evolutionary context (Makova et al. 2023). By leveraging these ape genomes, we can fully reconstruct the intricate tRNA gene landscape across hominid evolution and gain deeper insights into the interplay between tRNA gene conservation and neural expansion. Despite limited attention to tRNA copy number variation and diversity within Eukaryotes, our findings unveil a multifaceted landscape of tRNA gene copy number variation, shedding light on its significance in human-primate neural gene expansion.

2.2 Results

2.2.1 Nearly 100 additional tRNA gene loci in newer Human Genome Assemblies

After running tRNAscan-SE on the new human T2T-CHM13v2.0(hs1) genome assembly, we observed almost 100 additional tRNA gene loci when compared to the two prior assemblies GRCh38(hg38)/ GRCh37(hg19) (Table 2.1). In the high confidence set, this number increases from 429 to 521 (an increase of 92); in the total set, this number increases from 619 to 715 (an increase of 96). We next examined if any tRNA genes were changed in sequence in the new assembly. While most tRNAs maintained the same sequence, a few notable examples changed between assembly versions.

tRNA sequence divergence occurs in 56 syntenic tRNA genes when compared against those located at the same loci in the previous assembly (hg38) (Table S2.1). tRNA-Ser-GCT-4-4 (previously tRNA-Ser-GCT-6-1) is one such example where two base changes in the stem of the tRNA at positions 2

Genome Assembly	High Confidence	Total Set
hg19	416	596
hg38	429	619
hs1	521	715

Table 2.1: Human tRNA Counts

Human hg19, hg38, and hs1 tRNA counts.

and 3 increase the tRNAs tRNAscan-SE score (based on the Eukaryotic domain covariance model) from 71.6 to 88.4 (Figure 2.1A). A score above 50 is typically associated with the high-confidence set with higher relative scores conforming more closely to an archetypical translationally active tRNA. This change is associated with increased stability of the tRNA due to an A to C at position 3, allowing binding with G on the opposing side of the acceptor stem. Furthermore, this tRNA switches from being a unique genomic loci to becoming identical to several other Ser-GCT genes. tRNA-Ala-AGC-25-1 and tRNA-Ser-AGA-2-3 are other unique examples in that they both contain changes in position 35 of their anticodons changing from A to G and G to A, respectively. In the case of Ala-AGC, this causes an identity switch from Val-AAC on the previous assembly (hg38) while maintaining a high enough set of tRNAscan-SE scores to remain a high confidence set tRNA. tRNA-Ser-AGA-2-3 is particularly interesting because, in addition to changing its anticodon from Ser-AGA to Phe-AAA, its relative tRNA structure maintains identity elements of Ser, causing a large isotype prediction disagreement (IPD) of -134.60, the difference between the highest isotype score and the isotype score with the detected anticodon. This indicates a disagreement between the canonical predicted sequence of the hs1 tRNA gene loci being a Phe-AAA. This suggests it structurally looks more similar to a Ser-AGA despite the reflected anticodon (Phe-AAA).

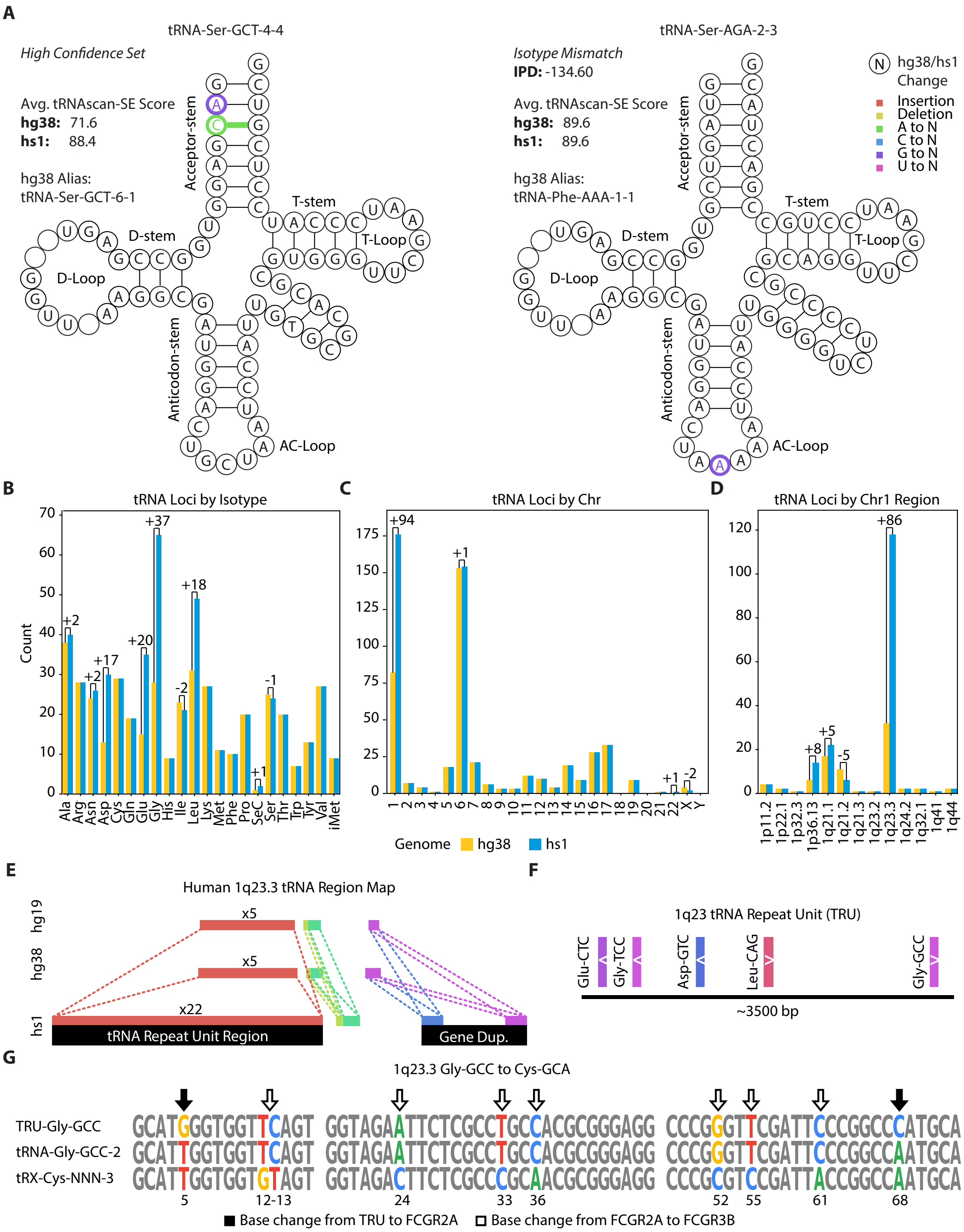

A
tRNA-Ser-GCT-4-4
High Confidence Set
Avg. tRNAscan-SE Score
hg38: 71.6
hs1: 88.4
hg38 Alias:
tRNA-Ser-GCT-6-1
Acceptor-stem
D-stem
D-Loop
T-stem
T-Loop
Anticodon-stem
AC-Loop
tRNA-Ser-AGA-2-3
Isotype Mismatch
IPD: -134.60
Avg. tRNAscan-SE Score
hg38: 89.6
hs1: 89.6
hg38 Alias:
tRNA-Phe-AAA-1-1
N hg38/hs1 Change
Insertion
Deletion
A to N
C to N
G to N
U to N
B
tRNA Loci by Isotype
Count
+2
+2
+17
+20
+37
-2
+18
+1
-1
Ala Arg Asn Asp Cys Gln Glu Gly His Ile Leu Lys Met Phe Pro SeC Ser Thr Trp Tyr Val iMet
C
tRNA Loci by Chr
+94
+1
+1
-2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
Genome hg38 hs1
D
tRNA Loci by Chr1 Region
+86
+8
+5
-5
1p11.2 1p22.1 1p32.3 1p36.13 1q21.1 1q21.2 1q21.3 1q23.2 1q23.3 1q24.2 1q32.1 1q41 1q44
E
Human 1q23.3 tRNA Region Map
hg19
hg38
hs1
x5
x5
x22
tRNA Repeat Unit Region
Gene Dup.
F
1q23 tRNA Repeat Unit (TRU)
Glu-CTC
Gly-TCC
Asp-GTC
Leu-CAG
Gly-GCC
~3500 bp
G
1q23.3 Gly-GCC to Cys-GCA
TRU-Gly-GCC GCATGGGTGGTTCAGT GGTAGAATTCTCGCCTGCCACGCGGGAGG CCCGGGTTCGATTCCCGGCCCATGCA
tRNA-Gly-GCC-2 GCATTGGTGGTTCAGT GGTAGAATTCTCGCCTGCCACGCGGGAGG CCCGGGTTCGATTCCCGGCCAATGCA
tRX-Cys-NNN-3 GCATTGGTGGTGTAGT GGTAGACTTCTCGCCCGCAACGCGGGAGG CCCGCGTCCGATTACCGGCCAATGCA
5 12-13 24 33 36 52 55 61 68
■ Base change from TRU to FCGR2A □ Base change from FCGR2A to FCGR3B

Figure 2.1

Figure 2.1: Human GRCh38/hg38 and CHM13v2.0/hs1 tRNA Changes

(A) *Schematic of tRNAs mapped to the same genomic loci but have changes in their sequence between hg38 and hs1 assemblies of the human genome, with new base pair change/s shown in orange.* ***(B-D)*** *Bar plots showing the relative counts of tRNAs found across* ***(B)*** *isotype,* ***(C)*** *chromosome, and* ***(D)*** *chromosome 1 cytoband locations in hg38 (yellow) and hs1 (blue).* ***(E)*** *Schematic representation of human 1q23.3 between hs1 and hg38/hg19 with different colored bars representing tRNA regions defined as tRNA genomic loci bound by protein-coding genes, with tRNA repeat unit region and gene duplications highlighted in black.* ***(F)*** *Schematic representation of the five tRNA genomic loci found in each 1q23.3 tRNA repeat cluster, directionality shown in white and spaced to scale.* ***(G)*** *Sequence alignment of TRU Gly-GCC consensus, tRNA-Gly-GCC-2-1 (proximal to FCGR2A), and tRX-Cys-NNN-3-1/2 (proximal to FCGR3B), with solid and hollow arrows to distinguish positions of changes between the tRNA gene sequences.*

The relative concentration of DNA variants between hs1 and hg38 shows that the mutation rate could be higher in the anticodon stem and stems relative to D/T-arms of the tRNA, possibly due to higher conservation of the A and B boxes (tRNA promoter region) (Figure S2.1A). DNA variants are more common in Ala (9), Lys (6), and Cys (5) isotypes. Secondary structure scores (predicting the tRNA's ability to form internal pairing structure) show that about half of the 56 tRNAs with alternate bases have score increases and half have score decreases with the most extreme change ∼9-10 in either direction. This typically matches whether a base change in a stem/loop will facilitate better/worse binding and create a more/less canonical tRNA structure. In addition to looking at tRNA variants between the two assemblies, we next wanted to determine what was causing the large increase in tRNA genes.

Asp, Glu, Gly, and Leu isotypes had the highest increase in gene copy number with gains of 17, 20, 37, and 18 loci, respectively (Figure 2.1B-D). When broken down by isoacceptors, Gly seemed to split evenly between GCC (17) and TCC (17), with CCC (3) accounting for only a minor portion of the pool's increase (Figure S2.1B). Asp only had changes in its major isoacceptor (GTC), and Glu had the majority of changes in CTC (17) rather than TTC (3). Interestingly, the tRNA isotype Leu has the most diverse set of possible anticodons in Humans. Yet, nearly all loci increases were found in CAG (17), with a single other change in CAA. Since a specific set of tRNAs was increasing rather than a random distribution, which would be expected if only caused by SNPs and population variance, we wanted to determine where these increases occurred.

Looking at tRNA loci on a chromosome basis, we found that chromosome 1 had 94 additional tRNA loci, accounting for 98% of the total additional tRNAs in hs1 (Figure 2.1C). On hg38, chromosome 6 contained most tRNA genes followed by chromosome 1; however, this no longer seems to be the case with the marked increase in chromosome 1 tRNA genes. Interestingly, the highest number of variants was found on chromosome 6 (23 out of 56) despite only having one additional tRNA in hs1 (Figure S2.1A). We then determined a large increase in Chromosome 1 and localized to the 1q21 and 1q23.3 regions, with 1q23.3 accounting for the greatest tRNA gene increase by far (Figure 2.1D).

The high resolution of the telomere-to-telomere assembly helped to resolve sequences and scaffolds that were unclear within the 1q21 region in previous assemblies. tRNA genes often fall into proximal clusters on the genome and were defined as tRNA clusters based on continuous regions of tRNA genes bound by protein-coding genes. In 1q21, 16 distinct tRNA regions were found in hg38 and hs1 assemblies with two different regions that were inverted in the new assembly relative to the old one (Figure S2.2A). Since many tRNA genes in this region are found adjacent to either N2NL or NBPF genes (which share location and sequence), this set of rearrangements shows that the prior understanding of tRNAs in the region wasn't fully understood. This is particularly important as over 5% of high-confidence tRNA genes fall into 1q21 (28/521), but nearly 20% of tRNA pseudogenes and others filtered from the total set also reside there (37/194). Altogether, just short of 10% of all tRNA genes and pseudogenes fall within this region in hs1 (as well as in hg38). Many of these tRNA genes were predicted to be inactive on previous assemblies, hinting at alternative non-translation functions. Interestingly, many protein-coding genes in these regions have roles in neural developmental and neurogenetic disease (Brunetti-Pierri et al. 2008; Mefford et al. 2008).

tRNA gene loci in 1p36.13 also seem to more than double (6 to 14) relative to hg38. Looking at the region closely, a series of Asn-GTT, Glu-TTC, Gly-CCC, and Val-CAC are copied and rearranged within the region, forming semi-repeating units of 2-5 tRNA genes. This series of tRNA gene repeats flank NBPF1 on one end of the tRNA region and CROCC on the other, genes associated with neuroblastoma and primary autosomal recessive microcephaly (Dang and Schiebel 2022; L. Li et al. 2022). Further, 1p36 deletion syndrome, characterized by developmental delay/intellectual disability and other neurodevelopmental disorders, is one of the most common human disorders resulting from terminal autosomal deletion (Radio et al. 2021). Since tRNA copy number seems variable between hs1/hg38, we wanted to look at other regions of tRNA CNV.

In contrast to 1q21, the 1q23.3 region is more clearly defined and marked by four distinct tRNA regions in hg19/hg38, with an extra region appearing in hs1. One tRNA region in 1q23.3 has a series of 5 tRNAs Glu-CTC (-), Gly-TCC(-), Asp-GTC (-), Leu-CAG (+), and Gly-GCC (+) defined as a tRNA repeat unit (TRU) in variable copy across the genomes (Figure 2.1E,F). In prior assemblies, this repeat unit appeared 5x times, increasing to x22 in hs1 with the same series of 5 tRNAs in all units except on one outside TRU, which had the first Glu-CTC absent. Segmental duplications (via SEDEF) are predicted at over 99% similarity for the entire tRNA region, and the expansion of repeat copy number is likely in part due to the use of ultralong reads during assembly, which were most likely collapsed on older assemblies (Numanagic et al. 2018; Vollger et al. 2022).

It has been observed that an orthologous relationship between FCGR2A and FCGR3 genes is adjacent to the 1q23.3 TRU region throughout the primate lineage (Lejeune, Brachet, and Watier 2019). When looking at the gene evolution of FCGR2A to FCGR3B closely, we discovered that tRNA genes also got carried along with an Asp-GTC, Gly-GCC, Leu-CAG, Gly-TTC, and Asn-GTT (similar in sequences to the TRU) being adjacent both FCGR genes. The TRU Gly-GCC has an identical sequence across all 22 repeats in hs1, with only two base changes compared to the FCGR2A version. Both are in the high-confidence tRNA set (Figure 2.1G). In contrast, when comparing the FCGR2A Gly-GCC and FCGR3B Cys-GCA tRNA genes, the tRNAscan-SE model prediction scores are drastically lower, indicating complete degradation of these tRNA into pseudogenes. This suggests that much more evolutionary constraint is placed on the tRNA genes in the 1q23.3 TRU region than in the repeats associated with these FCGR genes.

2.2.2 Increased tRNA Diversity is Found in Neural Expansion Regions of Primate

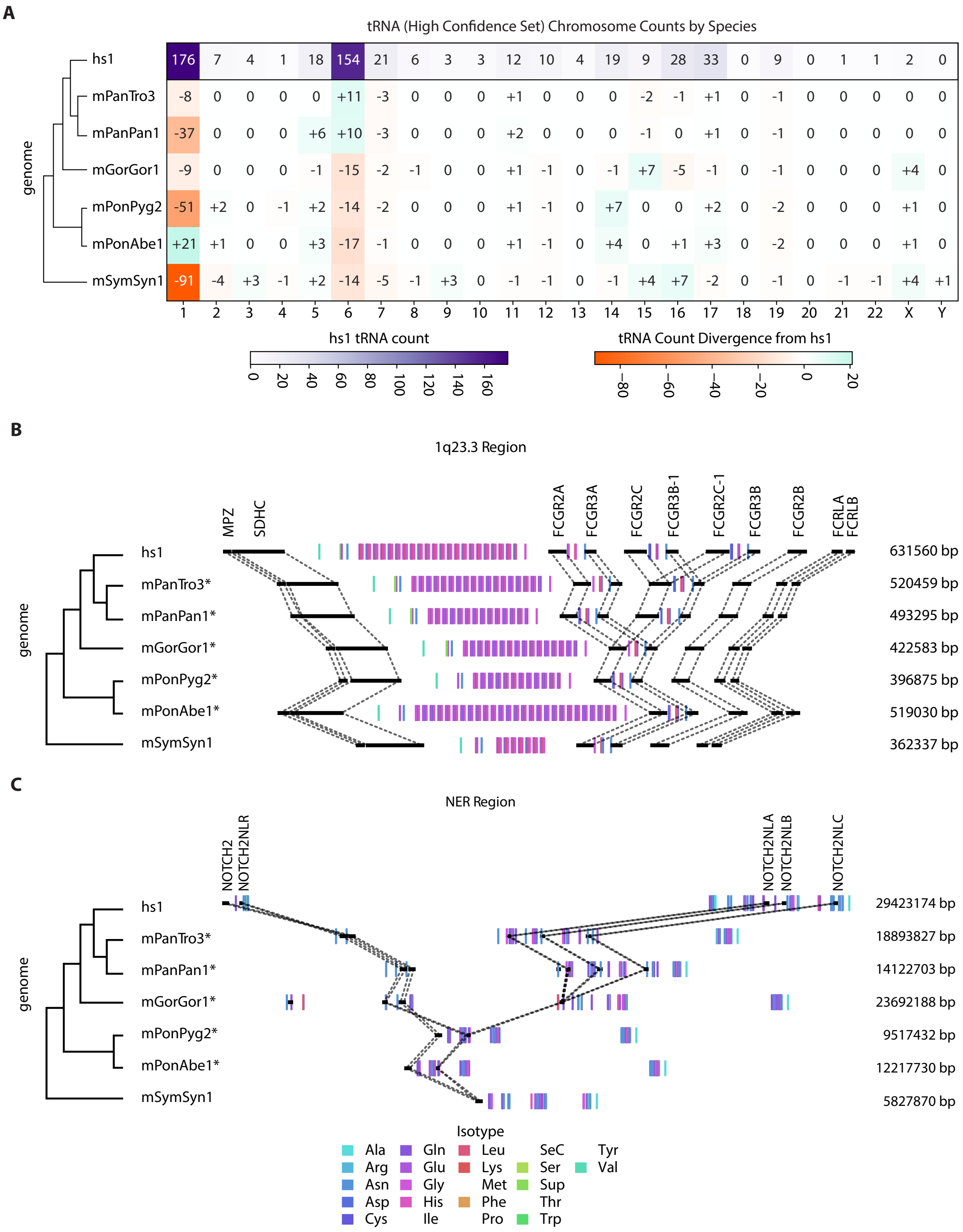
We examined the six new T2T ape assemblies aligned to hs1 in a cactus graph (HAL) to compare tRNA loci across these closely related species. From proximal to distal to human relation, the included genomes are *Pan troglodytes* (Chimpanzee) mPanTro3, *Pan paniscus* (Bonobo) mPanPan1, *Gorilla gorilla* (Gorilla) mGorGor1, *Pongo pygmaeus* (Bornean orangutan) mPonPyg2, *Pongo abelii* (Sumatran orangutan) mPonAbe1, and *Symphalangus syndactylus* (Siamang gibbon) mSymSyn1. In general, the number of tRNA genes in both the high confidence and total tRNA gene sets decreased with evolutionary distance from hs1, with the expectation being mPonAbe1, which has more tRNA genes than all other primates (Table 2.2). Only prior assemblies for Chimpanzee (Pan_tro 3.0), Gorilla (gorGor4), and Sumatran Orangutan (Susie_PABv2) have prior tRNA gene sets computed on gtRNAdb. In the case of prior assemblies, Sumatran orangutan also has a higher copy number of tRNA genes relative to gorillas but less than humans or chimpanzees. Prior Orangutan, Gorilla, and Chimpanzee genomes have many unresolved scaffolds (5299, 40729, and 45510, respectively) across many contigs, suggesting that tRNA gene copy number variation could be an artifact of poor genome assembly or could indicate that tRNA copy number is dynamic.

Species	Current Release	High Con.	Total	Prior Release	High Con.	Total
Human	hs1	521	715	hg38	429	619
Chimpanzee	mPanTro3	519	689	Pan_tro 3.0	430	622
Bonobo	mPanPan1	498	670	NA		
Gorilla	mGorGor1	496	679	gorGor4	390	571
Bornean orangutan	mPonPyg2	465	650	NA		
Sumatran orangutan	mPonAbe1	535	721	Susie_PABv2	405	596
Siamang gibbon	mSymSyn1	421	656	NA		

Table 2.2: Primate tRNA Counts

Primate tRNA counts compared to the prior release.

The primate tRNA genes were aligned to the human tRNA genes by chromosomes and found to follow the same trend of having the most tRNAs on Chromosome 1 and 6 (Figure 2.2A). The two most considerable discrepancies from hs1 seemed to be with tRNA counts on the mPonAbe1 and mSymSyn1 genomes, in particular with mPonAbe1 having more tRNA genes than Human on chr1 in contrast to all other primates and mSymSyn1 having more tRNA genes on chr15 and 16. When looking at the total tRNA set, this gap is greatly reduced in mPonAbe1 but grows even larger in mSymSyn; this suggests that at least on hs1 chr1 has far more tRNA pseudogenes than mPonAbe1 (Figure S2.3). Besides mSymSyn1, primates had precise mapping of the syntenic chromosomes, with the exception of two chromosome tRNA sets, both of which mapped to human chr2 and a known fusion event (Poszewiecka et al. 2022). mSymSyn, however, had a much more complicated set of tRNA mappings when compared to the other species with multiple mSymSyn1 chromosome tRNAs mapping to chr1, chr3, chr6, and chr16 and none mapping to chr21 and chr22 as well as tRNAs on various mSymSyn1 chromosomes mapping to multiple human chromosomes. This marks a drastic change in tRNA mappings between Gibbon and Orangutan that are then preserved in the ape lineage to Human. Finally, the amount of tRNAs relative to Human goes from negative to positive on chromosome 6 between the Gorilla and Chimpanzee/Bonobo branches, suggesting a tRNA gene translocation event happens between those species.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
hs1	176	7	4	1	18	154	21	6	3	3	12	10	4	19	9	28	33	0	9	0	1	1	2	0
mPanTro3	-8	0	0	0	0	+11	-3	0	0	0	+1	0	0	0	-2	-1	+1	0	-1	0	0	0	0	0
mPanPan1	-37	0	0	0	+6	+10	-3	0	0	0	+2	0	0	0	-1	0	+1	0	-1	0	0	0	0	0
mGorGor1	-9	0	0	0	-1	-15	-2	-1	0	0	+1	-1	0	-1	+7	-5	-1	0	-1	0	0	0	+4	0
mPonPyg2	-51	+2	0	-1	+2	-14	-2	0	0	0	+1	-1	0	+7	0	0	+2	0	-2	0	0	0	+1	0
mPonAbe1	+21	+1	0	0	+3	-17	-1	0	0	0	+1	-1	0	+4	0	+1	+3	0	-2	0	0	0	+1	0
mSymSyn1	-91	-4	+3	-1	+2	-14	-5	-1	+3	0	-1	-1	0	-1	+4	+7	-2	0	-1	0	-1	-1	+4	+1

Figure 2.2

Figure 2.2: Primate tRNA Genomic Loci Comparisons

(A) *Heatmap of high confidence tRNA counts found in hs1 and aligned primate chromosomes. The top row in purple shows total tRNA counts in hs1 (human) and the divergence in counts from Human in the primates from green (increase) to orange (decrease) relative to hs1. The phylogenetic tree is located on the left and arranged from most to least similar to humans.* ***(B-C)*** *Scaled schematic representation of the* ***(B)*** *1q23.3 and* ***(C)*** *Neural Expansion Region aligned across all T2T primates with tRNA genes colored by isotype. The signature genes of each region are shown in black, and the asterisks next to genomes denotes that the region is inverted relative to hs1.*

Across the apes, we noticed that a high copy number of tRNA genes fell into 1q23.3 with a region of variable TRUs bound by the SDHC gene on one side and the FCGR2A gene on the other (Figure 2.2B). Humans had the second-highest copy number of TRU, with mPonAbe1 containing an additional five repeats (27 total). In contrast, the other orangutan (mPonPyg2) only had 12 repeats within the region, only five more copies than the lowest (mSymSyn1 at 7 TRU). With the exception of mSymSyn1, all primates had the missing Glu-CTC in their first TRU (Figure 2.1F). In addition to the TRUs that expanded, a gene transfer event also increased the copy number of tRNA genes, as was determined by looking at tRNA genes adjacent to other genes in the region. The first gene duplication even appears to happen between the gorilla and chimpanzee/bonobo branch points, with FCGR2A and FCGR3B splitting into FCGR2C and FCGR3A, respectively, copying a series of 5-6 tRNA genes with them. One more duplication event occurs between the chimpanzee/bonobo and human lineage, with FCGR2C and FCGR3B becoming FCGR2C-1 and FCGR3B-1. Compared to the prior Human assembly hg38, these additional copies of FCGR2C/FCGR3B and the carried tRNA genes are absent.

We wanted to assess the changes in tRNAs in the neural expansion region (NER) of 1p11 and 1q21 because of their association with neural genes, changes in copy number, and conservation. Since this region expanded so much over the ape lineage, it’s hard to pinpoint whether this is actually two distinct regions in other primates, and it is still being determined if they are split across centromere as in humans (Figure 2.2C). From mSymSyn1 to hs1, the tRNAs in this region tend to move in parallel with N2NL and NBPF expansions as they fall upstream and downstream of these genes, similar to the FCGR adjacent tRNA expansion in 1q23. The region approximately doubles in size when going from Gibbon to Orangutan, splitting the tRNAs into two larger groups, one containing an extra copy of an N2NL-pseudogene (precursor to N2NLR on the 1p11 region of human) with associated tRNA gene cluster. In the jump to Gorilla, this region approximately doubles once more, splitting the N2NLR-pseudogene into 2 copies. The evolutionary split between Gorilla and Chimpanzee/Bonobo matches human with three N2NL-pseudogenes (precursors to A/B/C) carried over into Human. Over this expansion, the number of tRNA genes increases from 27 (mSymSyn1) to 32 (hs1) in the NER, but the amount of tRNA-pseudogenes drastically expands from 29 (mSymSyn1) to 44 (hs1) (Table S2.2). Since gene copy number and tRNA isotype variance differ drastically in these two regions, we wanted to examine the relationship between sequence conservation and TRUs more closely.

2.2.3 tRNA Codon Usage Bias Changes Based on Proximal Genes

We suspected that sequence conservation varies between the NER and 1q23.3 regions and found over 95% identity matching within the 1q23.3 TRUs, with the highest variance found in TRUs on the region's edges (Figure 2.3A). We determined this by looking at the average substitution rate across the TRUs compared to the consensus (most common base per position) and found the Asp-GTC located in the center had the highest sequence conservation, followed by the Gly-GCC and Leu-CAG. The Glu-CTC and Gly-TCC are very close to one another ($\sim$250bp) within the TRU and seemed to have a higher rate of substitutions overall, with hs1, mPanTro3, and mPonAbe1 having an average rate of 1 substitution per sequence between 19-42%, seeming to correlate with the number of TRU copies per genome. The tRNAscan-SE score of tRNAs in the TRU region also remained consistent, not degrading into tRNA-pseudogenes (Figure S2.4A). The Leu-CAG to Gly-GCC gap is the largest in the TRU ($\sim$1700bp), with the last TRU having an extension of $\sim$9250bp (In Human) to the Gly-GCC (closest to FCGR2A), since this was an entirely different Gly-GCC relative to the Gly-GCC in the TRU it was excluded from the substitution calculation.

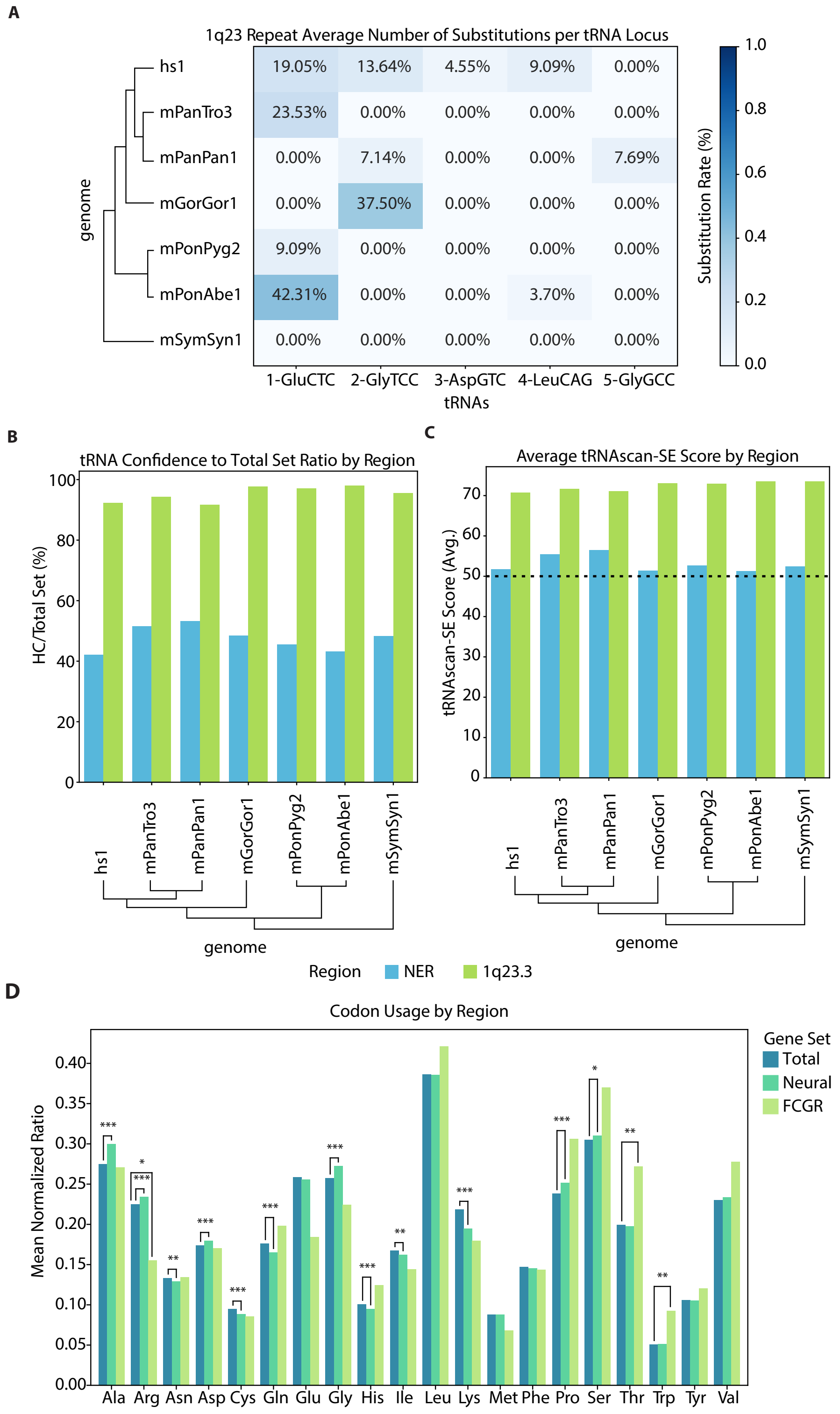

A
1q23 Repeat Average Number of Substitutions per tRNA Locus
genome
hs1
mPanTro3
mPanPan1
mGorGor1
mPonPyg2
mPonAbe1
mSymSyn1
19.05% 13.64% 4.55% 9.09% 0.00%
23.53% 0.00% 0.00% 0.00% 0.00%
0.00% 7.14% 0.00% 0.00% 7.69%
0.00% 37.50% 0.00% 0.00% 0.00%
9.09% 0.00% 0.00% 0.00% 0.00%
42.31% 0.00% 0.00% 3.70% 0.00%
0.00% 0.00% 0.00% 0.00% 0.00%
1-GluCTC
2-GlyTCC
3-AspGTC
4-LeuCAG
5-GlyGCC
tRNAs
Substitution Rate (%)
1.0
0.8
0.6
0.4
0.2
0.0
B
tRNA Confidence to Total Set Ratio by Region
HC/Total Set (%)
100
80
60
40
20
0
hs1
mPanTro3
mPanPan1
mGorGor1
mPonPyg2
mPonAbe1
mSymSyn1
genome
C
Average tRNAscan-SE Score by Region
tRNAscan-SE Score (Avg.)
70
60
50
40
30
20
10
0
hs1
mPanTro3
mPanPan1
mGorGor1
mPonPyg2
mPonAbe1
mSymSyn1
genome
Region
NER
1q23.3
D
Codon Usage by Region
Mean Normalized Ratio
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
***
*
***
**
***
***
***
***
***
**
***
***
*
**
**
Ala
Arg
Asn
Asp
Cys
Gln
Glu
Gly
His
Ile
Leu
Lys
Met
Phe
Pro
Ser
Thr
Trp
Tyr
Val
Gene Set
Total
Neural
FCGR

Figure 2.3

Figure 2.3: Primate tRNA Genomic Loci Comparisons

(A) *Heatmap of the five tRNA genomic loci found in each repeat cluster with the average substitution rate against the consensus per tRNA sequence across all T2T primates.* ***(B-C)*** *Barplot of the* ***(B)*** *ratio of tRNAs in the high confidence set vs the total set and* ***(C)*** *the average tRNAscan-SE score in the neural expansion and the 1q23.3 regions. The line at a score of 50 distinguishes the cutoff for secondary filtering of the tRNA high confidence set.* ***(D)*** *Codon usage by neural genes and FCGR genes (1q23.3) compared against the total NCBIRefSeq annotated gene set. Significance by independent t-test shown as ns (p <= 1), * (1e-02 < p <= 5e-02), ** (1e-03 < p <= 1e-02), *** (p <= 1e-03).*

Matching tRNAs by TRU and gene (as in 1q23.3) was far more difficult in the NER due to large changes and rearrangements from one primate genome to the next. We took tRNA genes by isoacceptor and analyzed them by which tRNA set they fell into (total vs. high confidence) and their tRNAscan-SE score (where less than 50 is considered a likely pseudogene) (Figure 2.3B,C; Table 2.3). When tRNAs have a high score, they are more likely to be predicted as functional tRNA genes used for translation, with a lower score often associated with alternative roles or an artifact of gene duplication. This is consistent with what we observed with the tRNAs in the NER, which appeared in the high confidence set nearly half as often as the total set on average. Additionally, the mean tRNAscan-SE score for tRNA genes in the NER is on the threshold of secondary filtering, indicating that the abundance of tRNA pseudogenes is much higher in the NER relative to 1q23.3. In general, tRNA genes seem far less conserved within the NER, with an example of Leu-TAA and Trp-CCA only appearing in Gorilla as 6 pseudogenes or Arg-CCG fluctuating in tRNAscan-SE score between primate branches before dissipating in hs1. tRNA pseudogenes for Ser-CGA and SeC-TCA entirely vanish between mSymSyn1 and hs1. Asn-GTT and Val-CAC appear to increase the most in humans relative to the other primates, rising from a mean of $\sim$19 to 29 copies (a $\sim$45% increase) for Asn-GTT and from $\sim$2 to 7 for Val-CAC (Figure S2.4B). The Asn isotype expansion falls in line with Asn being heavily favored in the 1q21 region, as $\sim$58% (26/45) of Asn isotype from the total set fall into 1q21 as well, suggesting that Asn copy number was favored across the neural expansion.

Neural Expansion Region (NER)				
Assembly	**High Con.**	**Total**	**High Con.%**	**tRNAscan-SE Score (Avg.)**
hs1	32	76	42.10	51.73
mPanTro3	36	70	51.43	55.45
mPanPan1	34	64	53.13	56.49
mGorGor1	31	64	48.44	51.36
mPonPyg2	25	55	45.45	52.72
mPonAbe1	25	58	43.10	51.32
mSymSyn1	27	56	48.21	52.42
1q23.3 TRU Region				
Assembly	**High Con.**	**Total**	**High Con.%**	**tRNAscan-SE Score (Avg.)**
hs1	118	128	92.19	70.79
mPanTro3	98	104	94.23	71.74
mPanPan1	77	84	91.67	71.15
mGorGor1	86	88	97.73	73.06
mPonPyg2	66	68	97.06	73.00
mPonAbe1	140	143	97.90	73.49
mSymSyn1	42	44	95.45	73.49

Table 2.3: tRNA Conservation in Chromosome 1 Region

Human hs1 tRNA high confidence to total set ratios and average domain scores for the NER and 1q23.3.

When comparing the NER and 1q23.3 regions, the secondary structure score is consistently lower in all primates in the NER (Figure S2.5A). A primary drop in secondary structure scores might indicate less functional translational efficiency. Pol III has roles in regulating the transcription of tRNA genes; however, some Pol III regulates Pol II transcription and chromatin binding, altering local chromatin structures and Pol II transcription rate (Jiang et al. 2022). The degradation in tRNAs between the NER and 1q23.3 regions might have more to do with Pol III as an insulator and the conservation of A/B boxes being more important for tRNA genesis for translation. Lastly, the assumption that tRNA rates are often tied to codon use might indicate that the abundance of Arg, Asn, Gly, Glu, and Val in 1p36.13, 1p11, 1q21, and 1q23.3 in neural active gene regions correlates to higher use of those amino acids in neural specific genes.

We thus wanted to test if codon usage was tied to abundance for the NER and 1q23.3 regions and how well tRNA gene copy number correlates with translational need. In the context of 1q23.3, the FCGR (Fc gamma receptor) and FCRL (Fc receptor-like) genes in the region encode a receptor for the Fc portion of immunoglobulin G (or a similar protein, respectively), which play roles in immune cells, their response, and autoimmune disease susceptibility when dysregulated (X. Li et al. 2009; Lassaunière and Tiemessen 2021). When comparing the codon usage of these Fc genes against the mean codon usage of all coding genes, we found no significant difference in Asp, Glu, Gly, and Leu (the TRU tRNAs) (Figure 2.3D). These immune response genes increase in number over the ape lineage alongside the increase in TRU copy number (with the exception of mPonAbe1), suggesting that transcription of TRU tRNAs is not dependent on copy number alone but may be indirectly regulated based on the need for immune response.

In the context of the NER, we found that the expansion of Asn tRNA genes and pseudogenes corresponds with a significant decrease in Asn codon usage in neural genes. Since these Asn tRNA genes are expanded over the hominids, but usage of these codons is decreased in a neural context, we conclude that they are conserved for a non-translational purpose. Arg, Gly, Glu, and Val are also commonly found in the NER, with a significant increase in codon usage of Arg and Gly tRNAs in neural genes. This suggests that the increase in Arg and Gly is driven by translation needs, whereas the increase and conservation of Asn pseudogenes is driven by a non-translational role. Due to the contrasting roles of these tRNAs, we conclude that the dynamics of tRNA genes and codon usage in the NER are shaped by a delicate interplay between translational requirements and non-translational functions.

2.3 Discussion

The new T2T assembly hs1 has added almost 100 new cytosolic tRNA gene loci to the human genome and has expanded our understanding of tRNA variability in populations. This has led to a greater understanding of tRNA repeat regions that were previously collapsed, as well as variability in neural gene regions containing tRNA genes in chromosome 1 and the equivalent regions in closely related primates. Furthermore, looking at changes to tRNAs on a base level indicates that tRNAs are dynamically changing between individuals and that “one reference” genome will not give a complete story about human tRNA diversity. A greater focus must be placed on the state of tRNA gene copy number, chromosomal location, and conservation of nucleotides across positions.

tRNAs’ extra-translation roles are often neglected, especially in the context of copy number variation and nearby gene regulation. It is still not understood if the regulation of genes adjacent to tRNA genes in 1q23.3 (such as FCGR2A) is dependent on POL III/II regulation pathways and how altering the tRNA genes to alternate ones, removing the regions entirely or altering the POL III promoters will have an effect. Further, the necessity of tRNA copy numbers being closely related to translation efficiency indicates that these regions might dynamically change tRNA gene transcription needed to operate at various stages of neural development.

Segmental duplication events involving tRNA gene loci are shared across the primate lineage, and this has changed how we define “static” and “conserved” tRNA gene loci. Segmental duplication events involving tRNA genes are common in the primate lineage, such as the fusion events of chr6 and chr2. Each time, highly concentrated regions of tRNAs are formed, but the translation activity of these regions is unclear. 1q23.3 has extremely high levels of conservation, whereas the human-primate NER tends to change drastically. Regardless of the surrounding base conservation, the amount and variety of tRNA genes in both regions change. It is unclear why some tRNA genes are preferentially conserved over others.

Tools like tRNA Activity Predictor (tRAP) (Thornlow et al. 2019) estimate similar activity levels as GRCh38/hg38 in hs1; however, highly duplicated/repetitive regions are uncertain. In regions where tRNA gene copy number changes, the necessity of new methods or experiments to determine the state of tRNA activity has never been more important. This can help establish whether tRNAs in these regions are important translationally, extra-translationally, or both. Additionally, the complicated dual roles of tRNAs found in the clusters of the NER reveal a need for disentangling tRNA functions based on their proximity to nearby coding genes.

Finally, by following the evolution of tRNA genes across species, they can be used as anchors in highly duplicated regions to help explain the recent evolution of important genes such as NOTCH2NL. These regions are notoriously difficult to understand because of segmental duplication events, dynamic tRNA copy numbers, and the potential for erroneous assembly. tRNA genes are relatively easy to find with software, and important databases like gtRNAdb provide context for finding syntenic tRNA genes. These tRNAs thus can be used as anchor points in the genome to find important evolutionary events and help distinguish between orthologous genes based on adjacent tRNAs. T2T assemblies going forward will become more standard, giving a better idea of the “true” makeup of repetitive regions, but in turn, this opens new questions about the stability of tRNA regions and the evolutionary roles they play.

2.4 Methods

2.4.1 Data Acquisition

Reference genomes were obtained from the “Telomere-to-Telomere consortium primates project” via GenBank for the latest assembly releases of mGorGor1, mPanTro3, mPanPan1, mPonAbe1, mPonPyg2, and mSymSyn1 (Table S2.3) (Makova et al. 2023). CHM13v2.0/hs1 and GRCh38/hg38 reference genomes were obtained from the UCSC goldenPath. tRNA annotations for GRCh38 were obtained from gtRNAdb. A HAL alignment of the human and primate genomes was created using CACTUS as part of the Comparative Genomics Toolkit (Armstrong et al. 2020). Gene annotations were obtained from NCBI for RefSeq and UCSC by using Liftoff and the Comparative Annotation Toolkit (CAT + Liftoff) on the hg38 annotated gene set (Fiddes, Armstrong, et al. 2018; Shumate and Salzberg 2021). Additionally, BLAST was used to find possible syntenic genes relative to Human across the primates accounting for gene evolution and drift and repetitive elements (Altschul et al. 1990; Camacho et al. 2009).

2.4.2 tRNA Analysis and Annotation

tRNAscan-SE 2.0 software was used for tRNA gene identification and annotation within the assemblies, followed by the EukHighConfidenceFilter to exclude tRNA pseudogenes and tRNA-derived SINEs (Chan, Lin, et al. 2019). The HAL alignment was used to perform a liftover of tRNA and gene annotations from one assembly to another, enabling direct comparisons and analysis of tRNA genes across different coordinate systems via CACTUS (Table S2.4). Output bed files were generated, and tRNA annotations matched across assemblies using Bedtools (Quinlan and Hall 2010). Syntenic chromosomes were aligned using the matching assembly hs1 chromosome in the NCBI assembly reports with the exception of mSymSyn1. These primates’ chromosomes were manually aligned based on the highest number of syntenic tRNA genes and related adjacent genes.

2.4.3 Evolutionary Analysis

1q23.3 Region tRNA repeats were annotated by manually viewing the region and aligning them sequentially as they occur in the genome. The average substitution rate in repeat clusters was calculated by counting the number of bases per position across the aligned tRNA gene sequences, with the highest scoring base at each position acting as the consensus. The number of mismatches to the consensus was then summed across each tRNA in the TRU and divided by the amount of the same isoacceptor tRNA in all TRU for that species to adjust for variable rates of TRU copies. 1q21 regional analysis was performed using a combination of NCBI RefSeq, Liftoff+CAT, and BLAST to find syntenic gene regions across the primates, accounting for gene size, location, and proximal gene annotations. 1q21 tRNA conservation analyses were performed by creating violin plots of the tRNAs found in the region using their tRNAscan-SE scores. Codon usage bias was calculated using coding genes (CDS) from NCBI RefSeq for hs1, then compared against the brain "Tissue enhanced" (At least four-fold higher mRNA level in the brain compared to the average level in all other tissues) gene annotations for neural genes found in the Human Protein Atlas (Sjöstedt et al. 2020).

2.4.4 Visualization and Computational Analysis

In addition to the methods described above, analysis was performed in Python using Seaborn, Pandas, and Numpy Packages via Jupyter Notebooks (Harris et al. 2020; Waskom 2021; Kluyver et al. 2016). TrackHubs were created for the UCSC Genome Browser to facilitate comparisons of complicated gene regions across species (Nassar et al. 2023). Specific tRNA regions were analyzed using custom Python scripts and Bedtools to bound tRNA genomic loci between coding genes found in NCBI RefSeq gene annotation sets for each primate.

Chapter 3

Computational Tools to Facilitate Multivariate and multi-species tRNA Analysis

3.1 Background

In recent years, direct small RNA sequencing has become much more widespread and accessible with specialized tRNA sequencing methods arising, such as OTTR-seq, ARM-seq, and DM-tRNA-seq (Cozen et al. 2015; Upton et al. 2021; Zheng et al. 2015). These techniques have allowed us to perform experiments that monitor specific tRNA genes. In particular, this has been tremendously helpful in accelerating discovery of specific tRNA genes and their roles within specific isotypes, such as in stress-dependent responses to cells (G. Li et al. 2022). Ensuring consistent use of specific tRNA genes between experiments has required the use of databases and tRNA identification tools.

As a ubiquitous and diverse class of RNA molecules, tRNAs have colossal diversity in copy numbers across all species. In higher-level metazoans like Humans, the number of tRNAs can range above 500-700 depending on the particular assembly, but in species such as *Saccharomyces cerevisiae*, this is closer to 250-300 copies (Todd M. Lowe and Chan 2016). When considering a species such as *Pyrococcus furiosus* (an Archea), this number can be around 46 tRNA genomic loci. This means that tRNAs in more complex organisms often have multiple redundant copies with identical genomic sequences in various locations. Despite this, there needs to be more understanding of the complexity of tRNA diversity across many species and why some tRNAs have higher levels of conservation than others. To facilitate the discovery of new tRNAs and catalog them, gtRNAdb (the Gemomic tRNA database) and tRNAscan-SE (a genomic tRNA search tool) were created (Todd M. Lowe and Chan 2016; T. M. Lowe and S. R. Eddy 1997).

tRNAscan-SE names tRNAs as they sequentially appear in the genome and based on their chromosome. To make these names easier to understand, gtRNAdb adopted a naming convention of tRNA followed by Isotype followed by Isodecoder, then numbered by highest scoring tRNA within that isotype that matches the covariance model and the genomic loci copy number. Thus, tRNAs can be quickly identified by isotype, isodecoder, covariance score, and amount of genomic loci copies. While this system makes it easier to interpret tRNAs in a human-readable manner, it, unfortunately, makes it more difficult to compare tRNAs across species as the highest scoring tRNA in one isotype is not necessarily the same for another species. For example, Arg-TCT-4-1 in Human hg38/GRCh38 and Mouse mm10/GRCm38 refer to two different tRNAs. This can make it very difficult to annotate syntenic tRNAs across species.

In addition to naming discrepancies, some species have many tRNA-derived SINEs (Short interspersed nuclear elements) and pseudogenes. For example, in Mice, over 30000 possible tRNA pseudogenes exist before filtering. tRNA pseudogenes are also poorly understood; sometimes, they become highly conserved and serve alternative purposes in gene regulation and pre-tRNA processing (Kaçar, Beier, and Gross 1995; Mabuchi et al. 2004). While gtRNAdb tries to filter and maintain consistent naming, it will be harder to keep up as whole genome assemblies get completed faster, more often, and in greater numbers than ever.

To assist with the abundance of tRNA-sequencing data across many species, computational tRNA pipelines such as tRAX (tRNA Analysis of eXpression), which easily combine information from tRNAscan-SE and gtRNAdb with sequencing data, were created (Holmes et al. 2022). tRAX often provides a great first analysis of tRNA sequencing data as it can take known tRNA annotations, assign reads (broken by uniquely and multi-mapping tRNAs), and generate figures of general expression information. While this is helpful for projects with single experimental conditions, it is unsuitable for multiple experimental conditions, such as an experimental time course with different conditions at each time point. Furthermore, tRAX can only be run on one tRNA database (genome) at a time, requiring further analysis when comparing tRNAs across species.

Taking advantage of this gtRNAdb and tRAX framework for tRNA analysis, I have created the tools tRNAgraph and tRNAmap to facilitate the analysis of complex tRNA sequencing experiments. This means creating standardized data objects that contain all tRNA sequencing information that can be combined and analyzed in parallel. They can use new nomenclature that works across species based on unique sequences (rather than loci and score). They can also leverage the high dimensionality of the data and single-cell-like computational methods to get deeper insights into tRNAs and their functions.

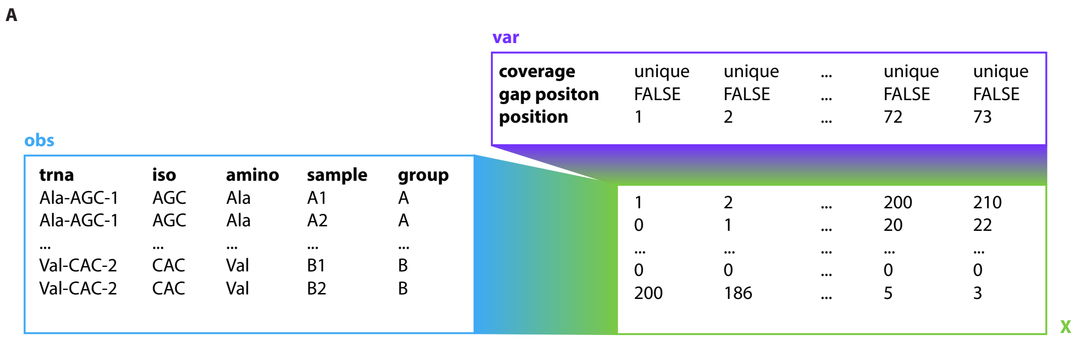
3.2 Results

3.2.1 tRNAgraph: advanced tRNA Sequencing Analysis

tRNAgraph is a tool for analyzing tRNA-seq data generated from tRAX. While tRAX generates a comprehensive set of results, it does not provide a way to visualize specific meta-data associated with a particular experiment. Additionally, tRAX cannot process multiple experimental conditions or offer insights into changes across these conditions. While the output of tRAX is useful the amount of knowledge that can be learned in these cases is limited and is broken into many hard-to-interpret files. tRNAgraph addresses this by creating a single AnnData (Annotated Data) database object from a tRAX output directory and any additional metadata provided about the samples in the sequencing experiment (Virshup et al. 2021). This single object can then be used to generate a variety of visualizations, including heatmaps, coverage plots, PCA plots, and more that are more specific to the experimental conditions of interest, as well as automated downstream analysis (Figure S3.1A).

tRNAgraph can be used with build, cluster, merge, graph, and tools commands. The build command generates an AnnData object (Virshup et al. 2021) from a tRAX coverage file. The graph command creates visualizations from the database object. The cluster command is used to cluster the database object. The merge command is used to merge two database objects. The build command will take any tRAX run as input and convert it to a single file. This object type was chosen because of its flexible nature and how it was designed to work well with large feature-heavy datasets (such as single-cell sequencing). The AnnData object is a type of hierarchical data format (HDF) used to store large amounts of data in multidimensional arrays, which is commonly used in scientific analysis. In the use case of tRNAgraph, this object has four major components: observations (obs), variables (var), unsorted (uns), and reads by position (X), each appearing as aligned data frames (Figure 3.1A).

The observations (obs) are the metadata categories derived from the sequencing data corre-



trna	iso	amino	sample	group
Ala-AGC-1	AGC	Ala	A1	A
Ala-AGC-1	AGC	Ala	A2	A
...	...	...	...	...
Val-CAC-2	CAC	Val	B1	B
Val-CAC-2	CAC	Val	B2	B

coverage	unique	unique	...	unique	unique
gap positon	FALSE	FALSE	...	FALSE	FALSE
position	1	2	...	72	73

1	2	...	200	210
0	1	...	20	22
...	...	...	...	...
0	0	...	0	0
200	186	...	5	3

Figure 3.1: tRNAgraph data structure

(A) *Schematic representation of the aligned data structure of a tRNAgraph file, with observations (obs) in blue, features (var) in purple, and data (X) in green.*

sponding to samples and conditions. By default, the tool will generate obs for tRNA, Isotype, Anticodon, Sample Name, Sample Group, and reads broken down into normalized and raw categories split by full-length, 5', 3', and other read counts based on whether a tRNA sequencing read is in the form of a fragment or full length read. These numbers are independent of the positional reads in the read data (X). Additionally, supplemental metadata, such as experimental conditions, can be added.

The variables (var) are commonly associated with "Features" in single-cell sequencing data, however, in the use case of tRNAgraph, instead, positional information is saved aligned to the read data (X). This includes all positional information from the sequencing data; since tRNAs can have many alternate positions and extensions, a position column is provided with aligned Sprinzl information (Sprinzl et al. 1996), and a gap column allows easy filtering between canonical positions and extended positions (gap positions) (Figure S3.2A). In general, these gap positions are skipped when plotting as they are non-uniform across tRNAs. Since tRAX provides coverage information that is unique and multimapping (due to the nature of tRNA sequence similarity), these categories are provided in parallel in the object. Further information is provided on tRNA positional information, such as whether a specific position is found in a 3' Half or the T-loop structure of the tRNA. Since all the read-at-position data is aligned against the var and obs data frames, it can quickly be filtered to draw any relevant information, such as a specific set of tRNAs' unique read coverage within the loop structures of tRNAs. Thus far, no other tRNA analysis software can focus on tRNAs at such specific and precise levels.

Some data generated by tRAX is important but not directly aligned with the observations (obs) and variables (var) of the read data (X). The following is unstructured data (uns), saved as independent data frames that can be recalled on command. This mostly includes precomputed data such as isotype and anticodon counts, non-tRNA small RNA reads, log-fold change of differential expressions cross conditions, as well as tRAX and tRNAgraph run information for reliable data reproduction (and prevention of merging data generated in different software versions). The merge function can combine these AnnData objects across multiple sequencing conditions or runs, which may be important if two different sequencing types are performed under the same experimental conditions. tRNAgraph facilitates this by automatically merging across the var, obs, and X-aligned parts and will combine the unstructured data appropriately.

An innate function can cluster the database object using UMAP and cluster using HDBSCAN. The default parameters used in tRNAgraph were designed to work well on ARMseq, DM-tRNAseq, and OTTRseq data; however, each dataset is different and may require fine-tuning to yield the best results. To address this, we allow many inputs to be used that can optimize the full functionality of both UMAP and HDBSCAN APIs. By default, clustering is performed across the unique coverage, read-starts, read-ends, mismatched bases, and deletions categories of the AnnData object. Clustering is performed on sample and group observations. In the case of samples, every set of reads for every single tRNA is used for clustering. In the case of groups, the mean of the reads is taken for each tRNA across the read categories and then used for clustering. This is done to reduce the number of samples used for clustering and to reduce the noise in the clustering.

tRNAgraph provides many functions to visualize data automatically as barplots (of the various tRNA coverage, tRNA, isodecoder, and isotype ratios), cluster projections (UMAP projections), correlation plots (Spearman/Pearson correlation plots of samples and coverage), coverage plots (read -coverage, -starts, and -ends), heatmaps (differential expression), sequence logos (tRNA coverage), PCA (of samples/groups), radar plots (tRNA isodecoder distributions), and volcano plots (differential tRNA expression). Each of these plots is highly customizable with various flags allowing one to filter by any metadata attributes provided. JSON config files also provide support for custom complex filtering and colors. This extensible data object is supported in various coding languages such as Python, Julia, and R for further downstream analysis.

3.2.2 tRNAmap: graph-based tRNA gene relationship exploration

Working with multiple species in a parallel analysis is difficult as gtRNAdb names do not match across species. In order to identify tRNAs that match across species, you would need to know both the tRNA gene sequence and its genomic coordinates, both prone to changing between species and even within species (Darrow and Chadwick 2014). To alleviate this monumental task, we aligned the unique sequences of all eukaryotic tRNAs found in gtRNAdb (25751 unique sequences) by their canonical positions and then calculated their sequence divergence from one another against each isotype using a Hamming Distance metric against the base pairs. We also assigned each tRNA sequence a unique sequence identifier (tRNAid) as each tRNA gtRNAdb name can be different across various species.

Arg tRNAs are a primary example of tRNA conservation patterns (because of their high copy-number and neural important tRNA-Arg-TCT-4-1), with five large subtypes appearing correlated to the five major isoacceptors with varying conservation of isodecoders therein (Figure 3.2A). The tRNAid Arg-TCT-ID25 is found across many mammals and matches the human tRNA-Arg-TCT-4-1, which is neurally important for development (Ishimura et al. 2014). In addition to this Arg-TCT, we found 15 other Arg-isodecoders found in greater than or equal to half of the eukaryotic species in the dataset (1787 unique Arg sequences), showing that a select few Arg iso-decoders are highly conserved, with many single nucleotide variations existing across species outside this core set. When looking across all isotypes, 25751 unique tRNA sequences exist in eukaryotes, 2815 exist in 2 or more species, 791 exist in 5 or more, 101 exist in 35 (about half the species), and 28 are found in (70 species). tRNAs often have low hamming divergence scores from these “core” sequences, implying subtle mutations across species or allowable SNP variations rather than degenerate rapid mutation associated with pseudogenes and those that would be rapidly transcribed, such as tRNAs.

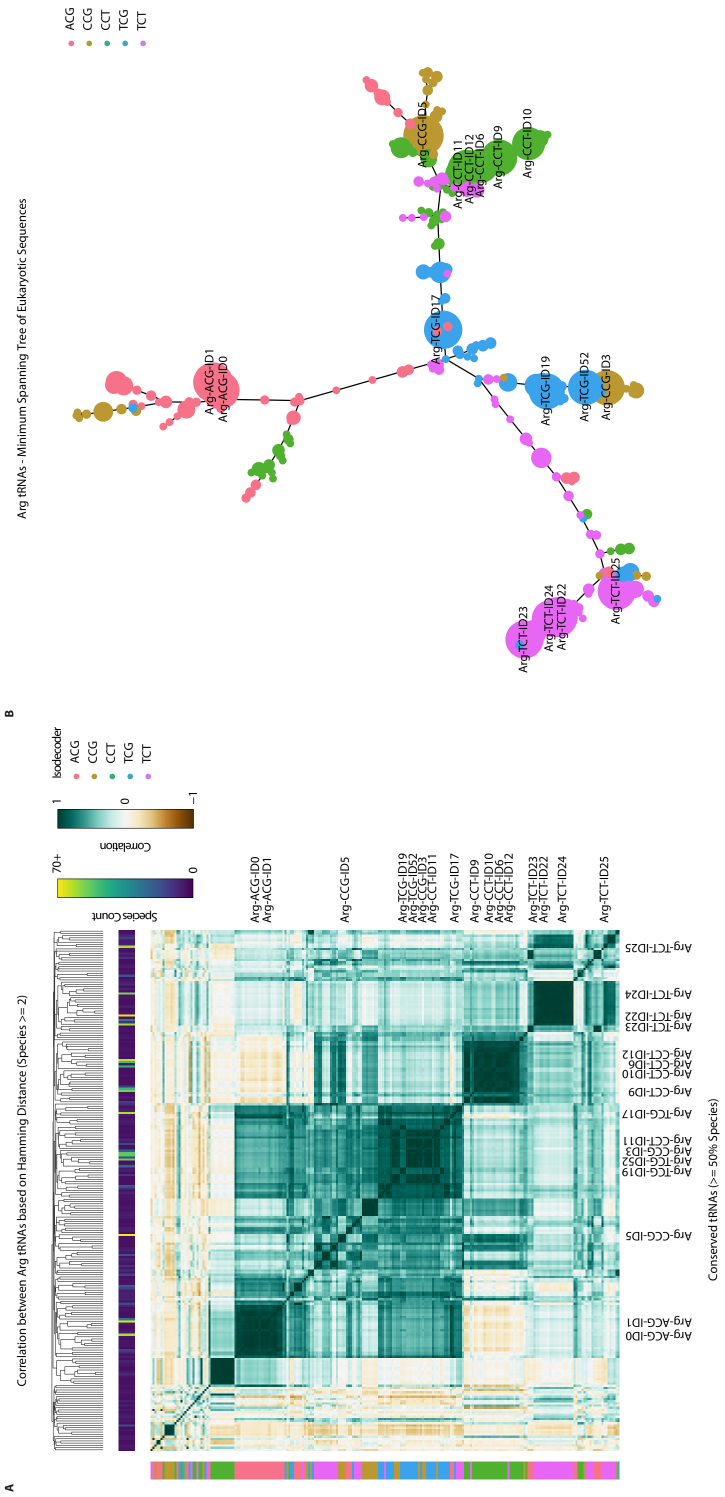

A
Correlation between Arg tRNAs based on Hamming Distance (Species >= 2)
Species Count
70+
0
Correlation
1
0
−1
Isodecoder
ACG
CCG
CCT
TCG
TCT
Arg-ACG-ID0
Arg-ACG-ID1
Arg-CCG-ID5
Arg-TCG-ID19
Arg-TCG-ID52
Arg-CCG-ID3
Arg-CCT-ID11
Arg-TCG-ID17
Arg-CCT-ID9
Arg-CCT-ID10
Arg-CCT-ID6
Arg-CCT-ID12
Arg-TCT-ID23
Arg-TCT-ID22
Arg-TCT-ID24
Arg-TCT-ID25
Conserved tRNAs (>= 50% Species)
B
Arg tRNAs - Minimum Spanning Tree of Eukaryotic Sequences
Arg-ACG-ID1
Arg-ACG-ID0
Arg-TCG-ID17
Arg-CCG-ID5
Arg-CCT-ID11
Arg-CCT-ID12
Arg-CCT-ID6
Arg-CCT-ID9
Arg-CCT-ID10
Arg-TCG-ID19
Arg-TCG-ID52
Arg-CCG-ID3
Arg-TCT-ID23
Arg-TCT-ID24
Arg-TCT-ID22
Arg-TCT-ID25
ACG
CCG
CCT
TCG
TCT

Figure 3.2

Figure 3.2: tRNAmap shows sequence relations of Arg-tRNAs

(A) *Clustermap of all Arg tRNA sequences in the Eukaryotic tRNA dataset that appear in 2 or more species, based on hamming distance from one another. The isodecoder group is shown on the y-axis, and the species count is on the x-axis.* ***(B)*** *A minimum spanning tree of the Arg tRNA sequences in the Eukaryotic tRNA dataset that appear in 2 or more species, based on hamming distance from one another. Labeled tRNAs (using tRNAid) are those found in half or more of the species in the dataset (>=35).*

We also applied Kruskal's algorithm to the hamming matrix to generate a DeBrujin graph minimum spanning trees (MST) for all isotypes. The purpose was to see the linkages between different tRNA sequences and to establish a graph structure of the data for exploration. This revealed that 1) sequence variation is less likely to happen to an anticodon as isodecoders tended to correlate highly with one another and 2) some tRNAs broke this trend, indicating evolutionarily divergent isodecoders. In the case of Arg tRNAs, nearly all conserved isodecoders (2 or more species) are closely related (less distance), with the exception of 3 groups of Arg-CCG (Figure 3.2B). The three groups each are more closely related to a different set of anticodons ACG, TCG, and CCT, with the latter two containing the two most commonly conserved Arg-CCG with tRNAid Arg-CCG-ID3 and Arg-CCG-ID5 respectively.

We wanted this data to be available and easily filtered, so we packaged it into a Plotly-based web application. In addition to containing the precomputed node graphs, we attached associated data for: tRNAid, human (hg38) syntenic name, isoacceptor, isotype, species counts and species list, tRNAscan-SE associated scores, GC content, and adjacency scores (number of proximal nodes). This list allows for easy filtering in the context of the web app as well as custom coloring for visualization. Additionally, the database was designed for easy upgradability as new species genomes are curated by using insertion into the hamming matrix to prevent high computational cost as well as new tRNAids sequentially being added to prevent clashing.

3.2.3 Multivariate tRNA Sequencing Analysis

When combined, both tools can be used for powerful cross-species analysis. We wanted to check the relative levels of tRNAs in three species: human (c305), chimpanzee (epi8919), and orangutan (jos3c1). These three species represent important great apes in the human primate lineage and have a massive expansion of the 1q21 region of human chromosome 1; a region containing many tRNAs and tRNA pseudogenes. The respective stem cells were grown into cerebral cortical organoids as a model to study tRNA changes in early human embryogenesis. A standard time course of culturing conditions was used to achieve neural induction and organoid growth. Cells were aggregated into embryoid bodies and differentiated into cortical tissues over a period of 10 weeks. Samples were taken for characterization and sequencing at the developmental time points of day 0 (D0 - stem cells), day 35 (D35 - emergence of deep layer neurons), and day 70 (D70 - continued generation of cortical projection neurons, emergence of outer radial glia) based on the emergence of key cell types in the course of dorsal cortex development in previous studies (Pollen et al. 2019).

Isodecoder	**Human**	**Chimpanzee**	**Orangutan**
AAT	15	13	13
GAT	3	2	3
TAT	5	5	6

Table 3.1: IF-Staining Antibodies
List of primary and secondary antibodies concentrations used for IF-staining.

We then used OTTR-seq to get the expression levels of tRNAs and tDRs from these primate organoids, as it is a robust protocol for obtaining tRNA sequencing data (Gustafsson et al. 2022; Upton et al. 2021). When comparing the relative levels of isodecoders across the three species and three neurodevelopmental time points, we found that Ile diverged in chimpanzees relative to humans and orangutans (Figure 3.3A). In the case of humans and orangutans, GAT isodecoders were around $\sim$20-30% of the isodecoder pool at timepoint 0 but reduced to less than 5% in all neural timepoints (35/70). In the case of chimpanzees, however, GAT levels are higher on day 0 and remain consistently high throughout embryogenesis, with a severe reduction in AAT at day 70. The genomic loci for chimpanzee Ile tRNAs surprisingly have one less GAT genomic loci compared to the other two species, suggesting that copy number alone is indicative of the expression of a single isodecoder (Table 3.1).

We compared the neurally important human tRNA-Arg-TCT-4-1 to the primates (tRNA-Arg-TCT-3-1) we found broadly similar expression patterns across organoids (Figure 3.3B). The expression in each primate drastically increases as the cells differentiate into neurons. Additionally, coverage drops around

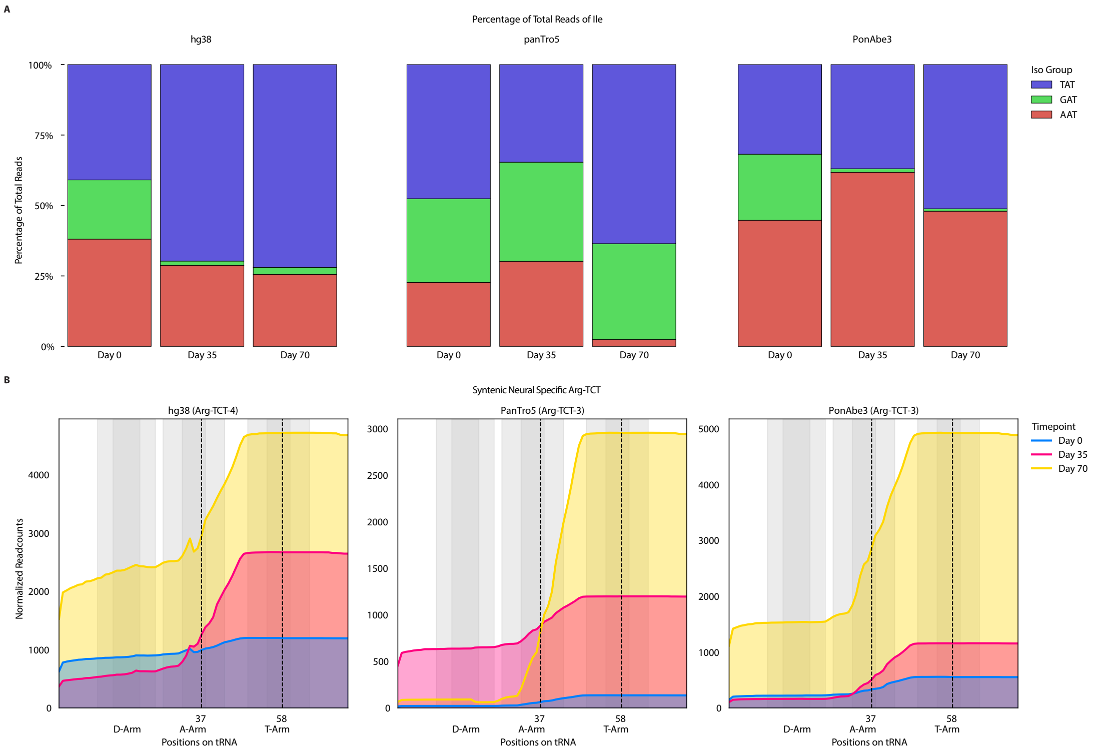


Figure 3.3: tRNAgraph facilitates multivariate analysis

(A) *Reads-counts distributions showing the relative ratio of Ile tRNA isodecoder across developmental time-course for human (hg38), chimpanzee (panTro5), and orangutan (ponAbe3) organoids.* ***(B)*** *Coverage profiles for the tRNA-Arg-TCT4-1 (human) and tRNA-Arg-TCT-3-1 (chimpanzee/orangutan) syntenic tRNA that is highly expressed over organoid neurogenesis. D-loop, anticodon-loop, and T-loop are shown in gray, with lines added at specific coverage breakpoints in each plot.*

position 40 in all primates, suggesting a consistent tRNA-derived small RNA (tDR) fragment is forming in each primate. tRNAs are heavily modified and associated with tDR fragmentation (Kuhle, Q. Chen, and Schimmel 2023), demonstrating that these primates' modification pathways are conserved.

After normalization and batch correction, the combined primate data can be clustered via UMAP to find commonalities in the tRNAs and tDR expression patterns. We found that tRNAs/tDRs tended to discriminate cleanly by isotype despite not being given any information about nucleotides (Figure 3.4A). This suggests that tRNAs/tDRs specific bases play roles in unique fragment formation and expression. We also looked at the distribution of data by species and time point and found that no biases existed within those. Since the majority of clusters are discriminated by sequence, we wanted to check the underlying modifications that are linked to these expression and fragmentation patterns.

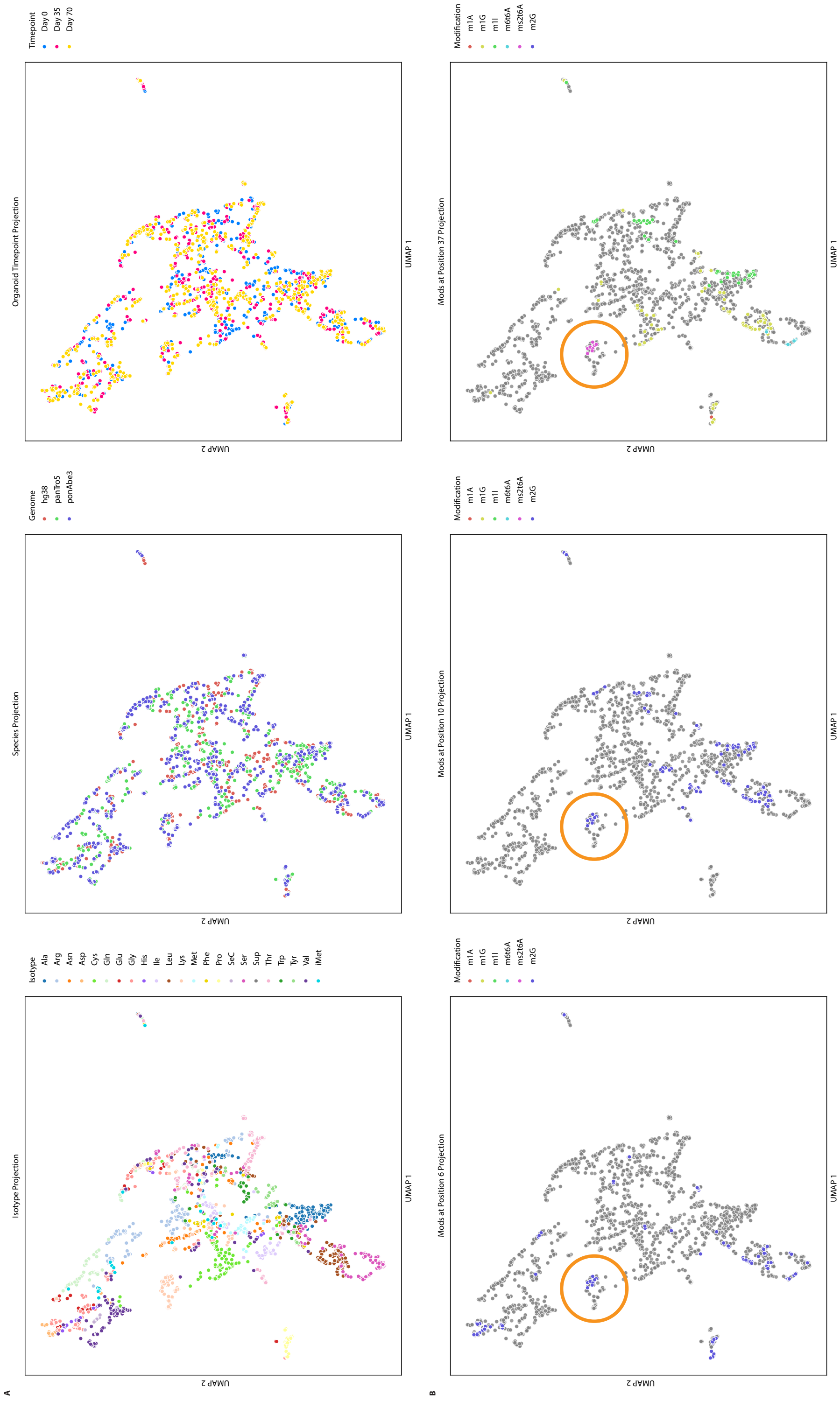

A
Isotype Projection
Isotype
Ala
Arg
Asn
Asp
Cys
Gln
Glu
Gly
His
Ile
Leu
Lys
Met
Phe
Pro
SeC
Ser
Sup
Thr
Trp
Tyr
Val
iMet
Species Projection
Genome
hg38
panTro5
ponAbe3
Organoid Timepoint Projection
Timepoint
Day 0
Day 35
Day 70
B
Mods at Position 6 Projection
Mods at Position 10 Projection
Mods at Position 37 Projection
Modification
m1A
m1G
m1I
m6t6A
ms2t6A
m2G
UMAP 1
UMAP 2

Figure 3.4

Figure 3.4: UMAP clustering of modification data

(A-B) *UMAP projections of tRNA sequencing profiles using unique coverage, 3' and 5' read ends (considering the tRNA structure), deletions, and mismatched bases at the position after HDBSCAN clustering.* ***(A)*** *Projections of isotype, species, and time point masked by HDBSCAN unannotated are also shown.* ***(B)*** *Projections of modifications at positions 6, 10, and 37 mapped to hg38 only. The region of interest is shown with an orange circle.*

We took observed Human modification data based on Modomics and misincorporations and found we could use it to infer those same modifications on similar primates based on matching expression patterns. We found one particular cluster of lysines broken into two subgroups (group cluster 5/6) with an m^2G at positions 6 and 10 and a ms^2t^6a at position 37 (Figure 3.4B; Table S3.1). The human tRNAs tRNA-Lys-TTT-1/3/4/6 fell into this group, as well as the chimpanzee (panTro5) tRNA-Lys-TTT-1/3/4/6 and orangutan (ponAbe3) tRNA-Lys-TTT-4/5. The other proximal Lys group (6) was made mostly of tRNA-Lys-CTT, only containing the common m1A at position 58. tRNAgraph with tRNAmap thus provides a useful framework for cross-species tRNA analysis.

3.3 Discussion

tRNAgraph and tRNAmap work together to achieve cross-species tRNA analysis. Creating a list of tRNAids based on tRNA gene sequence allows us to focus on tRNAs that are the same regardless of gene loci. tRNAs can often multimap to several locations, so to alleviate the burden of genomic location, focus can be placed on functional differences in the expression as they pertain to sequences independent of genomic loci. Eukaryotic studies of tRNAs focus mostly on yeast (*saccharomyces cerevisiae*), humans, and mice with different tRNA gene sets and gtRNAdb IDs. Neural critical tRNAs like tRNA-Arg-TCT-4-1 in humans are known as tRNA-Arg-TCT-3-1 in chimpanzees and orangutans. Instead, by using a unique tRNA-ID tied to sequence this ties all these tRNAs together.

The generation of tRNA sequencing data is becoming increasingly frequent, and there is a need for specific analysis that can discriminate between subtle differences. tRNAgraph handles much of the overhead of complicated dimensionality reduction by conveniently aligning and organizing all tRNAseq information before clustering the data. Additionally, tRNAgraph allows independent tRNA sequencing datasets to be combined and batch-corrected. Combined with tRNAids, the datasets can be combined across multiple species, allowing for complicated cross-species pattern matching, such as combining human cardiomyocyte cells and mouse heart tissue. As tDRs are becoming frequently recognized for translation alternative roles, they can thus be used as therapeutic targets found via this pipeline.

While tRNAmap improves understanding of tRNA genomic sequence frequency and evolutionary conservation, it does not contain information about genomic loci. We can understand the relative ratios of tRNA isodecoders in different primates, such as the chimpanzee having two Ile-GAT tRNA genes against human/orangutans with three Ile-GAT but the software alone cannot provide an explanation for why certain tRNAs are more abundant than others. These questions can be addressed by meaningfully combining more sequencing data. As greater importance is given to tRNA modification patterns, the need for tools to predict these modifications is apparent.

3.4 Methods

3.4.1 Organoid methods

Human GM12878-c305, Chimpanzee Epi-8919, and Orangutan Jos-3C1 induced pluripotent stem cells (iPSCs) were used to generate cerebral cortical organoids. Undifferentiated cells were grown in feeder-free conditions on matrigel (Corning) with mTeSR Plus (Stemcell Technologies) or on vitronectin with Essential-8 Flex media (ThermoFisher). Cerebral cortex organoids were generated using a protocol adapted from Kadoshima et al. 2013. 10,000 cells per embryoid body were aggregated using AggreWell-800 plates in AggreWell media (Stemcell Technologies) supplemented with 10 uM Y-27632 rock inhibitor (Stemcell Technologies) and transferred to low attachment 6-well dishes (Corning) on day 2. These methods supplemented the respective media with 10 uM SB431542 (SB, Millipore), and 1 uM IWR-1 (Millipore) for the first 14 days of differentiation. The media was then changed to Sasai II media (DMEM/F12 + glutamax supplemented with N2) on day 14. At this point, organoids were supplemented with 10ng/mL beta fibroblast growth factor (bFGF) and 10ng/mL epidermal growth factor (EGF) to improve survival in Sasai II media. On day 18 the organoids were transferred to an in-incubator orbital shaker (100 rpm) where they remained for the duration of the experiment. After day 35, all cultures were grown in Sasai

III media (Sasai II media supplemented with 50 mL of FBS (Hyclone)) supplemented with ten ng/mL brain-derived neurotrophic factor (BDNF) and 10 ng/mL of neurotrophin-3 (NT-3) (Kindberg et al. 2014).

3.4.2 RNA Isolation and PNK Treatment

Isolation of total RNA from cerebral cortical organoids and stem cells was performed using Direct-Zol RNA MiniPrep Kit (Zymo Research) with TRI Reagent (Molecular Research Center, Inc.). Since a single organoid would yield far less total RNA, approximately 3-8 organoids would be pooled in Trizol depending on organoid size for each sample replicate. The manufacturer's recommended volume of TRI Reagent was added to each sample ($\sim$1 mL). For RNA purification of stem cells, Trizol was directly added to cell culture plates on ice, scraped, and homogenized via pipetting. Organoids in Trizol were broken down via pipetting inside a 1mL Eppendorf tube on ice via a syringe. All total RNA was processed using a MirVana miRNA Isolation Kit (Life Technologies), according to the manufacturer's instructions, to select for RNA <200 nt. This was followed by an RNA Clean and Concentrate-25 (Zymo Research). RNA 3' de-phosphorylation was carried out as previously described (Huppertz et al., 2014) using 1ug of total RNA from each sample. Briefly, samples were treated with T4 Polynucleotide Kinase (T4PNK; New England Biolabs) in a modified 5x reaction buffer (350 mM Tris-HCl, pH 6.5, 50 mM MgCl2, 5mM dithiothreitol) under low pH conditions in the absence of ATP for 30 mins.

3.4.3 OTTR-seq Library Preparation

OTTR-seq libraries were generated as previously described (Upton et al. 2021). Briefly, total PNK-treated RNA was 3' tailed using mutant BoMoC RT in buffer containing only ddATP for 90 minutes at 30°C, with the addition of ddGTP for another 30 minutes at 30°C. This was then heat-inactivated at 65°C for 5 minutes, and unincorporated ddATP/ddGTP was hydrolyzed by incubation in 5 mM MgCl2 and 0.5 units of shrimp alkaline phosphatase (rSAP) at 37°C for 15 minutes. 5 mM EGTA was added and incubated at 65°C for 5 minutes to stop this reaction. Reverse transcription was then performed at 37°C for 30 minutes, followed by heat inactivation at 70°C for 5 minutes. The remaining RNA and RNA/DNA hybrids were then degraded using 1 unit of RNase A at 50°C for 10 minutes. cDNA was then cleaned up using a MinElute Reaction CleanUp Kit (Qiagen). To reduce adaptor dimers, cDNA was run on a 9% UREA page gel, and the size range of interest was cut out and eluted into gel extraction buffer (300mM NaCl, 10mM Tris; pH 8.0, 1mM EDTA, 0.25% SDS) and concentrated using EtOH precipitation. Size-selected cDNA was then PCR amplified for 12 cycles using Q5 High-fidelity polymerase (NEB #M0491S). Amplified libraries were then run on a 6% TBE gel, and the size range of interest was extracted to reduce adaptor dimers further. Gel slices were eluted into gel extraction buffer (300mM NaCl, 10mM Tris; pH 8.0, 1mM EDTA) followed by concentration using EtOH precipitation. Final libraries were pooled and sequenced on an Illumina NextSeq 500 150-cycle high-output kit.

3.4.4 RNA sequencing and differential expression analysis

Libraries prepared for OTTR-seq were sequenced using Illumina NextSeq 500 150-cycle high-output kit. Single-ended reads were produced as FASTQ files that were analyzed using tRNA Analysis of eXpression (tRAX) (Holmes et al. 2022). Sequence adapters were trimmed, and pair-ended reads were merged using the tool trimadapters.py in tRAX. The reference database for tRAX was built with high-confidence tRNA predictions retrieved from the Genome tRNA Database (Chan and Lowe, 2016) and the sequences of human genome assembly GRCh38. Other gene annotations were obtained from Ensembl release 102. Biological replicates of organoids at each time point were grouped as sample replicates for tRAX inputs, and different time points were marked as pairs for differential expression comparison for each sequencing type.

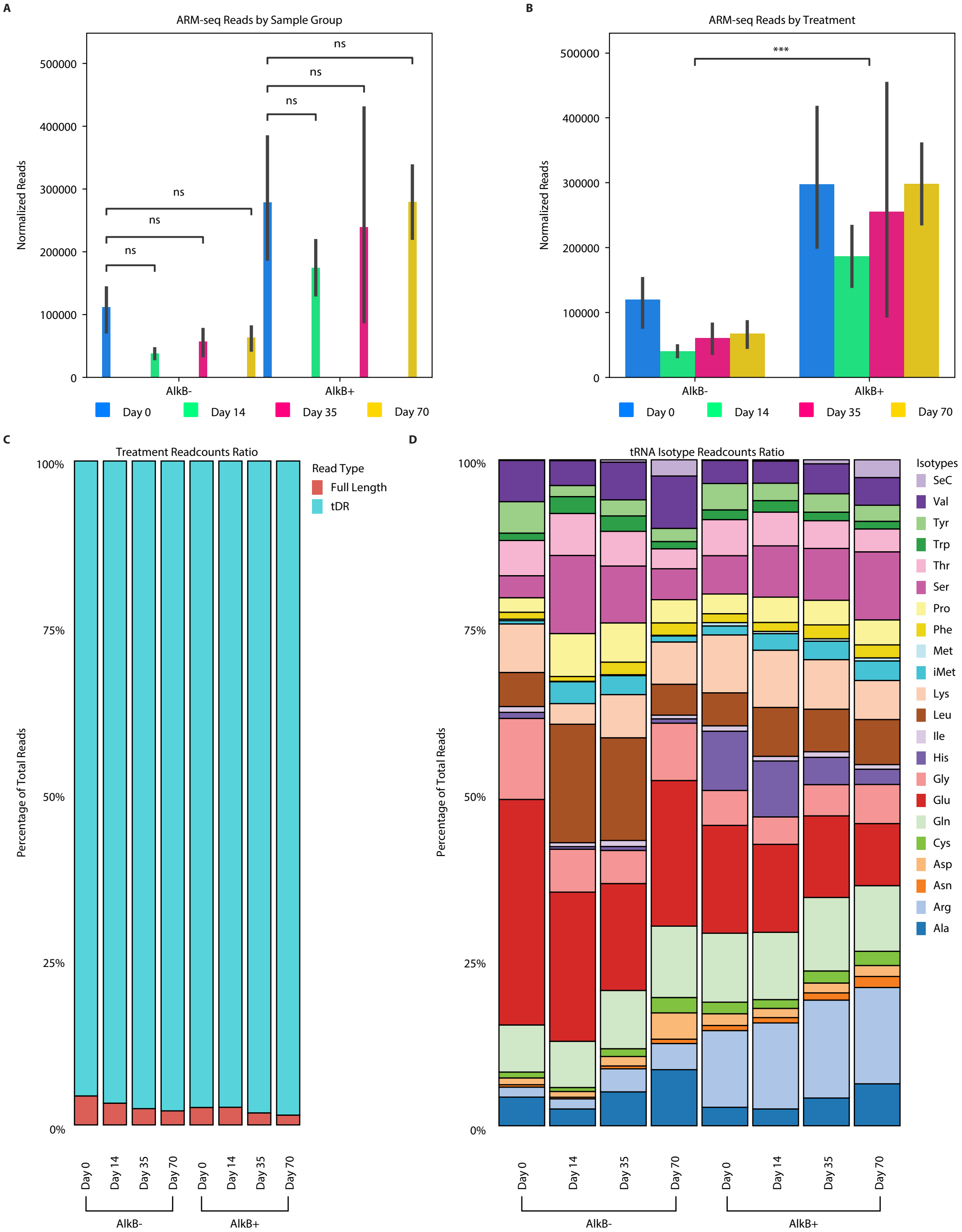
3.4.5 Multispecies tRNA Gene Alignment

Eukaryotic genomes were chosen by taking species available on gtRNAdb and then aligning the tRNA sequences based on their Sprinzl positions (Chan, Lin, et al. 2019; Chan and Todd M. Lowe 2016; Sprinzl et al. 1996). A hamming distance matrix for the sequences in the alignment table was created via the Pandas Python package (McKinney 2010). A directed node graph was then computed from the distance matrix using NetworkX, and a minimum spanning tree was generated using Kruskal’s algorithm (Hagberg, Swart, and Chult 2008; Kruskal 1956). These were then turned into an interactive Python web app via Plotly.

Part II

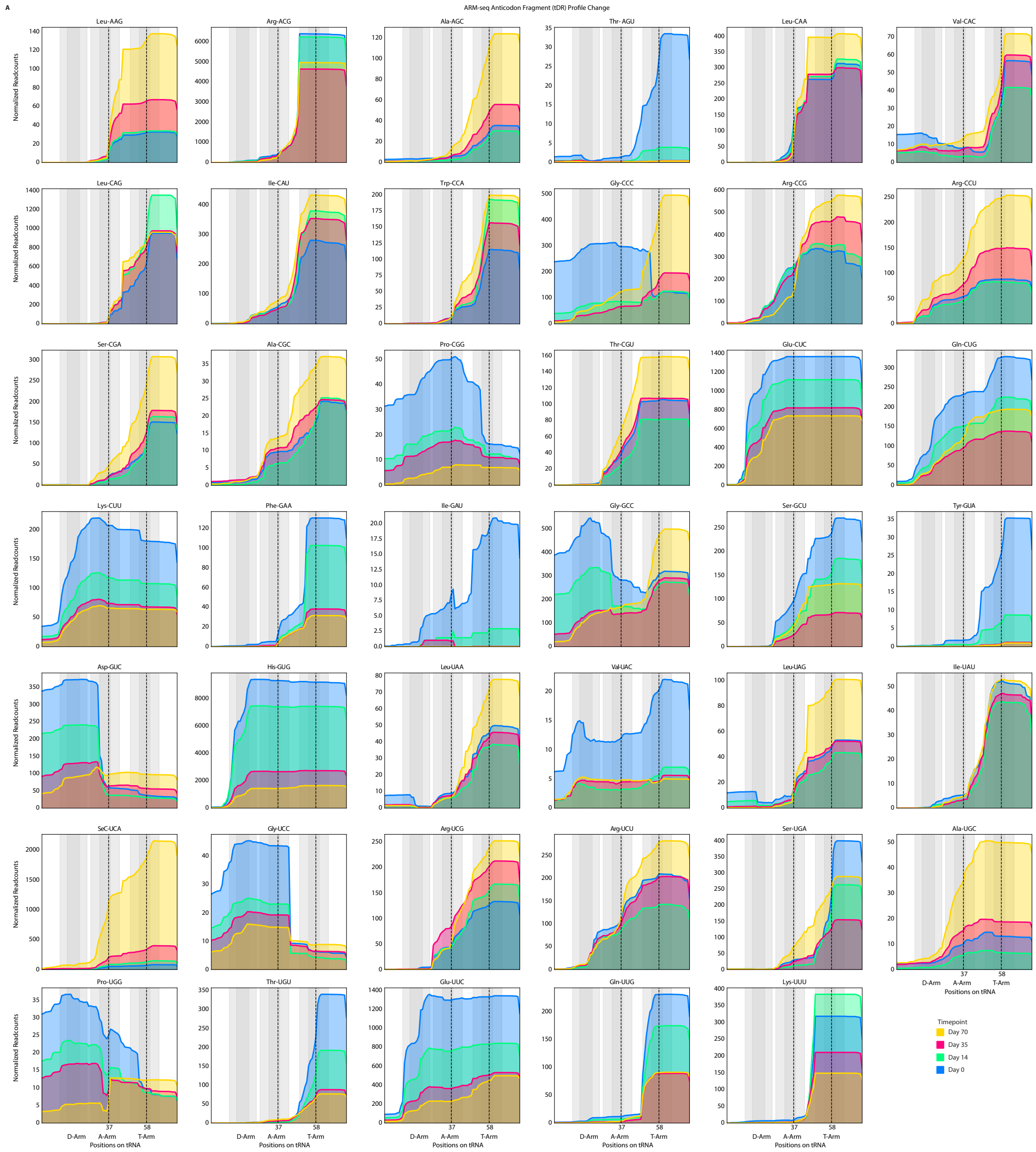
Supplemental Data

Chapter 1: Supplemental Figures


A
ARM-seq Reads by Sample Group
Normalized Reads
500000
400000
300000
200000
100000
0
ns
ns
ns
ns
ns
ns
AlkB-
AlkB+
Day 0
Day 14
Day 35
Day 70
B
ARM-seq Reads by Treatment
Normalized Reads
500000
400000
300000
200000
100000
0
***
AlkB-
AlkB+
Day 0
Day 14
Day 35
Day 70
C
Treatment Readcounts Ratio
Percentage of Total Reads
100%
75%
50%
25%
0%
Read Type
Full Length
tDR
Day 0
Day 14
Day 35
Day 70
Day 0
Day 14
Day 35
Day 70
AlkB-
AlkB+
D
tRNA Isotype Readcounts Ratio
Percentage of Total Reads
100%
75%
50%
25%
0%
Isotypes
SeC
Val
Tyr
Trp
Thr
Ser
Pro
Phe
Met
iMet
Lys
Leu
Ile
His
Gly
Glu
Gln
Cys
Asp
Asn
Arg
Ala
Day 0
Day 14
Day 35
Day 70
Day 0
Day 14
Day 35
Day 70
AlkB-
AlkB+

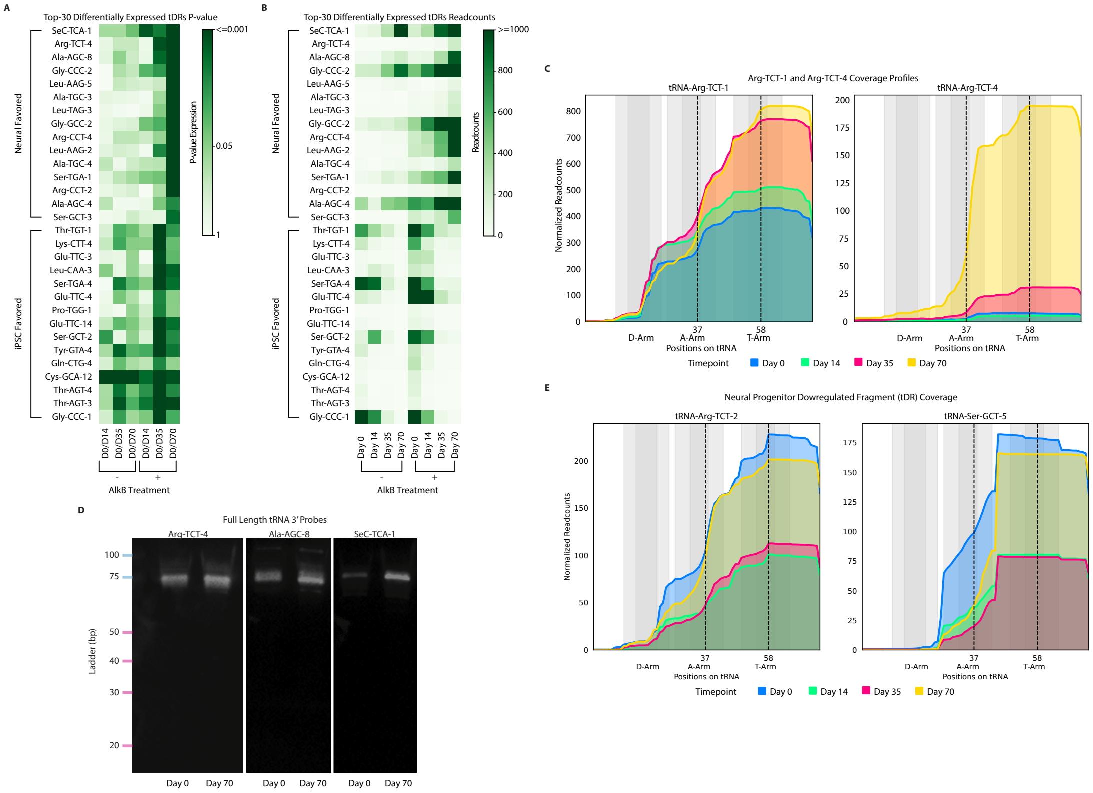
Supplemental Figure S1.1: ARM-seq expression patterns

(A-B) Normalized readcount distribution for tRNA reads by (A) AlkB treatment and (B) timepoint for ARM-seq with significance by Welch's t-test shown as ns (p <= 1), * (1e-02 < p <= 5e-02), ** (1e-03 < p <= 1e-02), *** (1e-04 < p <= 1e-03), and **** (P <= 1e-04). (C) Human tRNA and tDR reads-counts distributions showing relative ratio of fragments vs full-length tRNAs across organoid time-course, sequencing type and alkB treatment. (D) Human tRNA isotype reads-counts distributions showing relative ratio of tRNA isotypes across organoid time-course, sequencing type and alkB treatment.



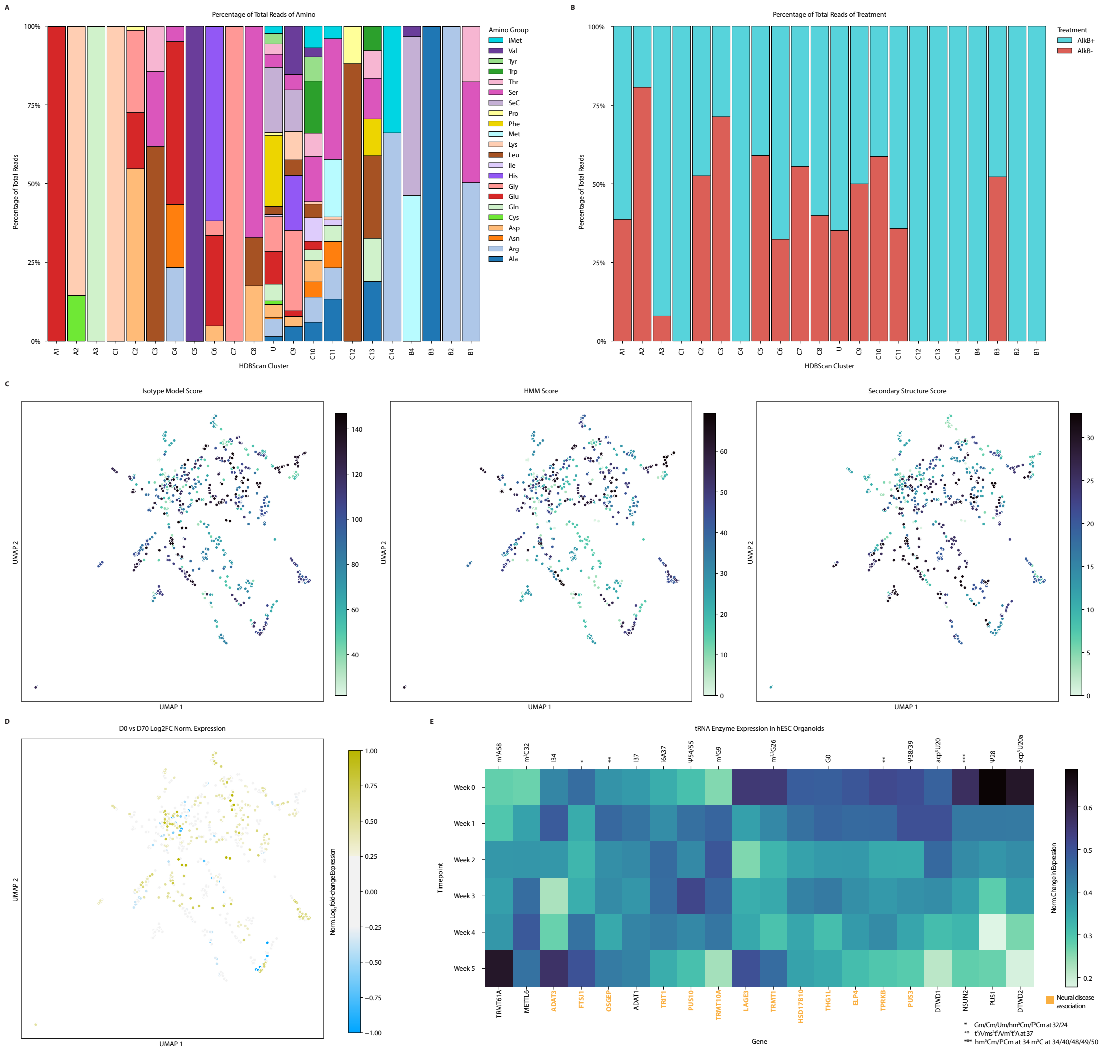
Supplemental Figure S1.2: Maximized fragment distributions of tDR Isotypes

(A) Coverage profiles for tDRs over organoid neurogenesis. D-loop, anticodon-loop, and T-loop shown in gray, with lines added at specific coverage breakpoints in each plot. The mean value was taken across all coverage profiles within a specific isotype at each position, with isodecoders not expressing a mean >= 20 reads are not displayed.



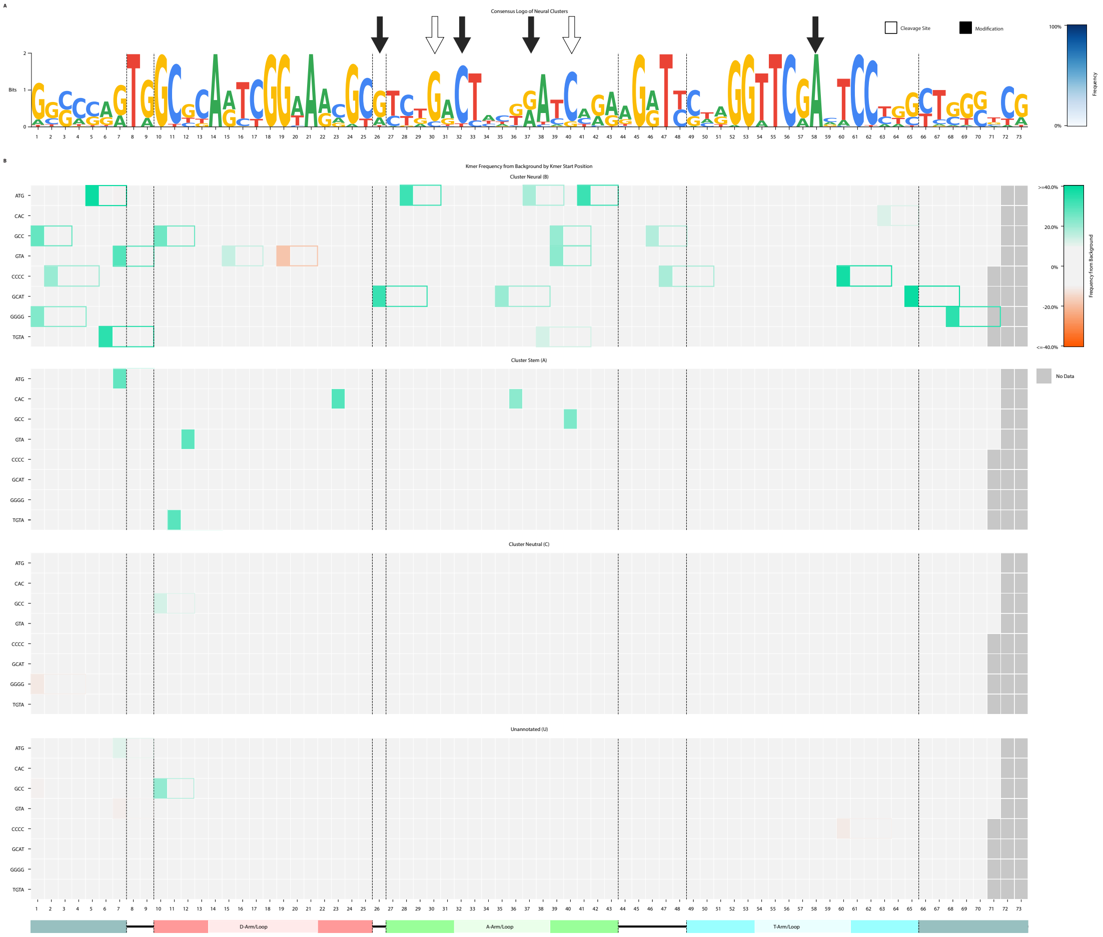
Supplemental Figure S1.3: tRNA expression in neural organoids

(A-B) Heatmap showing the top 15 neural and top 15 stem favored tDRs across all conditions by p-value (A) and readcounts (B). P-value and normalized read counts are shown in green with a cutoff of <=0.001 and >=1000 respectively. (C) Coverage profiles for Arg-TCT-1 and Arg-TCT-4 ARM-seq over organoid neurogenesis. D-loop, anticodon-loop, and T-loop shown in gray, with lines added at specific coverage breakpoints in each plot. (D) Northern blot analysis of Arg-TCT-4, Ala-AGC-8 and SeC-TCA-1 probs at full length mature tRNAs between day 0 and day 70 timepoints.(E) Coverage profiles for the tDRs that appear to downregulate with neural progenitors (D14 and D35). D-loop, anticodon-loop, and T-loop shown in gray, with lines added at specific coverage breakpoints in each plot.



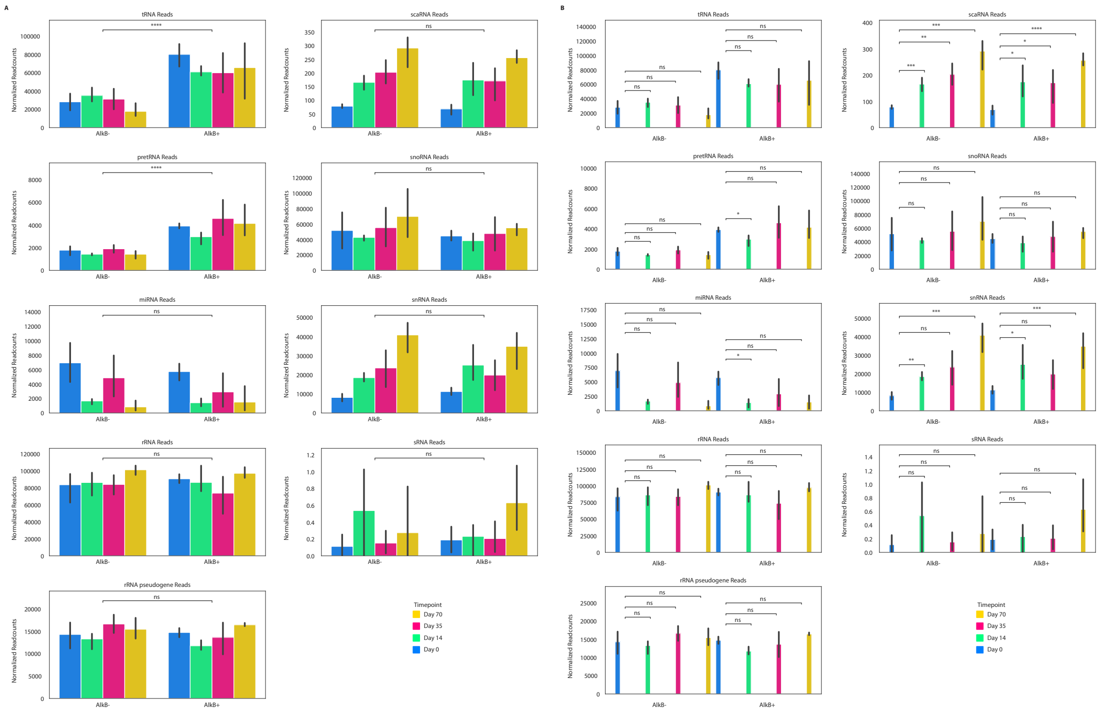
Supplemental Figure S1.4: tDR clustering and associated enzyme expression

(A-B) Bar plots of each tDR class's relative isotype (A) and AlkB treatment (B) distribution by each hdbscan class. (C-D) UMAP projections of tRNA sequencing profiles using unique coverage, 3' and 5' read ends (with consideration to the tRNA structure), deletions, and mismatched bases at position after hdbscan clustering. Projections of (C) tRNAscan-SE covariance and HMM as well as secondary structure (Infernal) scores and (D) Log2 fold-change normalized expression between day 0 and day 70 timepoints masked by unannotated (U) class are shown. (E) Normalized base-mean tRNA enzyme expression from human embryonic stem cells cultured into cerebral cortical organoids on a 5 week (day 35) time course.



Supplemental Figure S1.5: K-mer frequency distribution

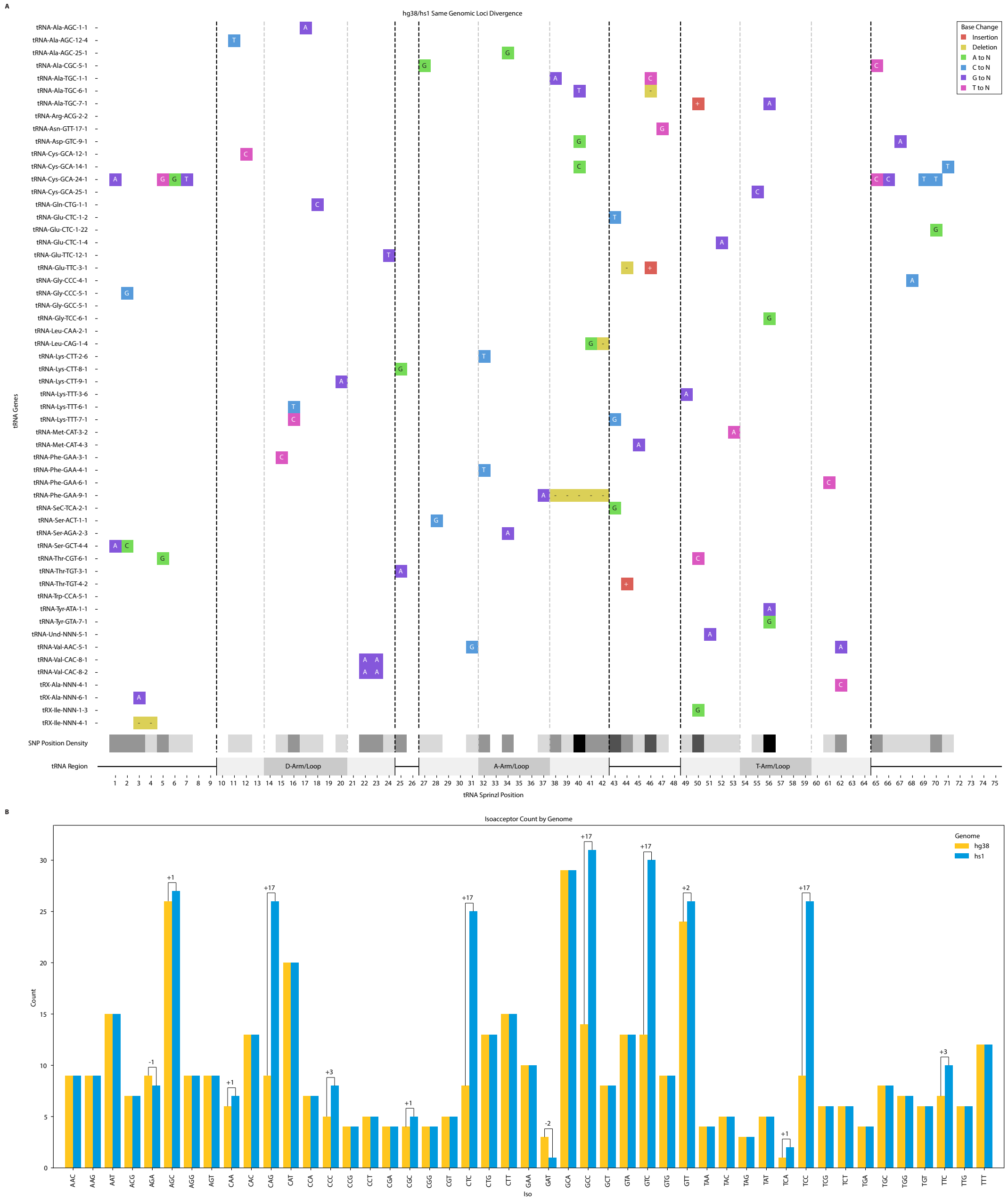
(A) Sequence logo representation of the read counts for tRNAs found in neural (B) clusters dropping non-expressed tRNAs (>20 reads) and treating each sequence as if the full sequence was transcribed to prevent 5'/3' bias. (B) Kmer frequency plot showing the percent frequency change from background kmer frequency for neural (B), neutral (C), stem (A) and unannotated (U) clusters, after filtering out low (>15) read counts.



Supplemental Figure S1.6: Mean expression of small RNA reads

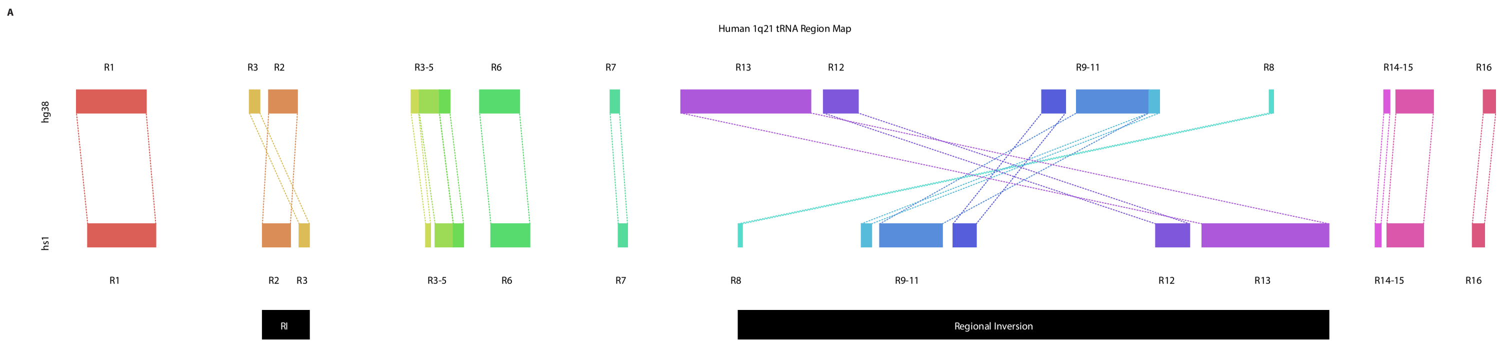
*(A-B) Normalized readcount distribution for tRNA reads by (A) AlkB treatment and (B) timepoint for ARM-seq with significance by Welch's t-test shown as ns (p <= 1), * (1e-02 < p <= 5e-02), * * (1e-03 < p <= 1e-02), * * * (1e-04 < p <= 1e-03), and * * * * (P <= 1e-04).*

Chapter 2: Supplemental Figures



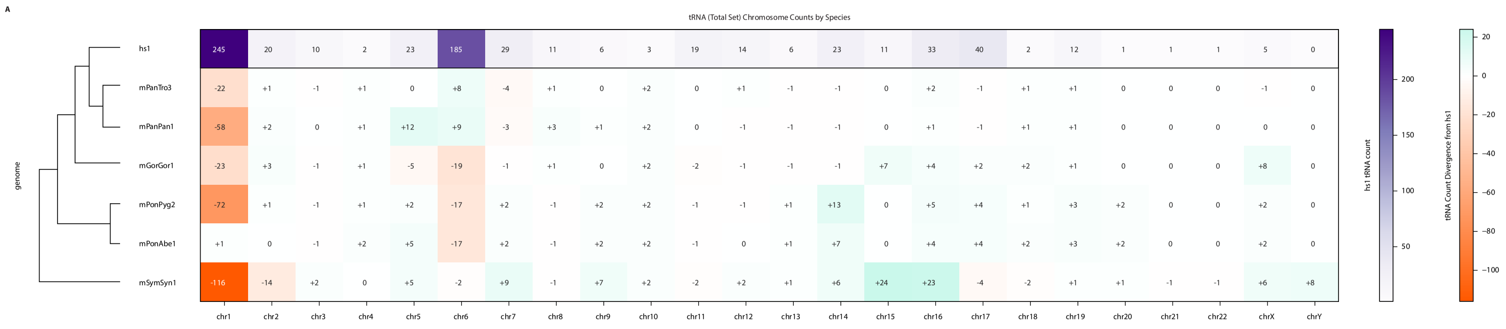
Supplemental Figure S2.1: Specific Human tRNA Base Changes

(A) Heatmap showing all tRNA genomic loci divergence between hg38 and hs1 aligned by canonical position and SNP density at each position normalized relative to all other SNP frequencies in the plot. (B) Bar plot showing the relative counts of tRNAs found across isoacceptor in hg38 (yellow) and hs1 (blue).


A
Human 1q21 tRNA Region Map
hg38
hs1
R1
R3
R2
R3-5
R6
R7
R13
R12
R9-11
R8
R14-15
R16
R1
R2
R3
R3-5
R6
R7
R8
R9-11
R12
R13
R14-15
R16
RI
Regional Inversion

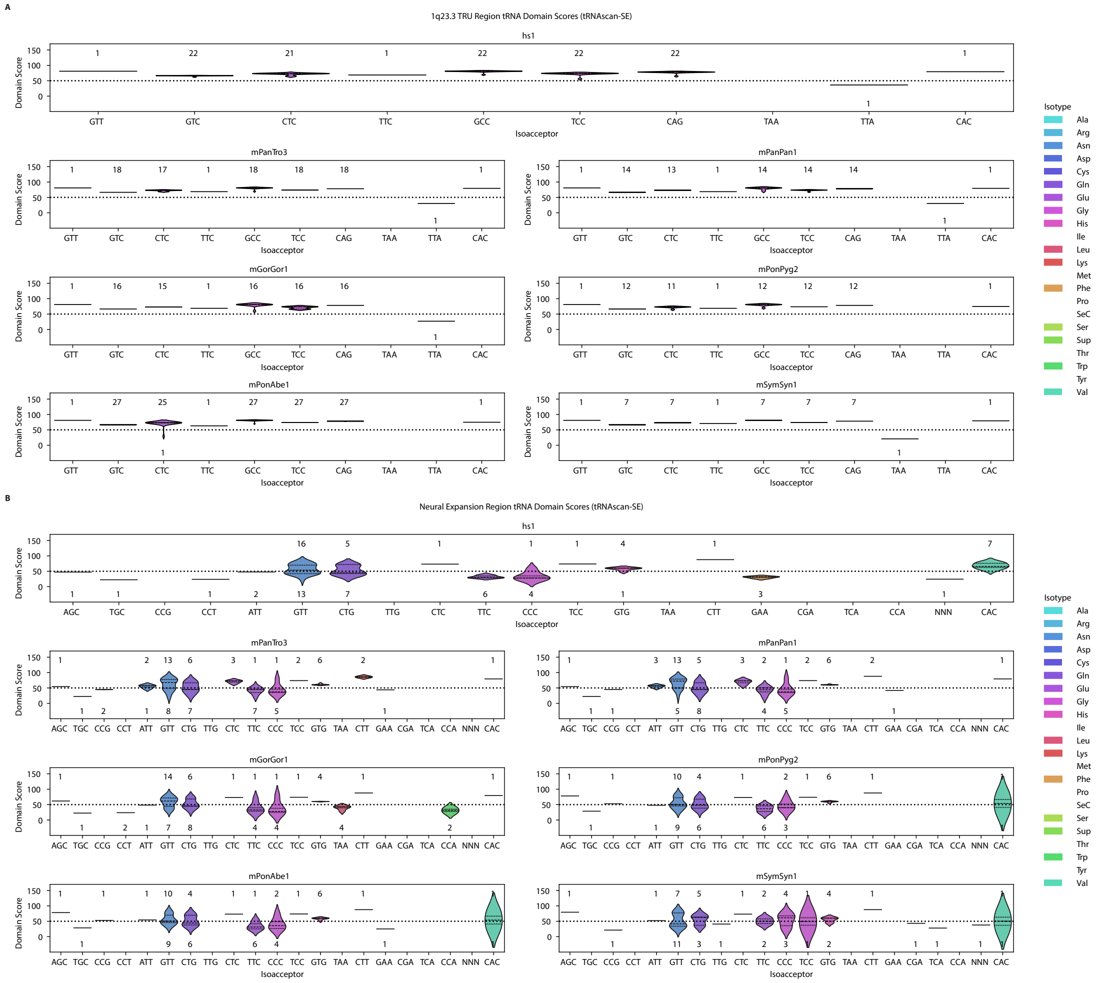
Supplemental Figure S2.2: Schematic of human 1q21 region between hs1/hg38

(A) Schematic representation of human 1q21 region between hs1 and hg38 with different colored bars representing tRNA regions defined as tRNA genomic loci bound by protein-coding genes and regional inversions (RI) highlighted in black.


A
tRNA (Total Set) Chromosome Counts by Species
genome
hs1: 245, 20, 10, 2, 23, 185, 29, 11, 6, 3, 19, 14, 6, 23, 11, 33, 40, 2, 12, 1, 1, 1, 5, 0
mPanTro3: -22, +1, -1, +1, 0, +8, -4, +1, 0, +2, 0, +1, -1, -1, 0, +2, -1, +1, +1, 0, 0, 0, -1, 0
mPanPan1: -58, +2, 0, +1, +12, +9, -3, +3, +1, +2, 0, -1, -1, -1, 0, +1, -1, +1, +1, 0, 0, 0, 0, 0
mGorGor1: -23, +3, -1, +1, -5, -19, -1, +1, 0, +2, -2, -1, -1, -1, +7, +4, +2, +2, +1, 0, 0, 0, +8, 0
mPonPyg2: -72, +1, -1, +1, +2, -17, +2, -1, +2, +2, -1, -1, +1, +13, 0, +5, +4, +1, +3, +2, 0, 0, +2, 0
mPonAbe1: +1, 0, -1, +2, +5, -17, +2, -1, +2, +2, -1, 0, +1, +7, 0, +4, +4, +2, +3, +2, 0, 0, +2, 0
mSymSyn1: -116, -14, +2, 0, +5, -2, +9, -1, +7, +2, -2, +2, +1, +6, +24, +23, -4, -2, +1, +1, -1, -1, +6, +8
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX
chrY
hs1 tRNA count
200
150
100
50
tRNA Count Divergence from hs1
20
0
−20
−40
−60
−80
−100

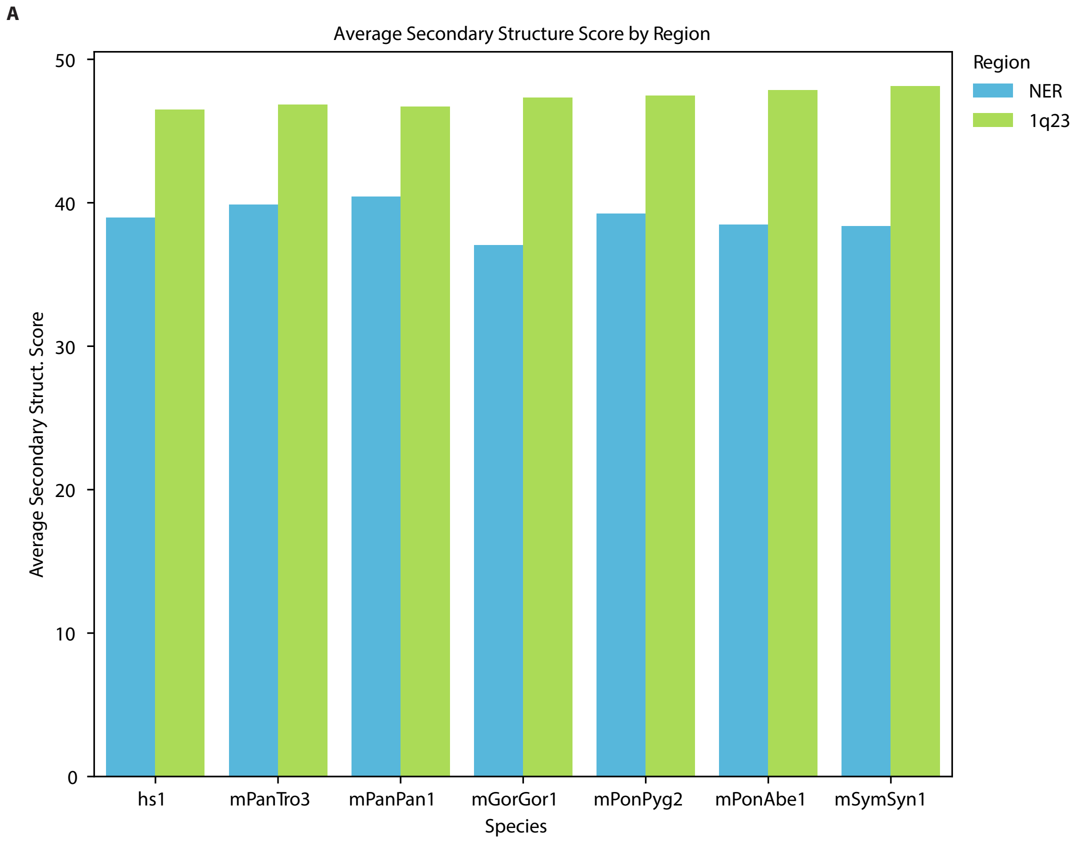
Supplemental Figure S2.3: Total Predicted tRNAs and Pseudogenes in T2T Primates

(A) Heatmap of all predicted tRNA counts found in hs1 and aligned primate chromosomes. The top row in purple shows total tRNA counts in hs1 (human) and the divergence in counts from Human in the primates from green (increase) to orange (decrease) relative to hs1. The phylogenetic tree is located on the left and arranged from most to least similar to humans.



Supplemental Figure S2.4: Sequence Logos of T2T Apes in 1q23.3 tRNA Repeats

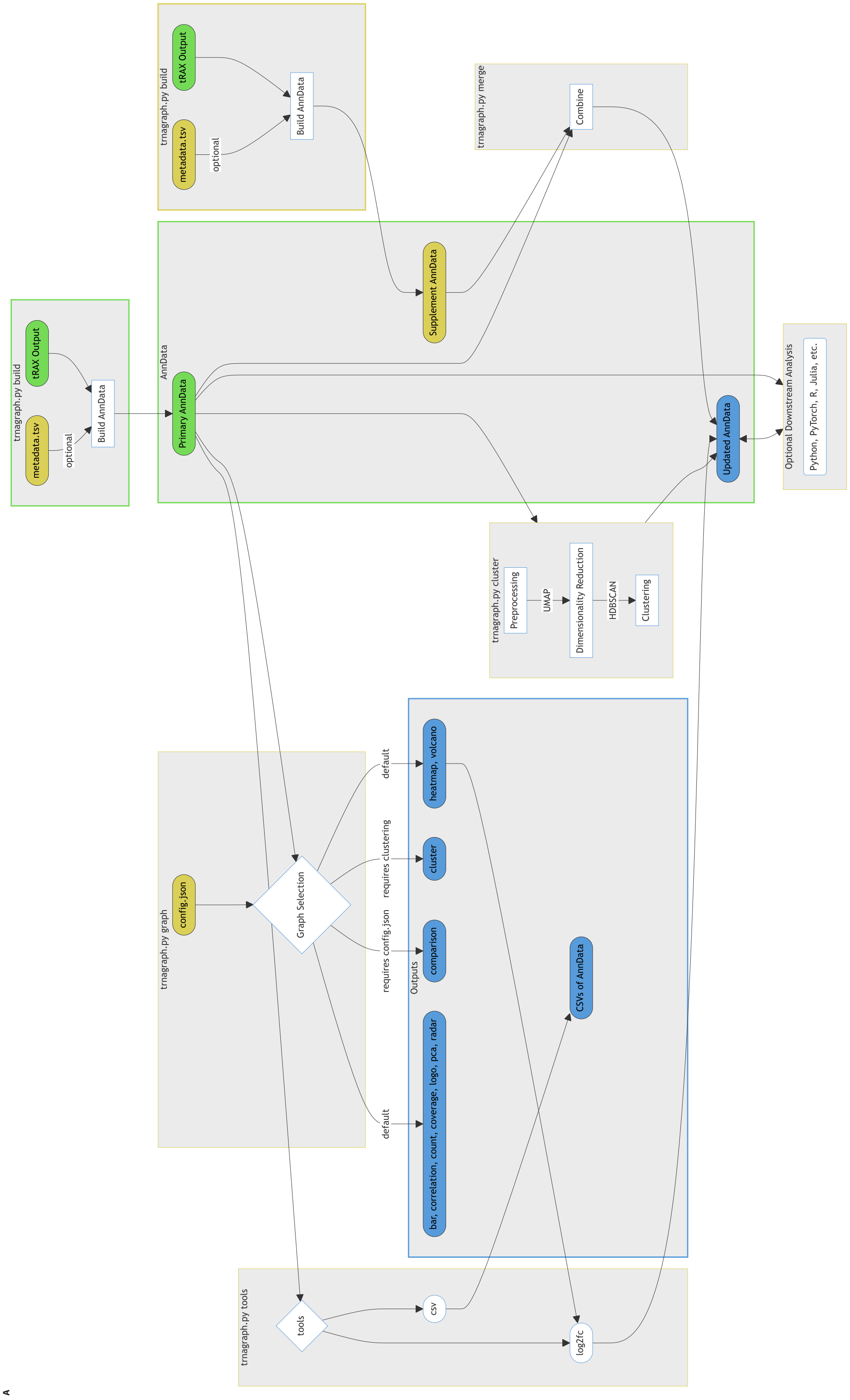
(A-B) Violin plots of the tRNA genomic loci found in (A) 1q23.3 TRU and (B) Neural Expansion Regions across the T2T primates. The dotted line at score 50 split tRNAs into predicted tRNAs and likely pseudogenes based on the tRNAscan-SE score.



Supplemental Figure S2.5

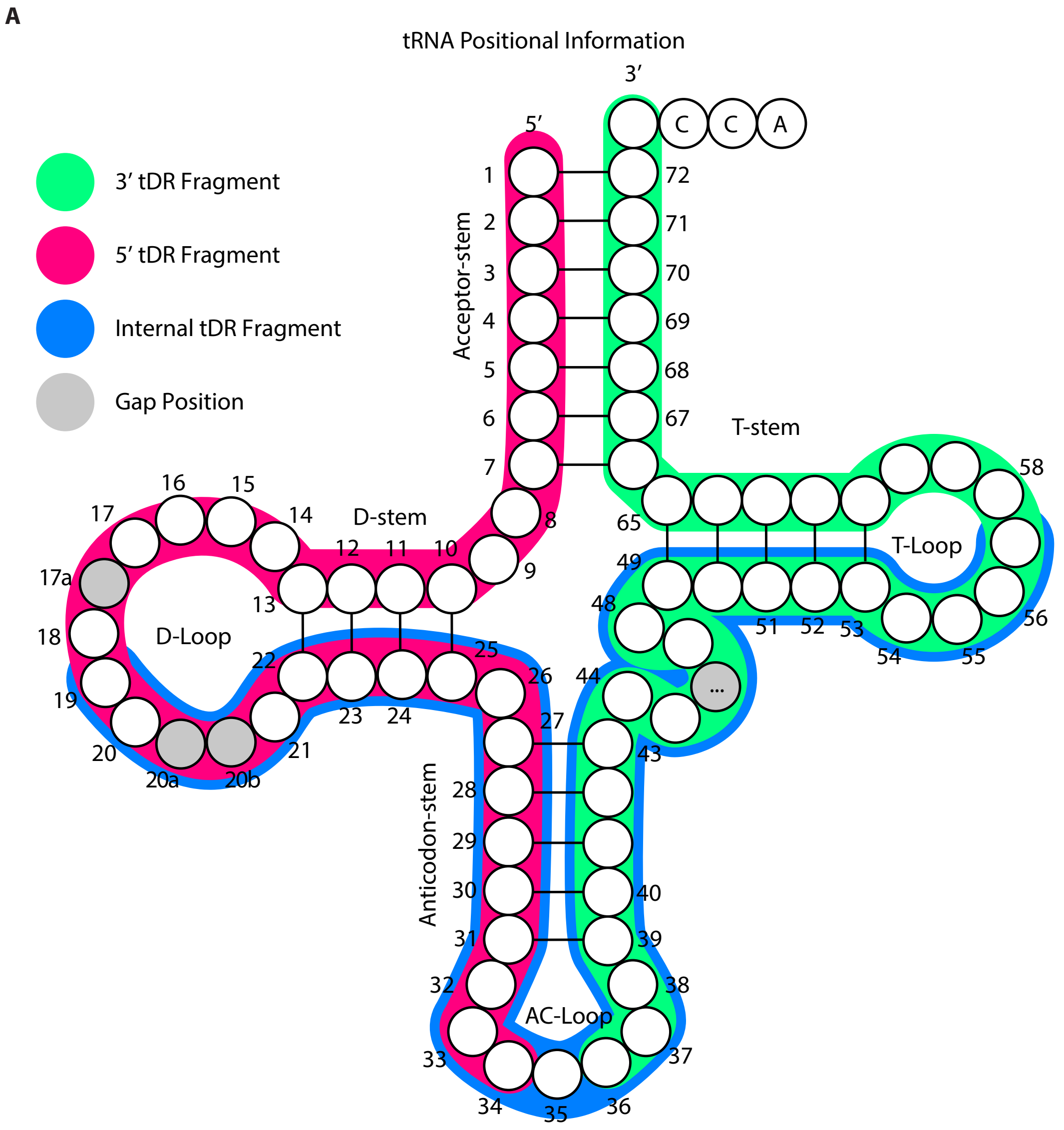
(A) Barplot of the average tRNA secondary structure score in the neural expansion and the 1q23.3 regions.

Chapter 3: Supplemental Figures


A
trnagraph.py build
tRAX Output
metadata.tsv
optional
Build AnnData
trnagraph.py build
tRAX Output
metadata.tsv
optional
Build AnnData
AnnData
Primary AnnData
Supplement AnnData
Updated AnnData
trnagraph.py merge
Combine
Optional Downstream Analysis
Python, PyTorch, R, Julia, etc.
trnagraph.py cluster
Preprocessing
UMAP
Dimensionality Reduction
HDBSCAN
Clustering
trnagraph.py graph
config.json
Graph Selection
default
requires clustering
requires config.json
default
Outputs
heatmap, volcano
cluster
comparison
bar, correlation, count, coverage, logo, pca, radar
CSVs of AnnData
trnagraph.py tools
tools
csv
log2fc

Supplemental Figure S3.1: tRNAgraph pipeline

(A) Detailed schematics of the tRNAgraph pipeline. The initial tRAX output directory is shown in green, and optional metadata and config files are shown in gold. Blue represents outputs from the pipeline either as visualizations or updates/creations of the hd5f AnnData file.



Supplemental Figure S3.2

(A) Detailed schematic of tRNA structure with positional information, including gap/extensions, 3′/5′, and internal fragment ranges.

Chapter 1: Supplemental Tables

Supplemental Table S1.1: Differential Expression of Small RNAs

List of small RNAs picked up in the ARM-seq data, with associated read counts, log2 fold-changes and p-values for developmental time points.

1.supp.1.csv

Supplemental Table S1.2: ARM-seq Differential Expression of tRNAs

List of tRNAs and tDRs picked up in the ARM-seq data, with associated read counts, log2 fold-changes and p-values for developmental time points of AlkB+ and AlkB- treatments.

1.supp.2.csv

Supplemental Table S1.3: tRNAs Classes and tRNAs therein

List of tRNA clusters that were manually annotated from UMAP/hdbscan clustering. Associated metadata, isotype, amino group, organoid developmental time point, alkB treatment, number of read counts, reference tRNA gene.

1.supp.3.csv

Primary	Dilution	Secondary	Dilution
ctip2 (ab18465)	1:1000	rat (ab150155) anti-donkey	1:250
tbr2/eomes (ab23345)	1:1000	rb (A10040) anti-donkey	1:250
pax6 (AB_528427)	1:1000	ms (ab150105)	1:250
sox2 (sc-398254)	1:1000	ms (ab150105)	1:250
vim (NB300-223SS)	1:1000	ch (ab150169)	1:1000
hopx (NBP1-97503)	1:1000	ms (ab150105)	1:1000
satb2 (ab34735)	1:1000	rb (A10040)	1:1000

Supplemental Table S1.4: IF-Staining Antibodies

List of primary and secondary antibodies concentrations used for IF-staining.

Chapter 2: Supplemental Tables

Supplemental Table S2.1: tRNA Sequence Divergence HS1-HG38

List of all syntenic tRNAs that have divergent sequences between hg38 and hs1 assemblies, their assigned gtRNAdb id, tRNAscan-SE score, hg38 gtRNAdb id alias if it changed, and notes about the tRNA.

2.supp.1.csv

Species	Assembly	High Con.	Total
Human	hs1	32	76
Chimpanzee	mPanTro3	36	70
Bonobo	mPanPan1	34	64
Gorilla	mGorGor1	31	64
Bornean orangutan	mPonPyg2	25	55
Sumatran orangutan	mPonAbe1	25	58
Siamang gibbon	mSymSyn1	27	56

Supplemental Table S2.2: tRNA NER Primate Counts

List of tRNAs in the NER by primate.

Supplemental Table S2.3: Samples and Metadata

List of all included genomes, their URL locations, accession codes for GenBank, and whether they are a primary, supplementary, or alternative haploid assembly.

2.supp.3.csv

Supplemental Table S2.4: tRNA Name Map

List of hs1 tRNAscan-SE names and assigned gtRNAdb IDs aligned to hg38 and all T2T primates.

2.supp.4.csv

Chapter 3: Supplemental Tables

Supplemental Table S3.1: Clustering results of combined Primate OTTR-seq

List of tRNA clusters that were identified from UMAP/HDBSCAN clustering. Associated metadata, isotype, amino group, organoid developmental time point, alkB treatment, number of read counts, reference tRNA gene.

3.supp.1.csv

Chapter 1: Data Accesss

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE259250. All visualizations were created using tRNAgraph (https://github.com/alba1735/tRNAgraph) and custom Python scripts, available upon request.

Chapter 2: Data Accesss

The FASTA files for assemblies and NCBI RefSeq genes can be located via NCBI RefSeq Accessions: hs1 (GCF_009914755.1), mGorGor1 (GCF_029281585.2), mPanTro3 (GCF_028858775.2), mPanPan1 (GCF_029289425.2), mPonAbe1 (GCF_028885655.2), mPonPyg2 (GCF_028885625.2), and mSymSyn1 (GCF_028878055.2).

Chapter 3: Data Accesss

Data available upon request.

Appendix A

Glossary

1p36: Human Chromosome 1 neural region of interest.

1p11: Human P-arm region proximal to the centromere that contains NOTCH2 and NOTCH2NLR that comprises part the "Neural expansion region" (NER).

1q21: Human Q-arm region proximal to the centromere that contains NOTCH2NLA/B/C that comprises part of the "Neural expansion region" (NER).

1q23: tRNA expansion region found in human chromosome 1 and syntenic region in mammals.

AnnData: Annotated Data file type, a type of Hierarchical Data Formats (HDF) files used by tRNAgraph.

gtRNAdb: Genomic tRNA Database

Isoacceptor: tRNA molecules which are acylated by the same amino acid but have divergent anticodons.

Isodecoder: tRNA molecules sharing the same anticodon but diverging elsewhere in their sequence.

Isotype: tRNA molecules which are acylated by the same amino acid.

MST: Minimum spanning tree

NBPF: Neuroblastoma break-point family genes, commonly found in the "Neural expansion region" and proximal to many tRNAs.

NER: Neural expansion region found across primates expanding greatly over the human-primate lineage.

NOTCH2: Notch2, a member of the Notch family of receptors.

NOTCH2NL: A paralog of the NOTCH2 receptor, with the ability to promote cortical progenitor maintenance (Fiddes, Lodewijk, et al. 2018).

SINE: Short interspersed nuclear element

tRAX: tRNA Analysis of eXpression (http://trna.ucsc.edu/tRAX/) (Holmes et al. 2022)

tRNA: Transfer RNA

tRNAgraph: A software pipeline for advanced and multivariate analysis of tRNA sequencing data.

tRNAid: Identification of unique tRNA sequences from tRNAmap

tRNAmap: A database and webtool for tRNA sequence graphical relationships across Eukaryotes.

tRNAnome: The collection of all tRNA sequences found across all species.

tDR: Transfer RNA-derived small

References

Altschul, S. F. et al. (Oct. 1990). “Basic local alignment search tool”. eng. In: *Journal of Molecular Biology* 215.3, pp. 403–410. ISSN: 0022-2836. DOI: 10.1016/S0022-2836(05)80360-2.

Armstrong, Joel et al. (Nov. 2020). “Progressive Cactus is a multiple-genome aligner for the thousand-genome era”. en. In: *Nature* 587.7833. Number: 7833 Publisher: Nature Publishing Group, pp. 246–251. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2871-y. URL: https://www.nature.com/articles/s41586-020-2871-y (visited on 12/14/2023).

Avcilar-Kucukgoze, Irem and Anna Kashina (Dec. 2020). “Hijacking tRNAs From Translation: Regulatory Functions of tRNAs in Mammalian Cell Physiology”. In: *Frontiers in Molecular Biosciences* 7, p. 610617. ISSN: 2296-889X. DOI: 10.3389/fmolb.2020.610617. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7773854/ (visited on 07/19/2021).

Begley, Ulrike et al. (Dec. 2007). “Trm9-catalyzed tRNA modifications link translation to the DNA damage response”. eng. In: *Molecular Cell* 28.5, pp. 860–870. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2007.09.021.

Berg, O. G. and C. G. Kurland (July 1997). “Growth rate-optimised tRNA abundance and codon usage”. eng. In: *Journal of Molecular Biology* 270.4, pp. 544–550. ISSN: 0022-2836. DOI: 10.1006/jmbi.1997.1142.

Bermudez-Santana, Clara et al. (Apr. 2010). “Genomic organization of eukaryotic tRNAs”. eng. In: *BMC genomics* 11, p. 270. ISSN: 1471-2164. DOI: 10.1186/1471-2164-11-270.

Bratkovič, Tomaž et al. (Mar. 2018). “Neuronal differentiation induces SNORD115 expression and is accompanied by post-transcriptional changes of serotonin receptor 2c mRNA”. en. In: *Scientific Reports* 8.1. Number: 1 Publisher: Nature Publishing Group, p. 5101. ISSN: 2045-2322. DOI: 10.1038/s41598-018-23293-7. URL: https://www.nature.com/articles/s41598-018-23293-7 (visited on 02/19/2021).

Brunetti-Pierri, Nicola et al. (Dec. 2008). “Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities”. en. In: *Nature Genetics* 40.12, pp. 1466–1471. ISSN: 1061-4036, 1546-1718. DOI: 10.1038/ng.279. URL: http://www.nature.com/articles/ng.279 (visited on 03/10/2020).

Camacho, Christiam et al. (Dec. 2009). “BLAST+: architecture and applications”. eng. In: *BMC bioinformatics* 10, p. 421. ISSN: 1471-2105. DOI: 10.1186/1471-2105-10-421.

Camp, J. Gray et al. (Dec. 2015). “Human cerebral organoids recapitulate gene expression programs of fetal neocortex development”. en. In: *Proceedings of the National Academy of Sciences* 112.51. Publisher: National Academy of Sciences Section: Biological Sciences, pp. 15672–15677. ISSN:

0027-8424, 1091-6490. DOI: 10.1073/pnas.1520760112. URL: https://www.pnas.org/content/112/51/15672 (visited on 05/28/2020).

Chaisson, Mark J. P. et al. (Jan. 2015). "Resolving the complexity of the human genome using single-molecule sequencing". eng. In: *Nature* 517.7536, pp. 608–611. ISSN: 1476-4687. DOI: 10.1038/nature13907.

Chan, Patricia P., Brian Y. Lin, et al. (Apr. 2019). "tRNAscan-SE 2.0: Improved Detection and Functional Classification of Transfer RNA Genes". en. In: *bioRxiv*, p. 614032. DOI: 10.1101/614032. URL: https://www.biorxiv.org/content/10.1101/614032v1 (visited on 03/03/2020).

Chan, Patricia P. and Todd M. Lowe (Jan. 2016). "GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes". In: *Nucleic Acids Research* 44.Database issue, pp. D184–D189. ISSN: 0305-1048. DOI: 10.1093/nar/gkv1309. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4702915/ (visited on 01/17/2020).

Chen, J. and J. R. Patton (Oct. 2000). "Pseudouridine synthase 3 from mouse modifies the anticodon loop of tRNA". eng. In: *Biochemistry* 39.41, pp. 12723–12730. ISSN: 0006-2960. DOI: 10.1021/bi001109m.

Cheng, Yu-Heng et al. (May 2019). "Hydro-Seq enables contamination-free high-throughput single-cell RNA-sequencing for circulating tumor cells". en. In: *Nature Communications* 10.1, p. 2163. ISSN: 2041-1723. DOI: 10.1038/s41467-019-10122-2. URL: https://www.nature.com/articles/s41467-019-10122-2 (visited on 09/22/2021).

Chujo, Takeshi and Kazuhito Tomizawa (Jan. 2021). "Human transfer RNA modopathies: diseases caused by aberrations in transfer RNA modifications". eng. In: *The FEBS journal*. ISSN: 1742-4658. DOI: 10.1111/febs.15736.

Cimadamore, Flavio et al. (Aug. 2013). "SOX2–LIN28/let-7 pathway regulates proliferation and neurogenesis in neural precursors". en. In: *Proceedings of the National Academy of Sciences* 110.32. Publisher: National Academy of Sciences Section: PNAS Plus, E3017–E3026. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1220176110. URL: https://www.pnas.org/content/110/32/E3017 (visited on 08/03/2021).

Coolen, Marion, Shauna Katz, and Laure Bally-Cuif (2013). "miR-9: a versatile regulator of neurogenesis". English. In: *Frontiers in Cellular Neuroscience* 0. Publisher: Frontiers. ISSN: 1662-5102. DOI: 10.3389/fncel.2013.00220. URL: https://www.frontiersin.org/articles/10.3389/fncel.2013.00220/full (visited on 08/03/2021).

Cozen, Aaron E. et al. (Sept. 2015). “ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments”. eng. In: *Nature Methods* 12.9, pp. 879–884. ISSN: 1548-7105. DOI: 10.1038/nmeth.3508.

Crécy-Lagard, Valérie de et al. (Mar. 2019). “Matching tRNA modifications in humans to their known and predicted enzymes”. en. In: *Nucleic Acids Research* 47.5. Publisher: Oxford Academic, pp. 2143–2159. ISSN: 0305-1048. DOI: 10.1093/nar/gkz011. URL: https://academic.oup.com/nar/article/47/5/2143/5304329 (visited on 03/04/2020).

Dang, Hairuo and Elmar Schiebel (Oct. 2022). “Emerging roles of centrosome cohesion”. In: *Open Biology* 12.10, p. 220229. ISSN: 2046-2441. DOI: 10.1098/rsob.220229. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9597181/ (visited on 05/09/2024).

Darrow, Emily M. and Brian P. Chadwick (June 2014). “A novel tRNA variable number tandem repeat at human chromosome 1q23.3 is implicated as a boundary element based on conservation of a CTCF motif in mouse”. In: *Nucleic Acids Research* 42.10, pp. 6421–6435. ISSN: 0305-1048. DOI: 10.1093/nar/gku280. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4041453/ (visited on 05/01/2024).

Dou, Shengqian, Yirong Wang, and Jian Lu (Feb. 2019). “Metazoan tsRNAs: Biogenesis, Evolution and Regulatory Functions”. eng. In: *Non-coding RNA* 5.1. ISSN: 2311-553X. DOI: 10.3390/ncrna5010018.

Duret, L. (July 2000). “tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes”. eng. In: *Trends in genetics: TIG* 16.7, pp. 287–289. ISSN: 0168-9525. DOI: 10.1016/s0168-9525(00)02041-2.

Eiraku, Mototsugu et al. (Nov. 2008). “Self-organized formation of polarized cortical tissues from ESCs and its active manipulation by extrinsic signals”. eng. In: *Cell Stem Cell* 3.5, pp. 519–532. ISSN: 1875-9777. DOI: 10.1016/j.stem.2008.09.002.

Fairchild, Corinne L. A. et al. (Oct. 2019). “Let-7 regulates cell cycle dynamics in the developing cerebral cortex and retina”. en. In: *Scientific Reports* 9.1. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Cell proliferation;Developmental neurogenesis Subject_term_id: cell-proliferation;developmental-neurogenesis, p. 15336. ISSN: 2045-2322. DOI: 10.1038/s41598-019-51703-x. URL: https://www.nature.com/articles/s41598-019-51703-x (visited on 08/03/2021).

Fiddes, Ian T., Joel Armstrong, et al. (July 2018). “Comparative Annotation Toolkit (CAT)—simultaneous clade and personal genome annotation”. In: *Genome Research* 28.7, pp. 1029–1038. ISSN: 1088-9051. DOI: 10.1101/gr.233460.117. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6028123/ (visited on 05/04/2024).

Fiddes, Ian T., Gerrald A. Lodewijk, et al. (May 2018). “Human-Specific NOTCH2NL Genes Affect Notch Signaling and Cortical Neurogenesis”. English. In: *Cell* 173.6, 1356–1369.e22. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2018.03.051. URL: https://www.cell.com/cell/abstract/S0092-8674(18)30383-0 (visited on 03/03/2020).

Fiddes, Ian T., Alex A. Pollen, et al. (July 2019). “Paired involvement of human-specific Olduvai domains and NOTCH2NL genes in human brain evolution”. en. In: *Human Genetics* 138.7, pp. 715–721. ISSN: 1432-1203. DOI: 10.1007/s00439-019-02018-4. URL: https://doi.org/10.1007/s00439-019-02018-4 (visited on 08/12/2019).

Goodarzi, Hani, Xuhang Liu, et al. (May 2015). “Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement”. eng. In: *Cell* 161.4, pp. 790–802. ISSN: 1097-4172. DOI: 10.1016/j.cell.2015.02.053.

Goodarzi, Hani, Hoang C. B. Nguyen, et al. (June 2016). “Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression”. eng. In: *Cell* 165.6, pp. 1416–1427. ISSN: 1097-4172. DOI: 10.1016/j.cell.2016.05.046.

Guimarães, Ana Rita et al. (Mar. 2021). “tRNAs as a Driving Force of Genome Evolution in Yeast”. English. In: *Frontiers in Microbiology* 12. Publisher: Frontiers. ISSN: 1664-302X. DOI: 10.3389/fmicb.2021.634004. URL: https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2021.634004/full (visited on 04/27/2024).

Gustafsson, H. Tobias et al. (Feb. 2022). *Deep sequencing of yeast and mouse tRNAs and tRNA fragments using OTTR*. en. Pages: 2022.02.04.479139 Section: New Results. DOI: 10.1101/2022.02.04.479139. URL: https://www.biorxiv.org/content/10.1101/2022.02.04.479139v1 (visited on 05/17/2024).

Hagberg, Aric, Pieter Swart, and Daniel Chult (Jan. 2008). “Exploring Network Structure, Dynamics, and Function Using NetworkX”. In.

Harris, Charles R. et al. (Sept. 2020). “Array programming with NumPy”. en. In: *Nature* 585.7825. Publisher: Nature Publishing Group, pp. 357–362. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2649-2. URL: https://www.nature.com/articles/s41586-020-2649-2 (visited on 05/04/2024).

Hashimoto, Y., T. Niikura, Y. Ito, et al. (Dec. 2001). “Detailed characterization of neuroprotection by a rescue factor humanin against various Alzheimer’s disease-relevant insults”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 21.23, pp. 9235–9245. ISSN: 1529-2401.

Hashimoto, Y., T. Niikura, H. Tajima, et al. (May 2001). “A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer’s disease genes and Abeta”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 98.11, pp. 6336–6341. ISSN: 0027-8424. DOI: 10.1073/pnas.101133498.

Haussecker, Dirk et al. (Apr. 2010). “Human tRNA-derived small RNAs in the global regulation of RNA silencing”. eng. In: *RNA (New York, N.Y.)* 16.4, pp. 673–695. ISSN: 1469-9001. DOI: 10.1261/rna.2000810.

Higgs, Paul G. and Wenqi Ran (Nov. 2008). “Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage”. eng. In: *Molecular Biology and Evolution* 25.11, pp. 2279–2291. ISSN: 1537-1719. DOI: 10.1093/molbev/msn173.

Holmes, Andrew D. et al. (July 2022). *tRNA Analysis of eXpression (tRAX): A tool for integrating analysis of tRNAs, tRNA-derived small RNAs, and tRNA modifications*. en. Pages: 2022.07.02.498565 Section: New Results. DOI: 10.1101/2022.07.02.498565. URL: https://www.biorxiv.org/content/10.1101/2022.07.02.498565v1 (visited on 06/05/2023).

Iben, James R. and Richard J. Maraia (July 2012). “tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote”. eng. In: *RNA (New York, N.Y.)* 18.7, pp. 1358–1372. ISSN: 1469-9001. DOI: 10.1261/rna.032151.111.

— (Feb. 2014). “tRNA gene copy number variation in humans”. en. In: *Gene* 536.2, pp. 376–384. ISSN: 0378-1119. DOI: 10.1016/j.gene.2013.11.049. URL: http://www.sciencedirect.com/science/article/pii/S0378111913015758 (visited on 03/04/2020).

Ishimura, Ryuta et al. (July 2014). “Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration”. en. In: *Science* 345.6195. Publisher: American Association for the Advancement of Science Section: Report, pp. 455–459. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1249749. URL: https://science.sciencemag.org/content/345/6195/455 (visited on 04/15/2020).

Ivanov, Pavel et al. (Dec. 2014). “G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 111.51, pp. 18201–18206. ISSN: 1091-6490. DOI: 10.1073/pnas.1407361111.

Jiang, Yongpeng et al. (Nov. 2022). “Cross-regulome profiling of RNA polymerases highlights the regulatory role of polymerase III on mRNA transcription by maintaining local chromatin architecture”. In: *Genome Biology* 23.1, p. 246. ISSN: 1474-760X. DOI: 10.1186/s13059-022-02812-w. URL: https://doi.org/10.1186/s13059-022-02812-w (visited on 05/11/2024).

Kaçar, Yasemin, Hildburg Beier, and Hans J. Gross (Apr. 1995). “The presence of tRNA pseudogenes in mammalia and plants and their absence in yeast may account for different specificities of pre-tRNA processing enzymes”. In: *Gene* 156.1, pp. 129–132. ISSN: 0378-1119. DOI: 10.1016/0378-1119(95)00079-L. URL: https://www.sciencedirect.com/science/article/pii/037811199500079L (visited on 04/26/2024).

Kadoshima, Taisuke et al. (Dec. 2013). “Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell–derived neocortex”. en. In: *Proceedings of the National Academy of Sciences* 110.50. ISBN: 9781315710112 Publisher: National Academy of Sciences Section: Biological Sciences, pp. 20284–20289. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1315710110. URL: https://www.pnas.org/content/110/50/20284 (visited on 05/04/2020).

Kapur, Mridu et al. (Oct. 2020). “Expression of the Neuronal tRNA n-Tr20 Regulates Synaptic Transmission and Seizure Susceptibility”. eng. In: *Neuron* 108.1, 193–208.e9. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2020.07.023.

Kelava, Iva and Madeline A. Lancaster (Dec. 2016). “Dishing out mini-brains: Current progress and future prospects in brain organoid research”. eng. In: *Developmental Biology* 420.2, pp. 199–209. ISSN: 1095-564X. DOI: 10.1016/j.ydbio.2016.06.037.

Kindberg, Abigail A. et al. (Dec. 2014). “An in vitro model of human neocortical development using pluripotent stem cells: cocaine-induced cytoarchitectural alterations”. In: *Disease Models & Mechanisms* 7.12, pp. 1397–1405. ISSN: 1754-8403. DOI: 10.1242/dmm.017251. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4257008/ (visited on 05/27/2020).

Kluyver, Thomas et al. (2016). “Jupyter Notebooks – a publishing format for reproducible computational workflows”. en. In: ed. by Fernando Loizides and Birgit Scmidt. IOS Press, pp. 87–90. DOI: 10.3233/978-1-61499-649-1-87. URL: https://eprints.soton.ac.uk/403913/ (visited on 07/07/2021).

Kruskal, Joseph B. (1956). “On the shortest spanning subtree of a graph and the traveling salesman problem”. en. In: *Proceedings of the American Mathematical Society* 7.1, pp. 48–50. ISSN: 0002-9939, 1088-6826. DOI: 10.1090/S0002-9939-1956-0078686-7. URL: https://www.ams.org/proc/1956-007-01/S0002-9939-1956-0078686-7/ (visited on 05/14/2024).

Kuhle, Bernhard, Qi Chen, and Paul Schimmel (Nov. 2023). “tRNA renovatio: Rebirth through fragmentation”. English. In: *Molecular Cell* 83.22. Publisher: Elsevier, pp. 3953–3971. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2023.09.016. URL: https://www.cell.com/molecular-cell/abstract/S1097-2765(23)00739-6 (visited on 01/29/2024).

Kundaje, Anshul et al. (Feb. 2015). “Integrative analysis of 111 reference human epigenomes”. en. In: *Nature* 518.7539, pp. 317–330. ISSN: 1476-4687. DOI: 10.1038/nature14248. URL: https://www.nature.com/articles/nature14248 (visited on 01/17/2020).

Kutter, Claudia et al. (Oct. 2011). “Pol III binding in six mammals shows conservation among amino acid isotypes despite divergence among tRNA genes”. en. In: *Nature Genetics* 43.10. Publisher: Nature Publishing Group, pp. 948–955. ISSN: 1546-1718. DOI: 10.1038/ng.906. URL: https://www.nature.com/articles/ng.906 (visited on 06/07/2024).

Lassaunière, Ria and Caroline T. Tiemessen (Dec. 2021). “FcγR Genetic Variation and HIV-1 Vaccine Efficacy: Context And Considerations”. In: *Frontiers in Immunology* 12, p. 788203. ISSN: 1664-3224. DOI: 10.3389/fimmu.2021.788203. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8714752/ (visited on 06/06/2024).

Lejeune, Julien, Guillaume Brachet, and Hervé Watier (2019). “Evolutionary Story of the Low/Medium-Affinity IgG Fc Receptor Gene Cluster”. eng. In: *Frontiers in Immunology* 10, p. 1297. ISSN: 1664-3224. DOI: 10.3389/fimmu.2019.01297.

Letzring, Daniel P., Kimberly M. Dean, and Elizabeth J. Grayhack (Dec. 2010). “Control of translation efficiency in yeast by codon-anticodon interactions”. eng. In: *RNA (New York, N.Y.)* 16.12, pp. 2516–2528. ISSN: 1469-9001. DOI: 10.1261/rna.2411710.

Li, Guoping et al. (2022). “Distinct Stress-Dependent Signatures of Cellular and Extracellular tRNA-Derived Small RNAs”. en. In: *Advanced Science* 9.17, p. 2200829. ISSN: 2198-3844. DOI: 10.1002/advs.202200829. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/advs.202200829 (visited on 08/15/2023).

Li, Lei et al. (Aug. 2022). “Oncogene or tumor suppressor gene: An integrated pan-cancer analysis of NBPF1”. In: *Frontiers in Endocrinology* 13, p. 950326. ISSN: 1664-2392. DOI: 10.3389/fendo.2022.950326. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9428449/ (visited on 05/09/2024).

Li, Xinrui et al. (July 2009). “Fcγ Receptors: Structure, Function and Role as Genetic Risk Factors in SLE”. In: *Genes and immunity* 10.5, pp. 380–389. ISSN: 1466-4879. DOI: 10.1038/gene.2009.35. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2830794/ (visited on 06/06/2024).

Love, Michael I., Wolfgang Huber, and Simon Anders (Dec. 2014). “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. In: *Genome Biology* 15.12, p. 550. ISSN: 1474-760X. DOI: 10.1186/s13059-014-0550-8. URL: https://doi.org/10.1186/s13059-014-0550-8 (visited on 04/14/2020).

Lowe, T. M. and S. R. Eddy (Mar. 1997). “tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence”. eng. In: *Nucleic Acids Research* 25.5, pp. 955–964. ISSN: 0305-1048. DOI: 10.1093/nar/25.5.955.

Lowe, Todd M. and Patricia P. Chan (2016). “tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes”. eng. In: *Nucleic Acids Research* 44.W1, W54–57. ISSN: 1362-4962. DOI: 10.1093/nar/gkw413.

Lyons, Shawn M., Marta M. Fay, and Pavel Ivanov (2018). “The role of RNA modifications in the regulation of tRNA cleavage”. en. In: *FEBS Letters* 592.17, pp. 2828–2844. ISSN: 1873-3468. DOI: 10.1002/1873-3468.13205. URL: https://febs.onlinelibrary.wiley.com/doi/abs/10.1002/1873-3468.13205 (visited on 04/15/2020).

Mabuchi, Kohji et al. (Sept. 2004). “Gene rearrangements and evolution of tRNA pseudogenes in the mitochondrial genome of the parrotfish (Teleostei: Perciformes: Scaridae)”. eng. In: *Journal of Molecular Evolution* 59.3, pp. 287–297. ISSN: 0022-2844. DOI: 10.1007/s00239-004-2621-z.

Mader, Marius Marc-Daniel and Patrick Czorlich (Dec. 2021). “The role of L-arginine metabolism in neurocritical care patients”. In: *Neural Regeneration Research* 17.7, pp. 1446–1453. ISSN: 1673-5374. DOI: 10.4103/1673-5374.327331. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8771107/ (visited on 11/02/2023).

Makova, Kateryna D. et al. (Dec. 2023). *The Complete Sequence and Comparative Analysis of Ape Sex Chromosomes*. en. Pages: 2023.11.30.569198 Section: New Results. DOI: 10.1101/2023.11.30.569198. URL: https://www.biorxiv.org/content/10.1101/2023.11.30.569198v1 (visited on 12/14/2023).

McInnes, Leland, John Healy, and Steve Astels (Mar. 2017). “hdbscan: Hierarchical density based clustering”. en. In: *Journal of Open Source Software* 2.11, p. 205. ISSN: 2475-9066. DOI: 10.21105/joss.00205. URL: https://joss.theoj.org/papers/10.21105/joss.00205 (visited on 01/27/2024).

McInnes, Leland, John Healy, and James Melville (Dec. 2018). “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction”. In: *arXiv:1802.03426 [cs, stat]*. arXiv: 1802.03426. URL: http://arxiv.org/abs/1802.03426 (visited on 02/24/2020).

McKinney, Wes (Jan. 2010). “Data Structures for Statistical Computing in Python”. In: pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a.

Mefford, Heather C. et al. (Oct. 2008). “Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes”. eng. In: *The New England Journal of Medicine* 359.16, pp. 1685–1699. ISSN: 1533-4406. DOI: `10.1056/NEJMoa0805384`.

Morisaki, Ikuko et al. (July 2021). “Modeling a human CLP1 mutation in mouse identifies an accumulation of tyrosine pre-tRNA fragments causing pontocerebellar hypoplasia type 10”. eng. In: *Biochemical and Biophysical Research Communications* 570, pp. 60–66. ISSN: 1090-2104. DOI: `10.1016/j.bbrc.2021.07.036`.

Nassar, Luis R. et al. (Jan. 2023). “The UCSC Genome Browser database: 2023 update”. eng. In: *Nucleic Acids Research* 51.D1, pp. D1188–D1195. ISSN: 1362-4962. DOI: `10.1093/nar/gkac1072`.

Nawrocki, Eric P. and Sean R. Eddy (Nov. 2013). “Infernal 1.1: 100-fold faster RNA homology searches”. eng. In: *Bioinformatics (Oxford, England)* 29.22, pp. 2933–2935. ISSN: 1367-4811. DOI: `10.1093/bioinformatics/btt509`.

Nie, Anzheng et al. (Nov. 2019). “Roles of aminoacyl-tRNA synthetases in immune regulation and immune diseases”. en. In: *Cell Death & Disease* 10.12, pp. 1–14. ISSN: 2041-4889. DOI: `10.1038/s41419-019-2145-5`. URL: `https://www.nature.com/articles/s41419-019-2145-5` (visited on 12/05/2019).

Novoa, Eva Maria et al. (Mar. 2012). “A role for tRNA modifications in genome structure and codon usage”. eng. In: *Cell* 149.1, pp. 202–213. ISSN: 1097-4172. DOI: `10.1016/j.cell.2012.01.050`.

Numanagic, Ibrahim et al. (Sept. 2018). “Fast characterization of segmental duplications in genome assemblies”. eng. In: *Bioinformatics (Oxford, England)* 34.17, pp. i706–i714. ISSN: 1367-4811. DOI: `10.1093/bioinformatics/bty586`.

Nurk, Sergey et al. (Apr. 2022). “The complete sequence of a human genome”. In: *Science* 376.6588. Publisher: American Association for the Advancement of Science, pp. 44–53. DOI: `10.1126/science.abj6987`. URL: `https://www.science.org/doi/10.1126/science.abj6987` (visited on 05/01/2024).

Palazzo, Alexander F. and Eliza S. Lee (2015). “Non-coding RNA: what is functional and what is junk?” eng. In: *Frontiers in Genetics* 6, p. 2. ISSN: 1664-8021. DOI: `10.3389/fgene.2015.00002`.

Parisien, Marc, Xiaoyun Wang, and Tao Pan (Dec. 2013). “Diversity of human tRNA genes from the 1000-genomes project”. eng. In: *RNA biology* 10.12, pp. 1853–1867. ISSN: 1555-8584. DOI: `10.4161/rna.27361`.

Patil, Ashish et al. (Oct. 2012). “Increased tRNA modification and gene-specific codon usage regulate cell cycle progression during the DNA damage response”. eng. In: *Cell Cycle (Georgetown, Tex.)* 11.19, pp. 3656–3665. ISSN: 1551-4005. DOI: 10.4161/cc.21919.

Pechmann, Sebastian and Judith Frydman (Feb. 2013). “Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding”. eng. In: *Nature Structural & Molecular Biology* 20.2, pp. 237–243. ISSN: 1545-9985. DOI: 10.1038/nsmb.2466.

Phizicky, Eric M. and Anita K. Hopper (Sept. 2010). “tRNA biology charges to the front”. eng. In: *Genes & Development* 24.17, pp. 1832–1860. ISSN: 1549-5477. DOI: 10.1101/gad.1956510.

Pinkard, Otis et al. (Aug. 2020). “Quantitative tRNA-sequencing uncovers metazoan tissue-specific tRNA regulation”. en. In: *Nature Communications* 11.1. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Next-generation sequencing;tRNAs Subject_term_id: next-generation-sequencing;trnas, p. 4104. ISSN: 2041-1723. DOI: 10.1038/s41467-020-17879-x. URL: https://www.nature.com/articles/s41467-020-17879-x (visited on 09/22/2021).

Pitts, Matthew W. et al. (Dec. 2014). “Selenoproteins in Nervous System Development and Function”. In: *Biological trace element research* 161.3, pp. 231–245. ISSN: 0163-4984. DOI: 10.1007/s12011-014-0060-2. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4222985/ (visited on 11/01/2023).

Plotkin, Joshua B. and Grzegorz Kudla (Jan. 2011). “Synonymous but not the same: the causes and consequences of codon bias”. eng. In: *Nature Reviews. Genetics* 12.1, pp. 32–42. ISSN: 1471-0064. DOI: 10.1038/nrg2899.

Pollen, Alex A. et al. (Feb. 2019). “Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution”. English. In: *Cell* 176.4, 743–756.e17. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2019.01.017. URL: https://www.cell.com/cell/abstract/S0092-8674(19)30050-9 (visited on 03/03/2020).

Poszewiecka, Barbara et al. (Aug. 2022). “Revised time estimation of the ancestral human chromosome 2 fusion”. In: *BMC Genomics* 23.6, p. 616. ISSN: 1471-2164. DOI: 10.1186/s12864-022-08828-7. URL: https://doi.org/10.1186/s12864-022-08828-7 (visited on 05/09/2024).

Qin, Chuan et al. (2020). “Pathological significance of tRNA-derived small RNAs in neurological disorders”. en. In: *Neural Regeneration Research* 15.2, p. 212. ISSN: 1673-5374. DOI: 10.4103/1673-5374.265560. URL: http://www.nrronline.org/text.asp?2020/15/2/212/265560 (visited on 03/03/2020).

Quinlan, Aaron R. and Ira M. Hall (Mar. 2010). “BEDTools: a flexible suite of utilities for comparing genomic features”. en. In: *Bioinformatics* 26.6. Publisher: Oxford Academic, pp. 841–842. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btq033. URL: https://academic.oup.com/bioinformatics/article/26/6/841/244688 (visited on 04/10/2020).

Radio, Francesca Clementina et al. (Mar. 2021). “SPEN haploinsufficiency causes a neurodevelopmental disorder overlapping proximal 1p36 deletion syndrome with an episignature of X chromosomes in females”. In: *The American Journal of Human Genetics* 108.3, pp. 502–516. ISSN: 0002-9297. DOI: 10.1016/j.ajhg.2021.01.015. URL: https://www.sciencedirect.com/science/article/pii/S000292972100015X (visited on 05/09/2024).

Schindelin, Johannes et al. (July 2012). “Fiji: an open-source platform for biological-image analysis”. en. In: *Nature Methods* 9.7. Number: 7 Publisher: Nature Publishing Group, pp. 676–682. ISSN: 1548-7105. DOI: 10.1038/nmeth.2019. URL: https://www.nature.com/articles/nmeth.2019 (visited on 05/30/2021).

Schneider, T D and R M Stephens (Oct. 1990). “Sequence logos: a new way to display consensus sequences.” In: *Nucleic Acids Research* 18.20, pp. 6097–6100. ISSN: 0305-1048. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC332411/ (visited on 07/20/2021).

Schneider, Valerie A. et al. (May 2017). “Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly”. eng. In: *Genome Research* 27.5, pp. 849–864. ISSN: 1549-5469. DOI: 10.1101/gr.213611.116.

Shaheen, Ranad et al. (July 2016). “A homozygous truncating mutation in PUS3 expands the role of tRNA modification in normal cognition”. In: *Human genetics* 135.7, pp. 707–713. ISSN: 0340-6717. DOI: 10.1007/s00439-016-1665-7. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5152754/ (visited on 01/29/2024).

Shigematsu, Megumi et al. (May 2017). “YAMAT-seq: an efficient method for high-throughput sequencing of mature transfer RNAs”. eng. In: *Nucleic Acids Research* 45.9, e70. ISSN: 1362-4962. DOI: 10.1093/nar/gkx005.

Shumate, Alaina and Steven L Salzberg (July 2021). “Liftoff: accurate mapping of gene annotations”. In: *Bioinformatics* 37.12, pp. 1639–1643. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btaa1016. URL: https://doi.org/10.1093/bioinformatics/btaa1016 (visited on 05/04/2024).

Sjöstedt, Evelina et al. (Mar. 2020). “An atlas of the protein-coding genes in the human, pig, and mouse brain”. In: *Science* 367.6482. Publisher: American Association for the Advancement of Science, eaay5947. DOI: 10.1126/science.aay5947. URL: https://www.science.org/doi/10.1126/science.aay5947 (visited on 06/05/2024).

Soares, Ana Raquel and Manuel Santos (2017). "Discovery and function of transfer RNA-derived fragments and their role in disease". en. In: *WIREs RNA* 8.5, e1423. ISSN: 1757-7012. DOI: 10.1002/wrna.1423. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wrna.1423 (visited on 04/09/2020).

Sprinzl, Mathias et al. (Jan. 1996). "Compilation of tRNA Sequences and Sequences of tRNA Genes". en. In: *Nucleic Acids Research* 24.1. Publisher: Oxford Academic, pp. 68–72. ISSN: 0305-1048. DOI: 10.1093/nar/24.1.68. URL: https://academic.oup.com/nar/article/24/1/68/2360954 (visited on 04/10/2020).

Su, Zhangli et al. (Dec. 2019). "tRNA-derived fragments and microRNAs in the maternal-fetal interface of a mouse maternal-immune-activation autism model". en. In: *bioRxiv*, p. 2019.12.20.884650. DOI: 10.1101/2019.12.20.884650. URL: https://www.biorxiv.org/content/10.1101/2019.12.20.884650v1 (visited on 01/16/2020).

Sun, Chunxiao et al. (Feb. 2018). "Roles of tRNA-derived fragments in human cancers". en. In: *Cancer Letters* 414, pp. 16–25. ISSN: 0304-3835. DOI: 10.1016/j.canlet.2017.10.031. URL: http://www.sciencedirect.com/science/article/pii/S0304383517306687 (visited on 05/06/2020).

Tareen, Ammar and Justin B. Kinney (Apr. 2020). "Logomaker: beautiful sequence logos in Python". eng. In: *Bioinformatics (Oxford, England)* 36.7, pp. 2272–2274. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btz921.

Thornlow, Bryan P. et al. (Dec. 2019). "Predicting transfer RNA gene activity from sequence and genome context". en. In: *Genome Research*, gr.256164.119. ISSN: 1088-9051, 1549-5469. DOI: 10.1101/gr.256164.119. URL: http://genome.cshlp.org/content/early/2019/12/19/gr.256164.119 (visited on 03/03/2020).

Tuorto, Francesca et al. (Sept. 2012). "RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis". eng. In: *Nature Structural & Molecular Biology* 19.9, pp. 900–905. ISSN: 1545-9985. DOI: 10.1038/nsmb.2357.

Ullah, Rahat et al. (Oct. 2020). "Glycine, the smallest amino acid, confers neuroprotection against d-galactose-induced neurodegeneration and memory impairment by regulating c-Jun N-terminal kinase in the mouse brain". In: *Journal of Neuroinflammation* 17.1, p. 303. ISSN: 1742-2094. DOI: 10.1186/s12974-020-01989-w. URL: https://doi.org/10.1186/s12974-020-01989-w (visited on 11/02/2023).

Upton, Heather E. et al. (Oct. 2021). "Low-bias ncRNA libraries using ordered two-template relay: Serial template jumping by a modified retroelement reverse transcriptase". In: *Proceedings of the National Academy of Sciences* 118.42, e2107900118. DOI: 10.1073/pnas.2107900118. URL: https://www.pnas.org/doi/10.1073/pnas.2107900118 (visited on 06/30/2022).

Virshup, Isaac et al. (Dec. 2021). *anndata: Annotated data*. en. Pages: 2021.12.16.473007 Section: New Results. DOI: 10.1101/2021.12.16.473007. URL: https://www.biorxiv.org/content/10.1101/2021.12.16.473007v1 (visited on 01/27/2024).

Vollger, Mitchell R. et al. (Apr. 2022). "Segmental duplications and their variation in a complete human genome". In: *Science* 376.6588. Publisher: American Association for the Advancement of Science, eabj6965. DOI: 10.1126/science.abj6965. URL: https://www.science.org/doi/10.1126/science.abj6965 (visited on 05/10/2024).

Waskom, Michael L. (Apr. 2021). "seaborn: statistical data visualization". en. In: *Journal of Open Source Software* 6.60, p. 3021. ISSN: 2475-9066. DOI: 10.21105/joss.03021. URL: https://joss.theoj.org/papers/10.21105/joss.03021 (visited on 05/04/2024).

Xu, Yao et al. (Mar. 2013). "Non-optimal codon usage is a mechanism to achieve circadian clock conditionality". eng. In: *Nature* 495.7439, pp. 116–120. ISSN: 1476-4687. DOI: 10.1038/nature11942.

Yang, Xiang-Lei (Dec. 2014). "tRNA-derived G-quadruplex protects motor neurons". en. In: *Proceedings of the National Academy of Sciences* 111.51, pp. 18108–18109. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1420838111. URL: https://www.pnas.org/content/111/51/18108 (visited on 02/18/2020).

Ying, Guoguang et al. (June 2004). "Humanin, a newly identified neuroprotective factor, uses the G protein-coupled formylpeptide receptor-like-1 as a functional receptor". eng. In: *Journal of Immunology (Baltimore, Md.: 1950)* 172.11, pp. 7078–7085. ISSN: 0022-1767. DOI: 10.4049/jimmunol.172.11.7078.

Zhang, Kejia et al. (May 2024). *Human TRMT1 and TRMT1L paralogs ensure the proper modification state, stability, and function of tRNAs*. en. Pages: 2024.05.20.594868 Section: New Results. DOI: 10.1101/2024.05.20.594868. URL: https://www.biorxiv.org/content/10.1101/2024.05.20.594868v1 (visited on 06/07/2024).

Zhao, Chunnian et al. (Feb. 2010). "MicroRNA let-7b regulates neural stem cell proliferation and differentiation by targeting nuclear receptor TLX signaling". en. In: *Proceedings of the National Academy of Sciences* 107.5. Publisher: National Academy of Sciences Section: Biological Sciences, pp. 1876–1881. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0908750107. URL: https://www.pnas.org/content/107/5/1876 (visited on 08/03/2021).

Zheng, Guanqun et al. (Sept. 2015). "Efficient and quantitative high-throughput tRNA sequencing". eng. In: *Nature Methods* 12.9, pp. 835–837. ISSN: 1548-7105. DOI: 10.1038/nmeth.3478.