Mapping and quantification of tRNA genes

Building artificial tRNA genome

tRNAs were annotated using tRNAscan-SE [1] and GtRNAdb [2]. Using the predicted tRNAs, artificial and mature tRNA genomes were generated for each species. tRNAscan-SE, one of the commonly used tools to predict tRNA genes, provides a score assigned to each putative tRNA gene. The genes with high scores are likely bona fide tRNA genes, while those with a low score are likely pseudogenes. tReasure uses tRNAs with high confidence scores that most likely function in translation by assessing a combination of domain-specific, isotype-specific, and secondary structure scores. To build the artificial genome, premature and mature tRNA libraries were generated using the predicted tRNA genes (Fig 1a). Each of the premature tRNAs comprises the predicted tRNA sequence and 50 nt 3′ and 5′ flanking sequences, and each of the mature tRNAs was generated by appending 3′ CCA tails to the predicted tRNA sequence (Fig 1a). An artificial genome was generated by masking all annotated tRNA genes and appending the premature tRNA library as additional chromosomes to the tRNA-masked genome (Fig. 1b). The mature tRNA genome was generated using the predicted mature tRNA library (Fig 1b).

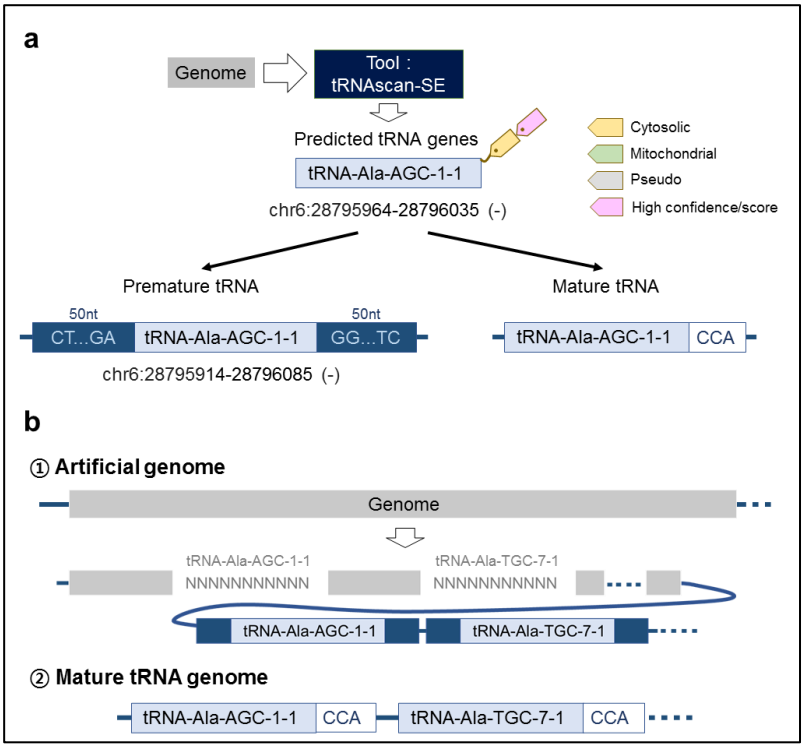

Fig.1 Building artificial tRNA genome and mature tRNA genome. a Generating premature and mature tRNA libraries from predicted tRNA genes. b Construction of the artificial genome and mature tRNA genome. Gray NNNNNN represents masking annotated tRNA gene in the genome.

Mapping mature tRNAs

For specific mapping of mature tRNA genes, tReasure provides a two-step mapping process based on a modification of a previous method [3] (Fig 2). First-round mapping was used to remove the non-tRNA and premature tRNA reads. Second-round mapping was used to detect mature tRNA genes.

1) First round mapping against the artificial genome

Preprocessed small RNA-seq reads were aligned against the artificial genome. The reads are classified into three types based on mapping location in the artificial genome: 1) reads that mapped to the tRNA-masked genomic region are classified as “non-tRNA”; 2) reads that map to the premature tRNA region of the additional chromosome (tRNA sequence with 50 nt flanking sequences) are classified as “premature tRNA”; and 3) reads that map to the pure tRNA gene sequence (without the flanking sequences) of the additional chromosome are classified as “mature tRNA”. Of the three types of reads, mature tRNAs were used for second-round mapping. In addition, < 30 nt non-tRNA reads are also used for the second round, as described elsewhere [4] (Fig. 2).

2) Second round mapping against mature tRNA genome

For final mapping of the mature tRNAs, the sequence reads from the first-round mapping were aligned against the mature tRNA genome. The reads that mapped to the mature tRNA genome were defined as mature tRNAs.

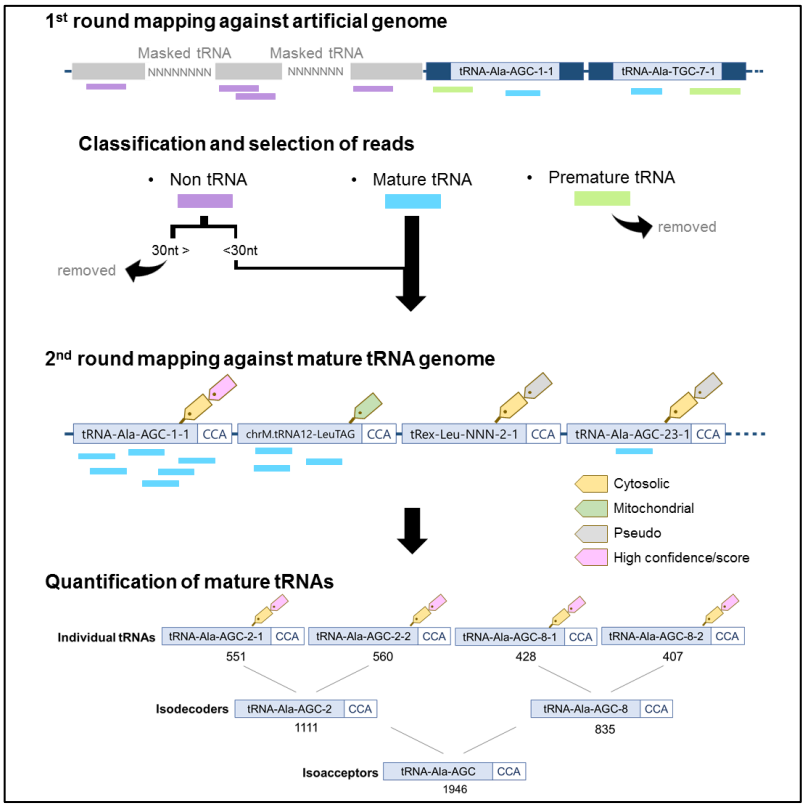

Fig 2. Mapping and quantification of mature tRNAs. The reads that mapped to the tRNA-masked genomic region are classified as “non-tRNA” (purple bars). The reads that map to the premature tRNA region of the additional chromosome are classified as

"premature tRNA" (blue bars). The reads that map to the pure tRNA gene sequence are classified as 'mature tRNA" (green bars). Only cytosolic and highly confident tRNAs are used for final quantification of mature tRNAs. The numbers below each individual tRNA, isodecoder, and isoacceptor represent the mapped read counts.

Quantification of mature tRNAs

For reliable and accurate quantification of mature tRNAs, tReasure uses only cytosolic tRNAs and tRNAs with high confidence values rather than the whole set of predicted mature tRNAs. The number of reads mapped to the individual tRNAs (cytosolic and high-confidence tRNAs) was counted (Fig. 2). tReasure can measure three different levels of mature tRNAs: individual tRNA, isodecoders, and isoacceptors. Isodecoders and isoacceptors were quantified by merging the counts of individual tRNAs (Fig. 2).

1. Chan, P.P.; Lowe, T.M. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol Biol* **2019**, *1962*, 1-14, doi:10.1007/978-1-4939-9173-0_1.
2. Chan, P.P.; Lowe, T.M. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* **2016**, *44*, D184-189, doi:10.1093/nar/gkv1309.
3. Hoffmann, A.; Fallmann, J.; Vilardo, E.; Morl, M.; Stadler, P.F.; Amman, F. Accurate mapping of tRNA reads. *Bioinformatics* **2018**, *34*, 1116-1124, doi:10.1093/bioinformatics/btx756.
4. Hernandez-Alias, X.; Benisty, H.; Schaefer, M.H.; Serrano, L. Translational efficiency across healthy and tumor tissues is proliferation-related. *Mol Syst Biol* **2020**, *16*, e9275, doi:10.15252/msb.20199275.