Supplemental Material

Predicting transfer RNA gene activity from sequence and genome context

Bryan P. Thornlow[1], Joel Armstrong[1,2], Andrew D. Holmes[1], Jonathan M. Howard[1], Russell B. Corbett-Detig[1,2], Todd M. Lowe[1,2]

[1]Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064; [2]Genomics Institute, University of California, Santa Cruz, CA 95064

Supplemental Material

SUPPLEMENTAL TEXT:

Possible tRNA gene classification and assembly errors. Ideally, we would observe that all species contained at least one tRNA gene per expected anticodon (excluding anticodons for which no active tRNA gene is observed in human or mouse). We found three exceptions to this rule. *C. hircus* (goat) is the only species without a predicted active tRNA-Ser-TGA gene. *C. hircus* is also one of only three species that has any tRNA-Ser-GGA genes, and it has 27 such genes, one of which was predicted active. These 27 tRNA-Ser-GGA genes do exhibit some sequence similarity, as 21 of these share a 10-base 3' flanking motif (AGAAGGAAAT). They may therefore reflect a very recent proliferation event. However, it is more likely that they represent assembly errors, because we have previously shown that the immediate 3' flanking regions of tRNA genes experience high mutation rates (Thornlow et al. 2018), these 27 tRNA genes are found on 15 different chromosomes, and Ser-GGA is not a standard eukaryotic tRNA (Grosjean et al. 2010). We also predict that the lone tRNA-Leu-TAA gene in *Erinaceus europaeus* (hedgehog) is inactive. The probability score for this prediction is 0.52, nearly the minimum possible score, indicating that it may be active. Additionally, the *Chinchilla lanigera* (chinchilla) genome in our Cactus graph (ChiLan1.0) does not contain any high-confidence selenocysteine (SeC) tRNA genes. This is likely due to an incomplete genome assembly, as selenocysteine tRNAs are required in mammals (Johansson et al. 2005).

We also identified several anomalous lineage-specific expansions that may represent SINEs or assembly errors. *E. europaeus* has 123 tRNA-Lys-CTT genes. All have high tRNAscan-SE general bit scores and are unlikely to be pseudogenes. We classified 96 of these as active. tRNA-derived SINEs can proliferate rapidly via retrotransposition, and a rapid expansion may have occurred in this species, targeting areas within or near regions of high transcription. While few of these tRNA genes have completely identical sequences, 43 have identical 3' flanking regions, extending 10 bases after the tRNA gene (CCATTTGTTG). An additional 24 genes share a different 3' flanking motif (CCAGATGTTG). The immediate flanking regions of actively transcribed tRNA genes are expected to be highly divergent due to transcription-associated mutagenesis (Thornlow et al. 2018). One possible explanation for this discrepancy is that these motifs are part of a fused SINE element responsible for the proliferation of the tRNA-derived SINE element. A similar expansion has occurred in *Microtus ochrogaster* (prairie vole), whose genome contains 131 tRNA-Lys-CTT genes, 66 of which we classified as active. 25 of these genes are exactly identical in tRNA gene sequence, suggesting a similarly recent expansion. These difficulties illuminate a benefit of our classifier in increased detection of SINE elements, as each of these passes through the high confidence filter in tRNAscan-SE, but may not represent truly actively transcribed tRNA loci.

We also noticed an expansion of Sup-CTA, -TCA and -TTA tRNA genes in cow, goat and orca, as these species have 15, 7 and 34 such Sup-tRNA genes respectively, all of which are predicted inactive (Supplemental Table S10). For comparison, no other species in our phylogeny has more than 2. Cow, goat and orca also have 11, 4 and 14 tRNA-Und-NNN genes, while no other species has more than 3, as well as 39, 40 and 35 Trp-CCA genes, respectively, while no other species has more than 12. We observe a similar expansion of tRNA-Gly-CCC genes, where cow, goat, orca, and relative outgroup *Dasypus novemcinctus* (nine-banded armadillo) have 36, 35, 38 and 40 genes, respectively, while no other species has more than 9. As is the case for the Lys-CTT and Ser-GGA genes mentioned above, many of these likely represent either assembly errors or the decreased ability of the tRNAscan-SE high confidence filter to detect SINE elements in these species. Conversely, several tRNA gene families, such as Glu-TTC, Cys-GCA and Ala-AGC, are present in many copies across all or most species in our phylogeny.

There are several other anomalous predictions. The goat genome has only 2 tRNA-Ala-AGC genes predicted active (and 10 predicted inactive) while all other species have at least 8 tRNA-Ala-AGC genes predicted active. Consistent with this, we do not observe ATAC-seq peaks for any of the tRNA-Ala-AGC genes in goat (Foissac et al. 2019). The naked mole rat genome (*Heterocephalus glaber*) also has 8 tRNA-SeC-TCA

genes predicted active while no other species has more than 2. However, we do not have any experimental data to validate this prediction.

More generally, we observe a relative abundance of inactive tRNA genes in the cow and goat genomes. We have noted several anomalous isotype distributions above, but find no abnormal patterns regarding their location in the genome. Many of these genes are most likely SINEs that the tRNAscan-SE high confidence filter was unable to filter out. We note that it is fairly typical for one or two tRNA chromosomes to contain the majority of high-confidence tRNA genes. In the cow genome, chromosome 3 has 108 tRNA genes (23 predicted inactive) and chromosome 23 has 161 tRNA genes (39 predicted inactive). For goat, chromosome 3 contains 84 tRNA genes (15 predicted inactive) and chromosome 19 has 54 tRNA genes (22 predicted inactive). For comparison, in the human genome, 136 of the genes in our training set are on chromosome 6, and 15 of these are inactive based on chromatin data.

Additionally, it is typical for tRNA genes to exist in clusters. For reference, the human chromosome 1 contains 32 tRNA genes in a 250 kilobase region, with 22 known to be inactive. However, many of these genes are found in segmental duplications. Likewise, in the cow genome, 42 tRNA genes (including those in segmental duplications) can be found within 200 kb of each other on chromosome 23, with 7 predicted to be inactive. On the goat chromosome 3, 53 tRNA genes (including those in segmental duplications) are present in a 350 kb region, with 4 predicted as inactive. It is common for tRNA genes within the same cluster to share the same binary activity level. For example, 79.6% of human tRNA genes are of the same activity class as their nearest neighbor. For the cow and goat genomes, based on our predictions, this is true of 75.9% and 78.8% tRNA genes, respectively.

To our knowledge, no actively transcribed suppressor tRNA gene has been demonstrated in a primate genome. However, we have predicted a tRNA-Sup-CTA in *Pan troglodytes* (chimpanzee) as active, although experimentation is necessary to support that this tRNA gene is indeed a suppressor and is also actively transcribed. While nonsense suppression via readthrough by near-cognate tRNAs has been demonstrated several times in metazoa (Jungreis et al. 2011; Yacoubi et al. 2012; Loughran et al. 2018; Roy et al. 2015) and suppressor tRNAs have been found in other species (Beier and Grimm 2001; Valle et al. 1987), evidence for an actively transcribed suppressor tRNA in a primate species is of great interest for future experimentation. This tRNA gene is a conserved ortholog of *TRQ-CTG6-1* (GtRNAdb ID: Gln-CTG-6-1), and exhibits several chimpanzee-specific nucleotide substitutions that are not expected to significantly impair the function of its transcript. However, this finding may merely reflect an assembly error. In some chimpanzee assemblies, this is a tRNA-Gln-CTG gene, while in others it is a tRNA-Sup-CTA gene. Therefore, this may in fact be an actively transcribed chimpanzee tRNA-Gln-CTG gene.

SUPPLEMENTAL METHODS:

Preparation of DM-tRNA-seq sequencing libraries from mouse brains and livers. Mouse brain and liver samples were obtained from 3 C57Bl/6 male 6-8 weeks old timepoints. Mice are housed in facilities with roughly 12 hr light-dark cycles (7AM-7PM light and 7PM-7AM dark). All samples were collected from said mice at 2-3 PM EST. All tissue samples were then flash-frozen in liquid nitrogen and stored in -80˚C freezer until use.

Isolation of total RNA from mouse brain and liver was performed using Direct-Zol RNA MiniPrep Kit (Zymo Research) with TRI Reagent (Molecular Research Center, Inc.). Briefly, appropriate amounts of TRI Reagents was added to each tissue sample, along with ~400 uL of zirconium silicate beads (1.0 mm), and samples were homogenized using a tissue homogenization centrifuge. 100 µg total RNA isolated from TRI Reagent protocol was used to isolate small RNA (RNA < 200 nts), using the MirVana miRNA Isolation Kit (Life Technologies), according to the manufacturer's instructions. RNA was DNase I-treated and concentrated using a RNA Clean and Concentrate-25 (Zymo Research). Small RNA samples were divided and treated as “minus-AlkB” control treatment and “plus-AlkB” experimental treatment. AlkB treatment was performed as

previously described (Cozen et al. 2015; Zheng et al. 2015). RNA was again concentrated after enzymatic treatment using an RNA clean and concentrator-5 kit (Zymo Research). Treated RNA was used for library preparation.

For DM-tRNA-seq, libraries were constructed with a 1 uµg of treated and control small RNA using the TGIRT™ Improved Modular Template-Switching RNA-seq Kit (InGex, LLC), which utilizes Illumina-compatible adapters for amplification and sequencing purposes. Libraries were amplified using PCR primers supplied with the NEBNext Small RNA library Prep Set (NEB) and purified using Agencourt AMPure XP beads (Beckman). Resulting purified sequencing libraries were quantified using NanoDrop and Agilent DNA High Sensitivity kit. Libraries were then pooled in equimolar amounts, and concentrated using a DNA Clean and Concentrate-5 (Zymo Research). Pooled PCR-amplified libraries were size-selected (140-250 nt) on a 6% non-denaturing TBE-acrylamide gel to remove unwanted primer dimer products. Pooled libraries were then eluted from gel pieces using Gel Elution Buffer (NEB) and precipitated using .3 M NaOAc, 80% Ethanol, and 1 uµL of Linear Acrylamide at final concentration. Samples were left in -80℃ freezer overnight to precipitate. Precipitated libraries were then pelleted, washed twice in 80% Ethanol, and resuspended in Pure H2O. Final processed pooled libraries were then quantified using NanoDrop and DNA High Sensitivity Bioanalyzer kit (Agilent Technologies) prior to sequencing.

Generating feature data. To generate phyloP data, we used the HAL (Hickey et al. 2013) and PHAST (Hubisz et al. 2011) toolkits. We extracted four-fold degenerate (4d) sites from each genome of interest with the hal4dExtract command. We reduced each set of 4d sites to 100,000 sites, and trained a phyloP model with this reduced set of 4d sites using the phyloFit command. We used these 4d sites because phyloP requires a neutral or nearly-neutral set of sites from which to generate a substitution matrix, so that it can properly weigh substitutions observed in the regions of interest and assign phyloP scores. We then extracted each genome of interest from the Cactus graph using the hal2fasta command, and applied tRNAscan-SE 2.0 (Chan et al. 2019) to each using the -o, -f, -s, -m, -b, -a, --detail, -H, and -y flags. We applied EukHighConfidenceFilter using the -r, -i, -s, -p, and -o flags. For classification purposes, we used all tRNAs in the data output by the EukHighConfidenceFilter, which includes high-confidence tRNA genes, as well as tRNA genes with high tRNAscan-SE bit scores and isotype mismatches. From there, we created custom Python scripts (available in Supplemental Code and at https://github.com/bpt26/tRNA_classifier) to generate .bed files and .fasta files including tRNA loci and their flanking regions, up to 350 nucleotides upstream and downstream of each gene. We applied the hal2maf and phyloP commands to generate phyloP data for these loci, as well as the phastCons command to identify tRNA genes with conserved flanking regions. For conserved elements no larger than 100 base pairs with phastCons scores of at least 200, we used the boundaries of the conserved element instead of the tRNA boundaries, as conserved flanking regions would hinder classification otherwise. We applied this process to obtain all feature data for all species, including human. For all other species, we trained a random forest classifier using ten-fold cross-validation in scikit-learn on the human data, and then applied it to feature data for each species of interest. Our final model in scikit-learn uses 250 estimators, a minimum of 2 samples required to split an internal node and a maximum depth of 4 nodes (Pedregosa et al. 2011).

To estimate the minimum free energy of canonical secondary structure, we used folding constraints output by tRNAscan-SE (Chan et al. 2019) as inputs to RNAfold (Lorenz et al. 2011), which estimates minimum free energy. Initially, we examined both RNAfold and tRNAscan-SE output. RNAfold output is significantly correlated with tRNAscan-SE secondary structure scores (Spearman's rank correlation, $p < 2\times10^{-4}$), indicating interchangeability. To determine which to include in the model, we created a separate model that included both scores, and found that RNAfold output had greater feature importance score (0.077 for RNAfold output compared to 0.014 for tRNAscan-SE secondary structure score). We also tested a model that includes tRNAscan-SE secondary structure, HMM, and isotype-specific bit scores. Upon their inclusion, the feature importance score of the tRNAscan-SE general bit score decreases slightly from 0.070 to 0.050, but it remains more important to the model than any of the other tRNAscan-SE scores (secondary structure, HMM

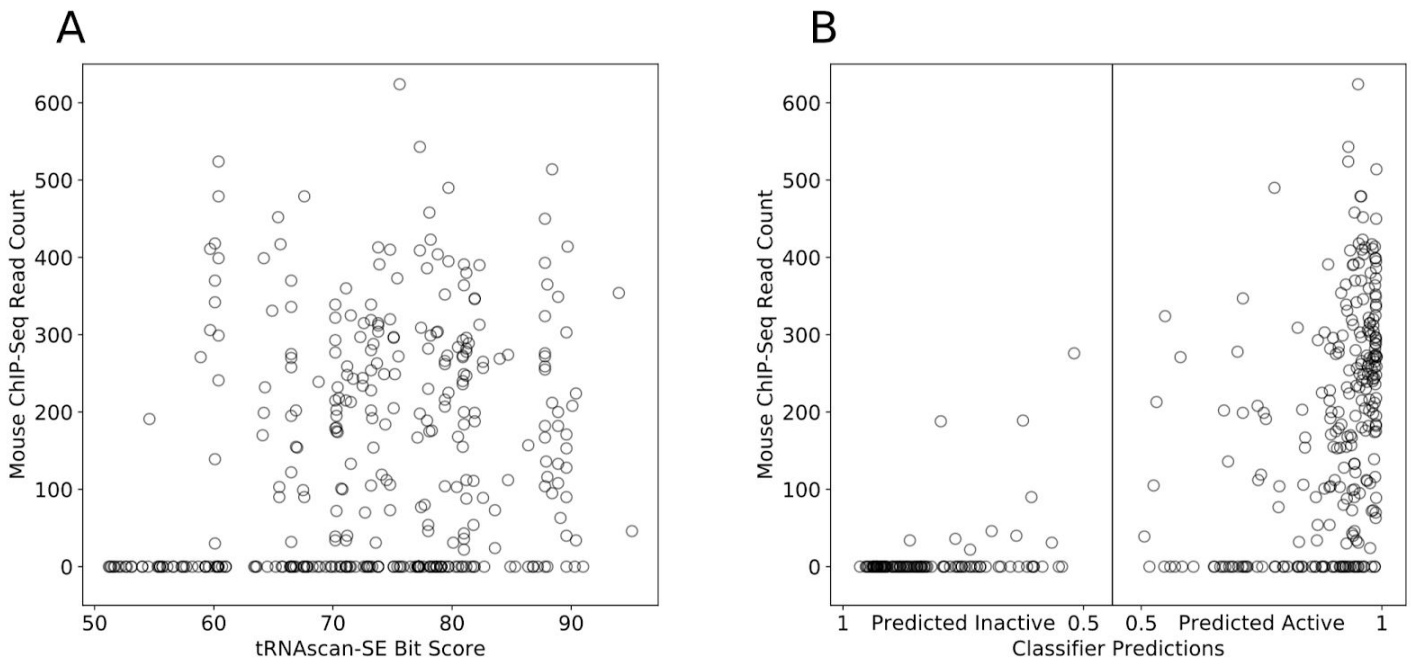
and isotype-specific bit scores had feature importance values of .015, .044 and .036, respectively), supporting its usage in the model.

Fitting to a Markov Model. For each species, we converted our ortholog set to an alignment, where for each species, ‘2’ indicated a predicted active tRNA, ‘1’ indicated a predicted inactive tRNA, and ‘0’ indicated no detected ortholog. We used RevBayes to fit this data, in conjunction with a phylogeny from TimeTree (Kumar et al. 2017), to a Markov model and allowed only the Q matrix to be changed over 10,000 generations. We then used the getTransitionProbabilities() command to estimate transition probabilities to and from each state over branch lengths of 1 million years.

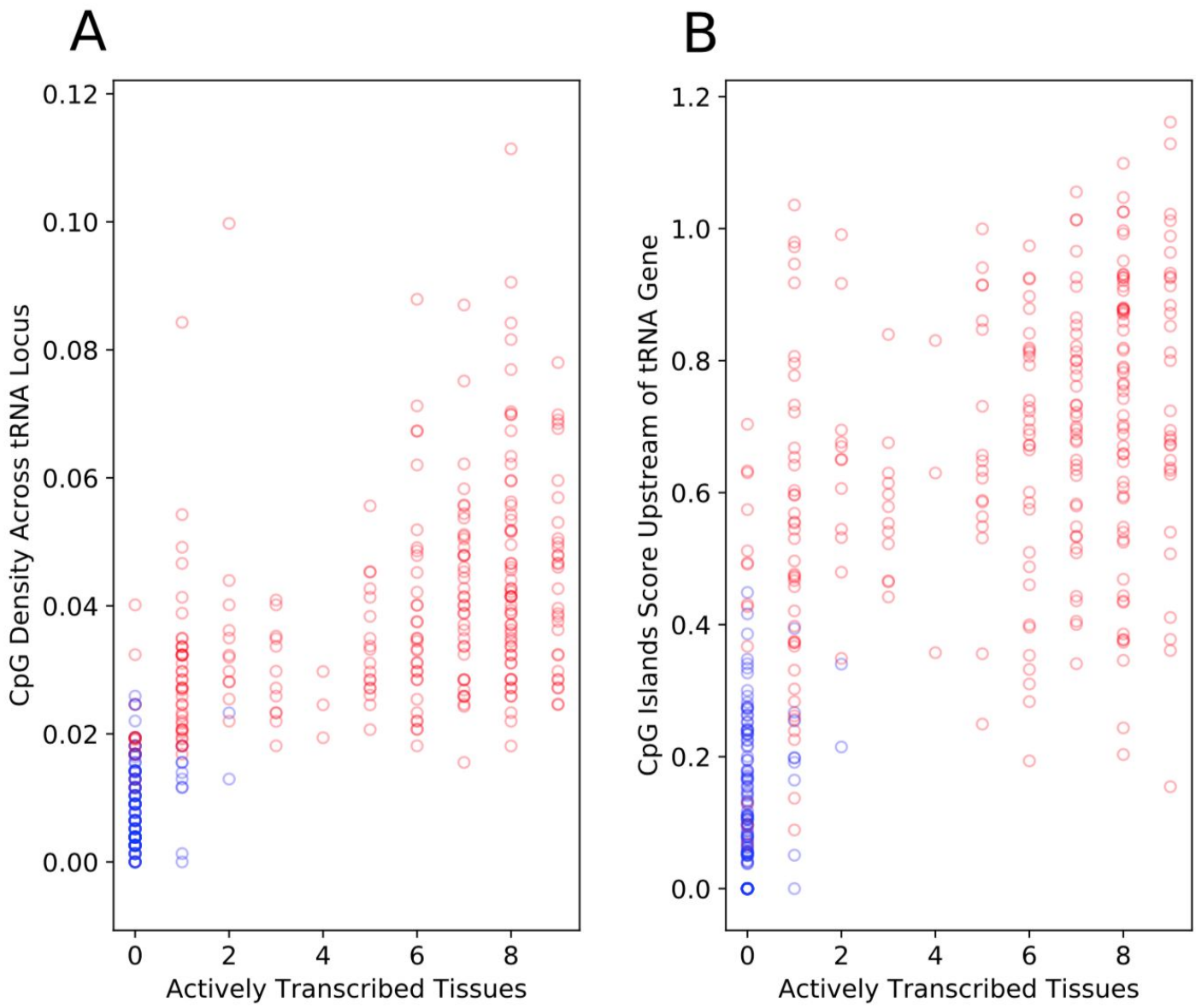
The guide tree for the Cactus graph was:

(((((((((Homo_sapiens:0.00655,Pan_troglodytes:0.00684)Anc32:0.00422,Gorilla_gorilla_gorilla:0.008964)Anc29:0.009693,Pongo_abelii:0.01894)Anc25:0.015511,Macaca_mulatta:0.043601)Anc20:0.08444,Aotus_nancymaae:0.08)Anc15:0.08,Microcebus_murinus:0.10612)Anc11:0.083494,((((Jaculus_jaculus:0.1,(Microtus_ochrogaster:0.14,(Mus_musculus:0.084509,Rattus_norvegicus:0.091589)Anc30:0.047773)Anc26:0.06015)Anc21:0.122992,(Heterocephalus_glaber:0.1,(Cavia_porcellus:0.065629,(Chinchilla_lanigera:0.06,Octodon_degus:0.1)Anc31:0.06)Anc27:0.05)Anc22:0.06015)Anc16:0.05,Marmota_marmota:0.1)Anc12:0.05,Oryctolagus_cuniculus:0.21569)Anc08:0.04)Anc04:0.040593,(((Sus_scrofa:0.12,(Orcinus_orca:0.069688,(Bos_taurus:0.04,Capra_hircus:0.04)Anc23:0.09)Anc17:0.045488)Anc13:0.02,((Equus_caballus:0.109397,(Felis_catus:0.098612,(Canis_lupus_familiaris:0.052458,Mustela_putorius_furo:0.08)Anc28:0.02)Anc24:0.049845)Anc18:0.02,(Pteropus_alecto:0.1,Eptesicus_fuscus:0.08)Anc19:0.033706)Anc14:0.03)Anc09:0.025,Erinaceus_europaeus:0.278178)Anc05:0.021227)Anc02:0.023664,((Loxodonta_africana:0.098842,Chrysochloris_asiatica:0.04)Anc06:0.05,Dasypus_novemcinctus:0.169809)Anc03:0.02)backbone_root;

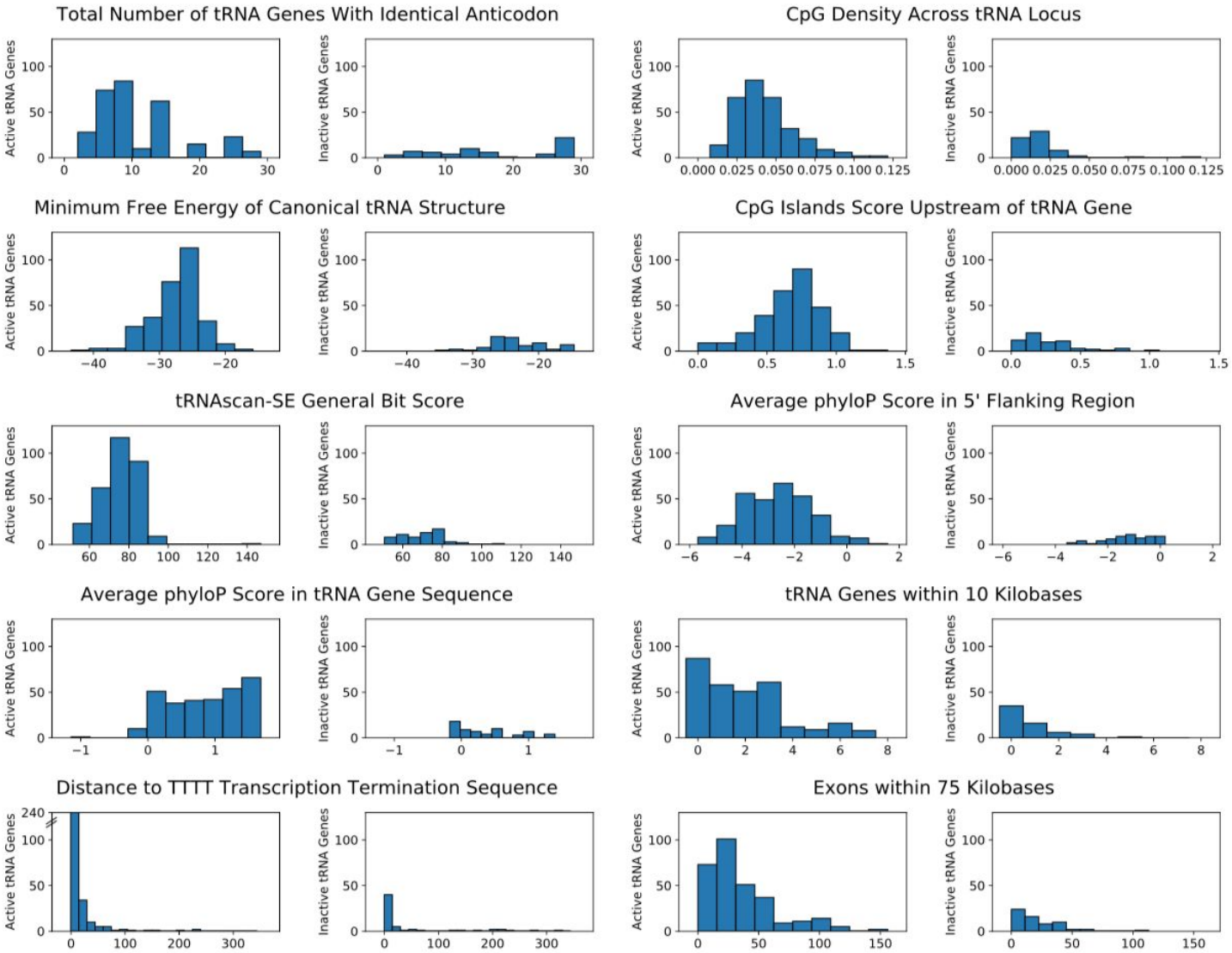
SUPPLEMENTAL FIGURES:



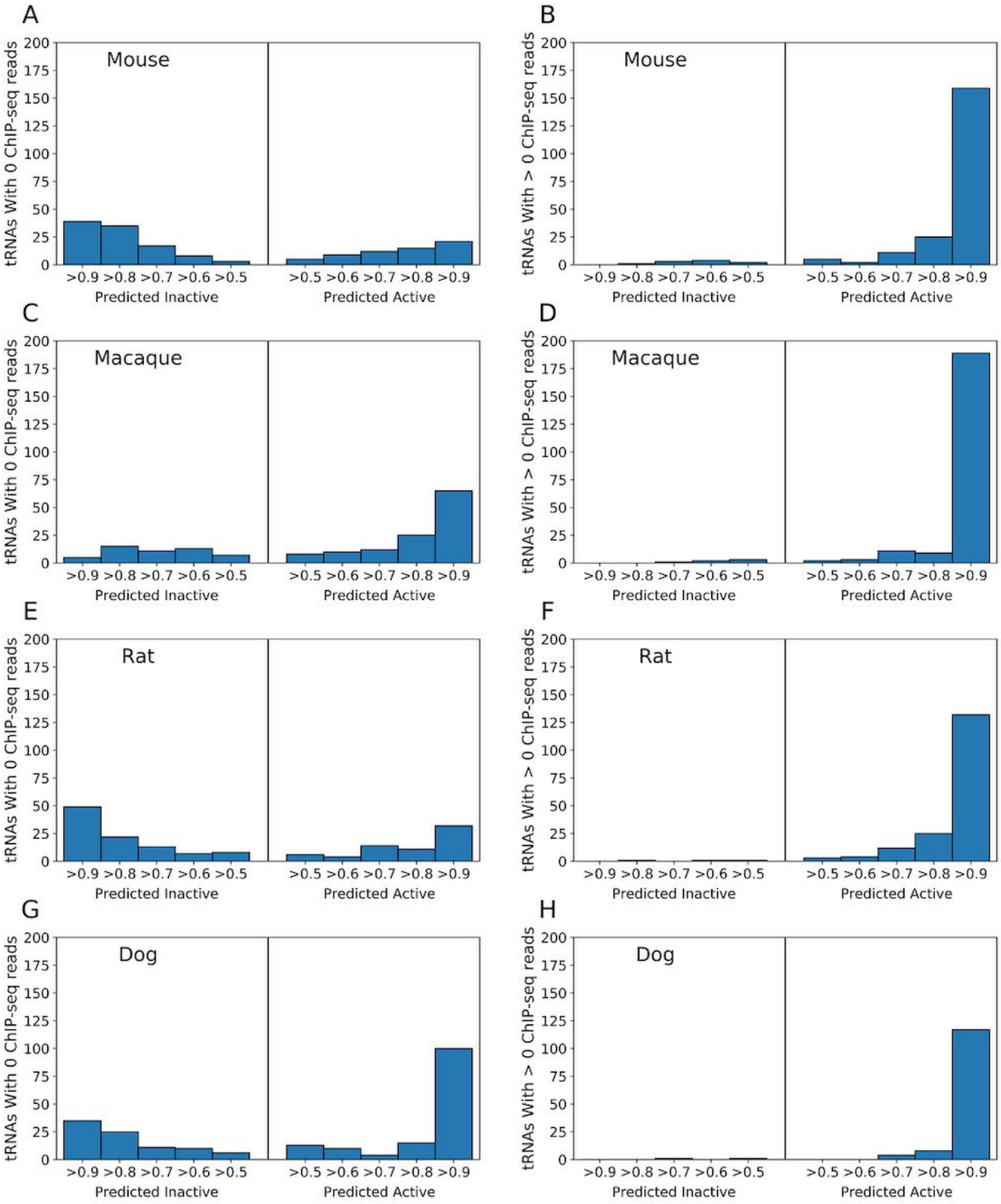
Supplemental Figure S1: Our classifier improves on tRNAscan-SE bit score for inferring transcriptional activity. A: The highest RNA Polymerase III ChIP-seq read count across mouse liver, muscle and testes (Kutter et al. 2011) is compared to the tRNAscan-SE general bit score for each tRNA gene. Many relatively high-scoring tRNA genes (greater than ~70 bits) have no evidence for activity. B: The same ChIP-Seq data is compared to probability scores output by our classifier. tRNA genes predicted as inactive with greater probability are furthest left, and tRNA genes predicted as active with greater probability are furthest right.



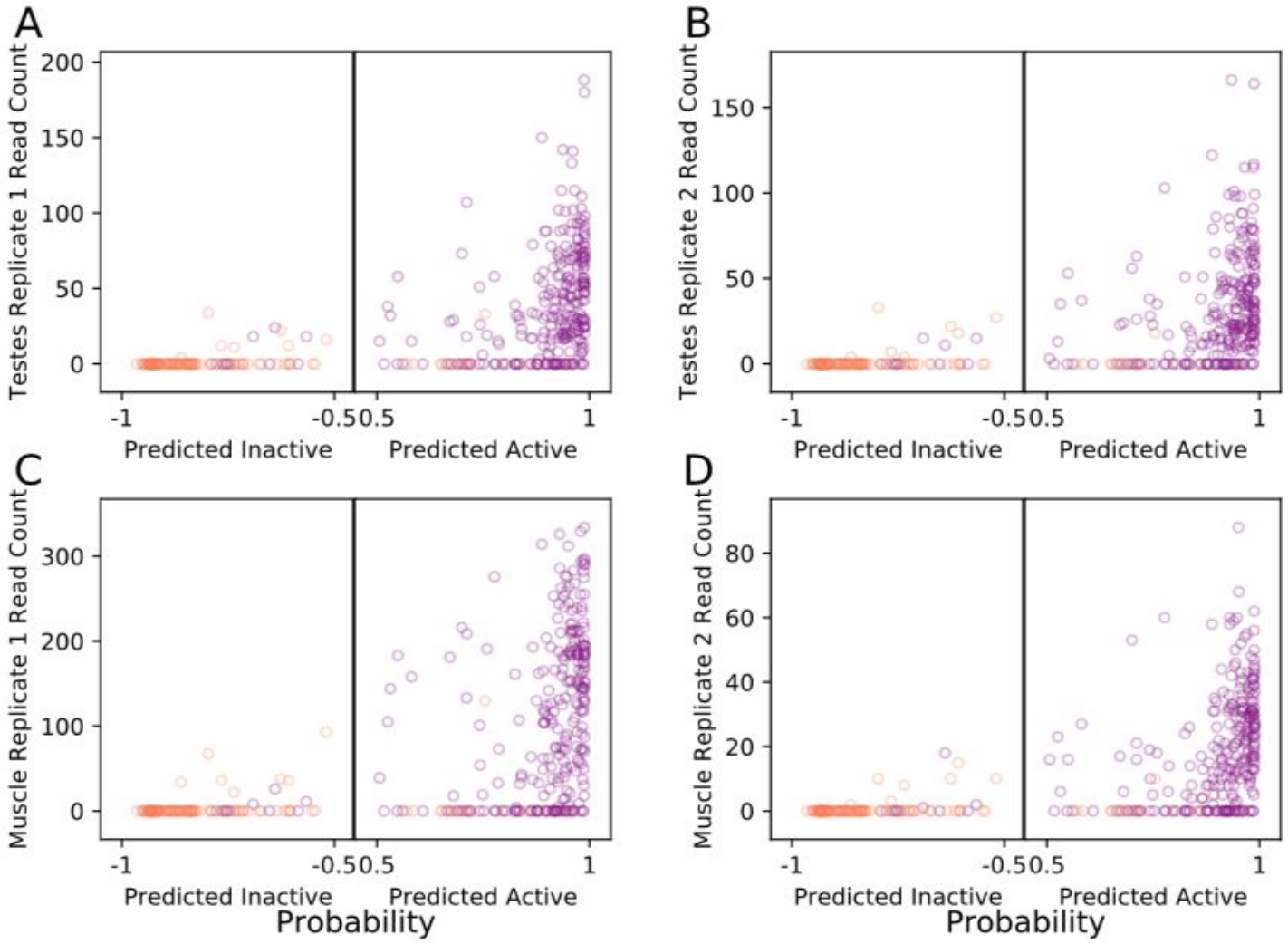
Supplemental Figure S2: CpG density and CpG Islands score vary with tissue-specificity in mouse. A: CpG density across each mouse tRNA locus included in our test set is compared to the number of tissues in which each tRNA gene is active based on epigenomic data from Bogu et al. 2016. B: The observed/expected CpG Islands score for the 350 nucleotide region upstream of each mouse tRNA gene in our test set is compared to the number of tissues in which each tRNA gene is active. tRNA genes in blue are predicted as inactive by our classifier, and tRNA genes in red are predicted as active.



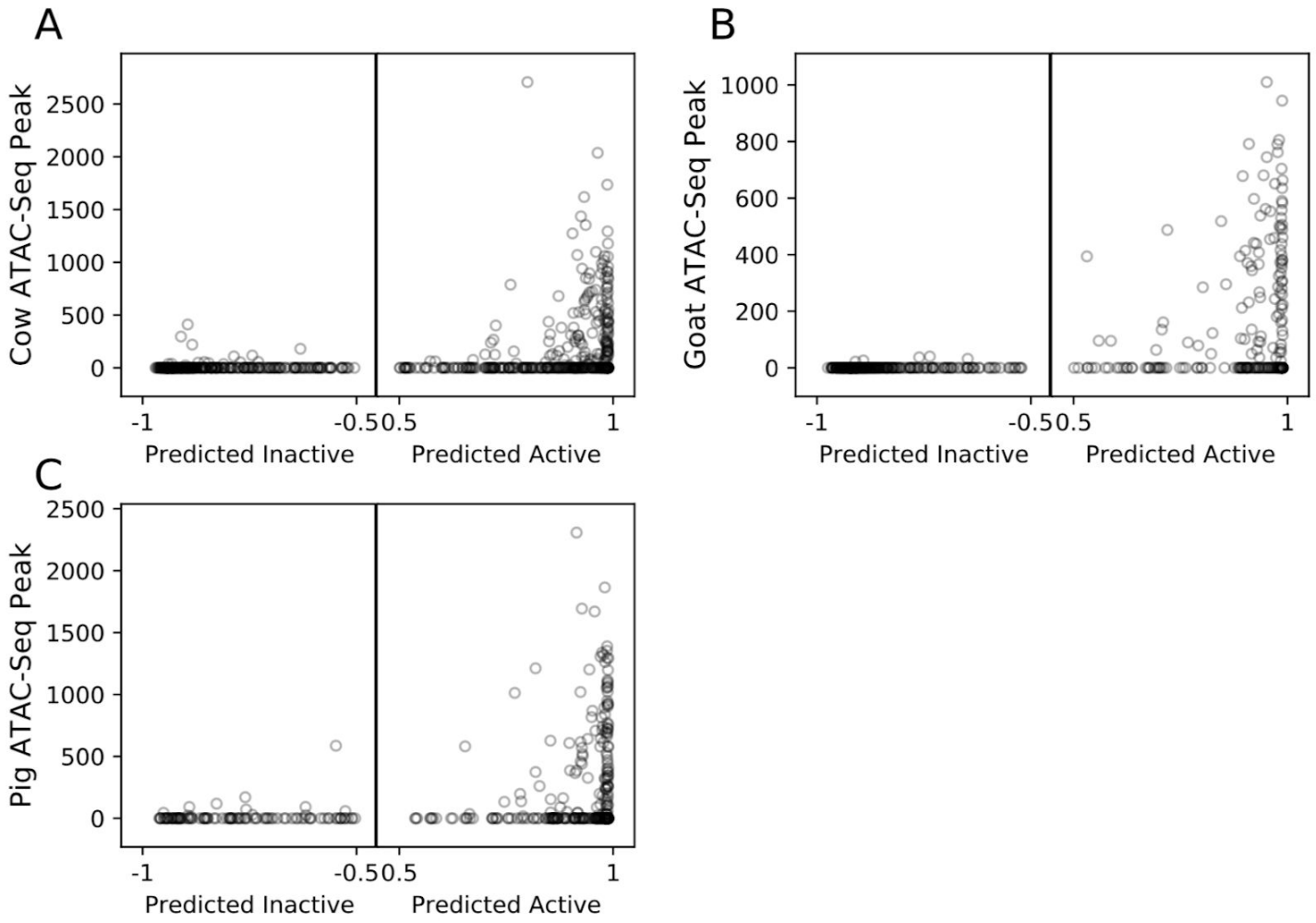
Supplemental Figure S3: Distributions of all features for active and inactive tRNA genes in the training set. For all intrinsic (left) and extrinsic (right) features in our model, the distribution across all active and inactive tRNA genes is shown. The mean and standard deviation for each feature can be found in Supplemental Table S5.



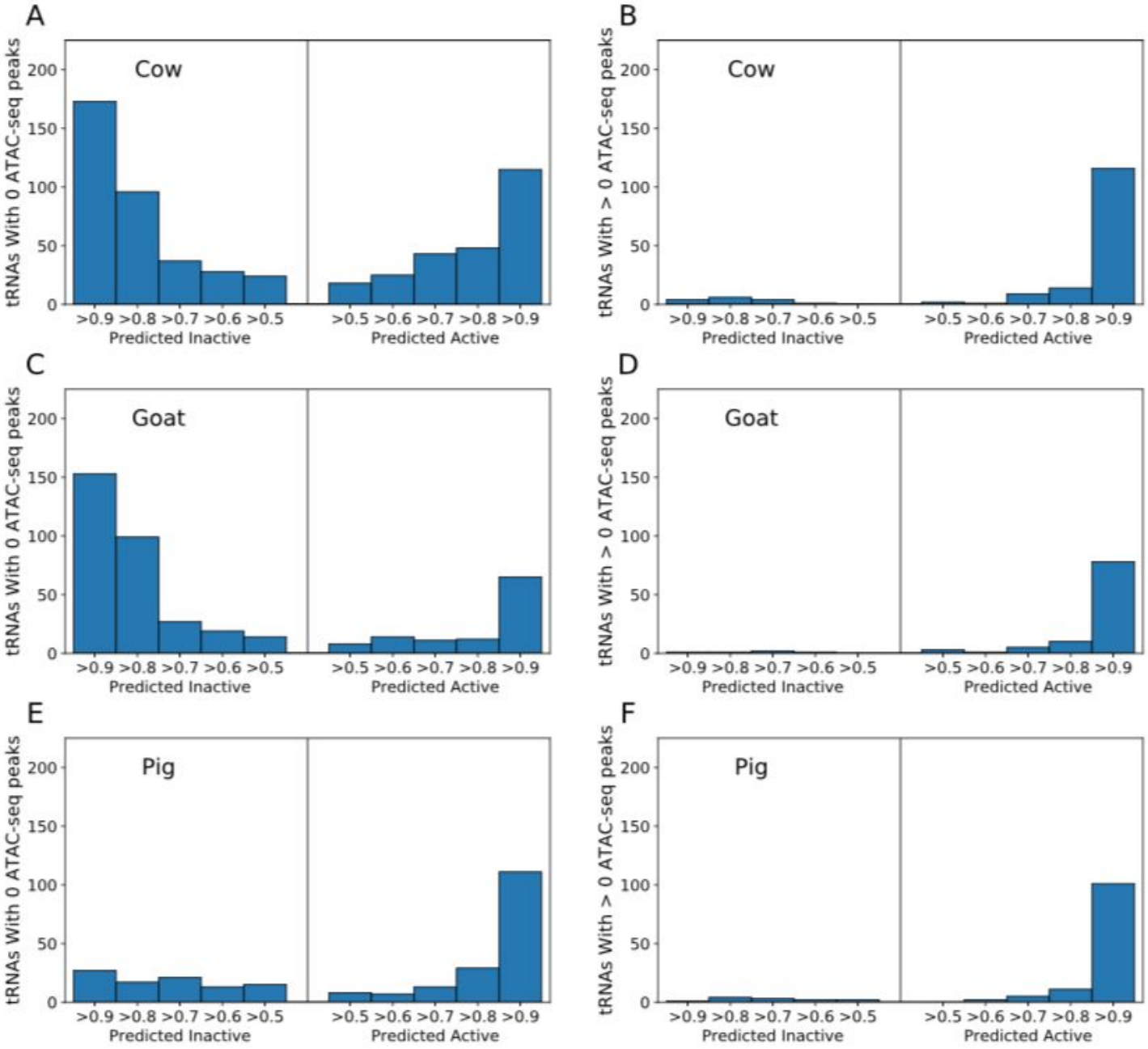
Supplemental Figure S4: Classifier predictions generally agree with Pol III occupancy in the livers of four species. All tRNAs with ChIP-seq read counts of 0 in mouse (A), macaque (C), rat (E) and dog (G) are shown with probability scores output by our classifier, with those predicted active on the right and those predicted inactive on the left. All tRNAs with ChIP-seq read counts greater than 0 are shown for mouse (B), macaque (D), rat (F) and dog (G) compared to their probability scores in the same manner. For mouse, we considered the maximum ChIP-seq read count across liver, testes and muscle, while in the other species, only data from liver was available (Kutter et al. 2011).



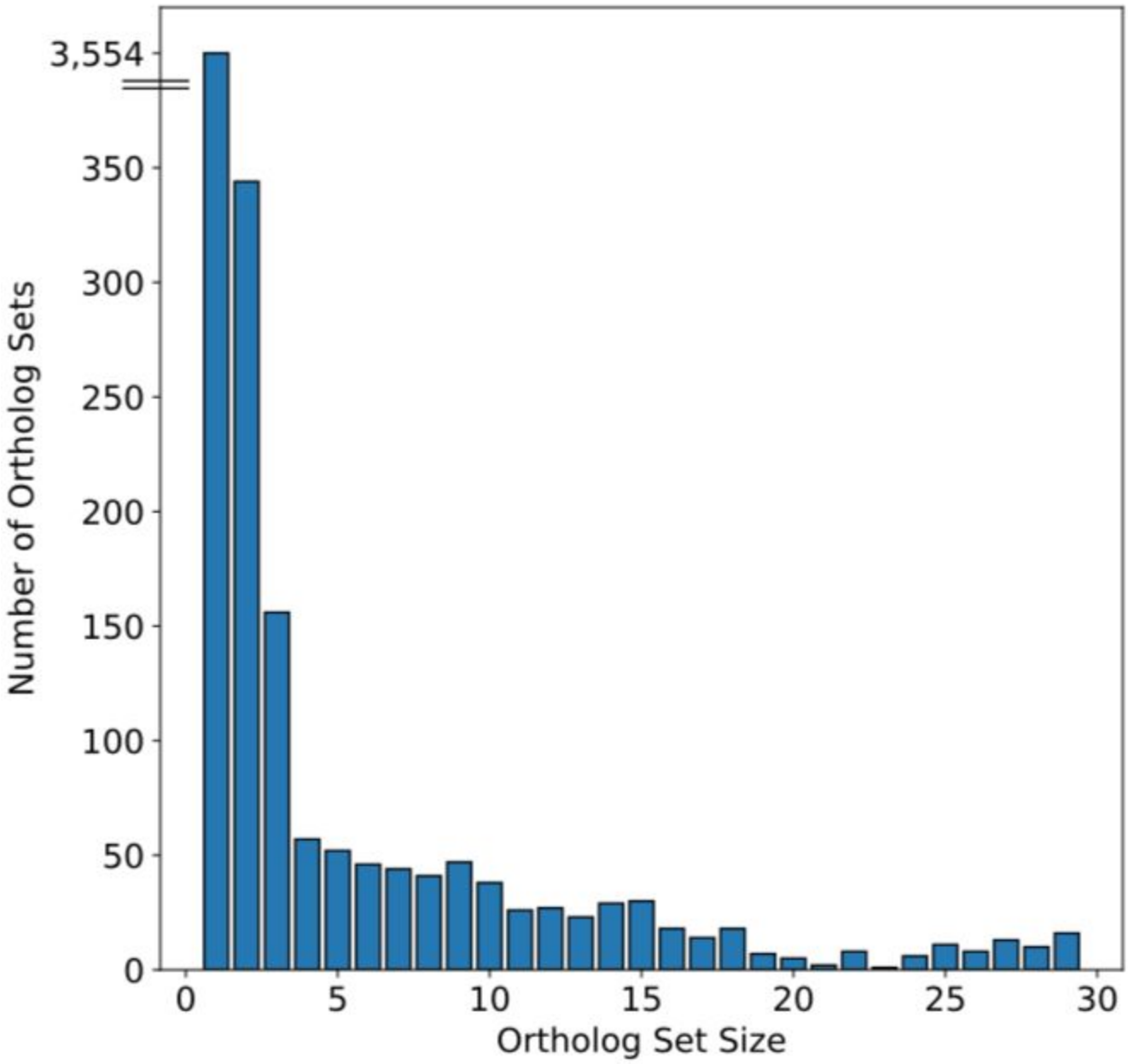
Supplemental Figure S5: Comparison of tRNA gene classifications to ChIP-seq data is consistent across tissues. Mouse ChIP-seq data compared to tRNA gene classifications and ChIP data for both testes replicates and both muscle replicates. Probability scores output by the classifier are shown on the horizontal axis where tRNA genes furthest left are predicted inactive with greater probability and tRNA genes furthest right are predicted active with greater probability. The vertical axis indicates read counts in (A, B) testes and (C, D) muscle from Kutter et al. 2011. Purple indicates measured activity in at least one tissue based on chromatin modification data from Bogu et al. 2016 while orange indicates no measured activity.



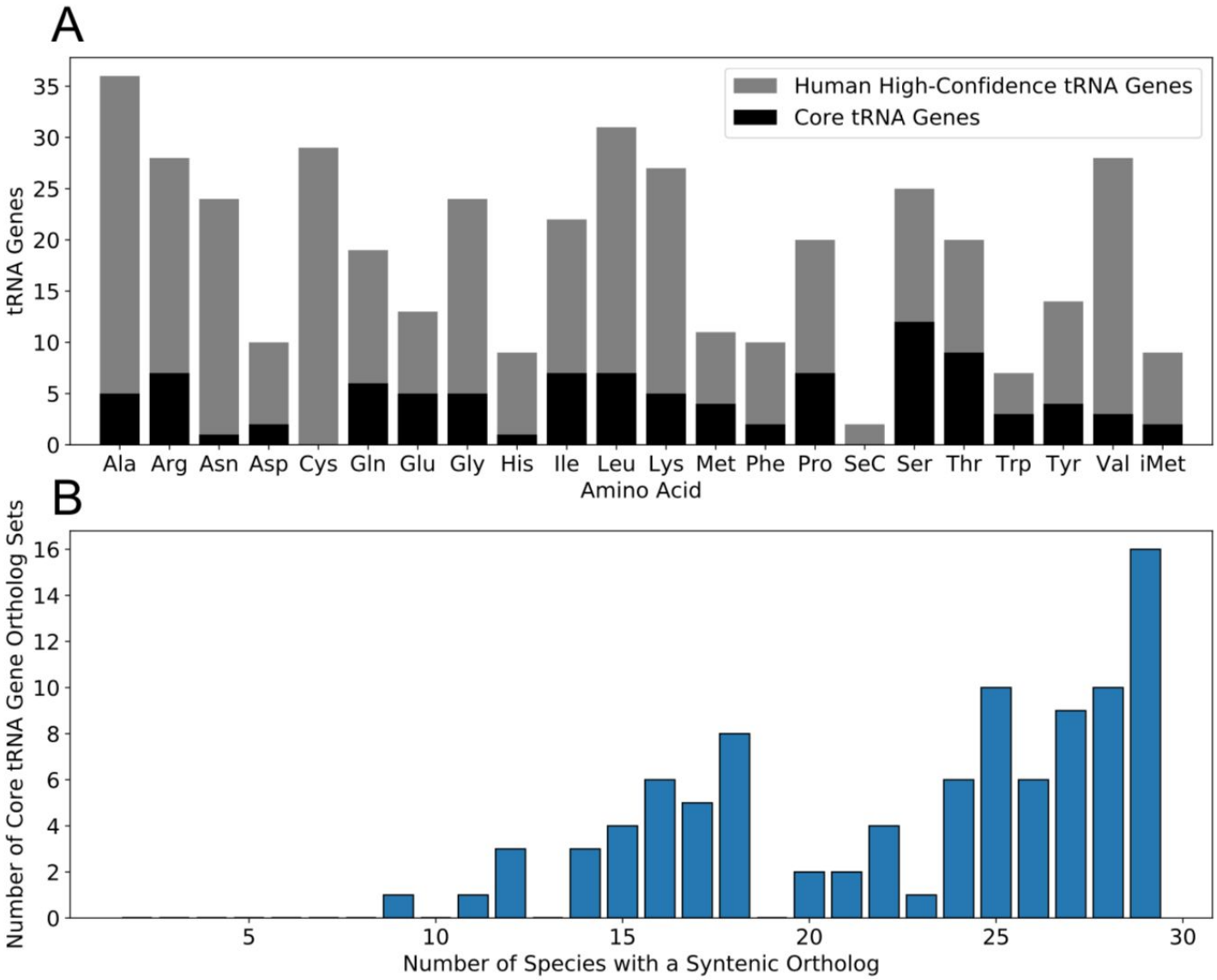
Supplemental Figure S6: Comparison of tRNA gene classifications to ATAC-seq data suggests similar accuracy across clades. The mean ATAC-seq peaks across liver, CD4 and CD8 cells in (A) cow, (B) goat and (C) pig within 250 base pairs of each tRNA gene are shown on the vertical axis. Probability scores output by the classifier are shown on the horizontal axis where tRNA genes furthest left are predicted inactive with greater probability and tRNA genes furthest right are predicted active with greater probability.



Supplemental Figure S7: Comparison of tRNA gene classifications to ATAC-seq data suggests similar accuracy across clades. All tRNAs with ATAC-seq peaks of 0 in cow (A), goat (C), and pig (E) are shown with probability scores output by our classifier, with those predicted active on the right and those predicted inactive on the left. All tRNAs with ChIP-seq read counts greater than 0 are shown for cow (B), goat (D), and pig (F) compared to their probability scores in the same manner.



Supplemental Figure S8: tRNA genes are generally either fairly deeply conserved or recently evolved. Size distribution of species-specific tRNAs (n=1), and all ortholog sets with tRNAs from two or more species. 3,554 of the 11,724 tRNA genes in our alignment are species-specific. The remaining 8,170 tRNA genes are condensed into 1,097 ortholog sets.



Supplemental Fig. S9: Cysteine is the only standard amino acid without a "core" tRNA. (A) All human active and inactive high-confidence tRNA genes are shown in gray for each amino acid, and those found in the core set of 97 tRNA genes are shown in black. Cysteine is one of the highest copy-number tRNA gene families, but is the only standard amino acid with zero core tRNA genes. (B) Histogram of sizes of the 97 core tRNA genes present in all 7 primate species.

A

Initial \ Final	Absent	Inactive	Active
Absent	0.994	0.004	0.002
Inactive	0.005	0.994	**0.001**
Active	0.015	**0.002**	0.983

B

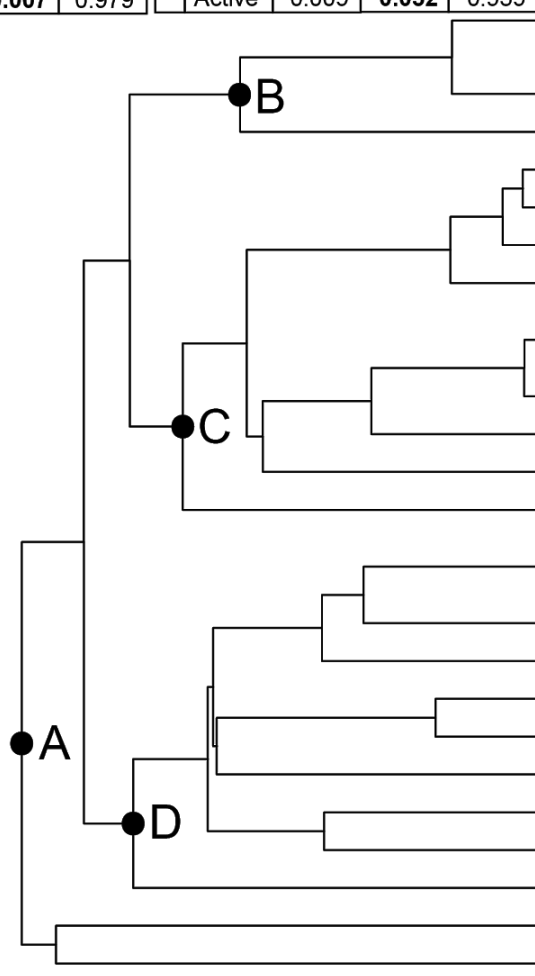
Initial \ Final	Absent	Inactive	Active
Absent	0.936	0.020	0.044
Inactive	0.074	0.899	**0.027**
Active	0.012	**0.145**	0.843

C

Initial \ Final	Absent	Inactive	Active
Absent	0.965	0.010	0.024
Inactive	0.010	0.973	**0.017**
Active	0.013	**0.007**	0.979

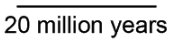
D

Initial \ Final	Absent	Inactive	Active
Absent	0.938	0.015	0.048
Inactive	0.007	0.950	**0.043**
Active	0.009	**0.032**	0.959

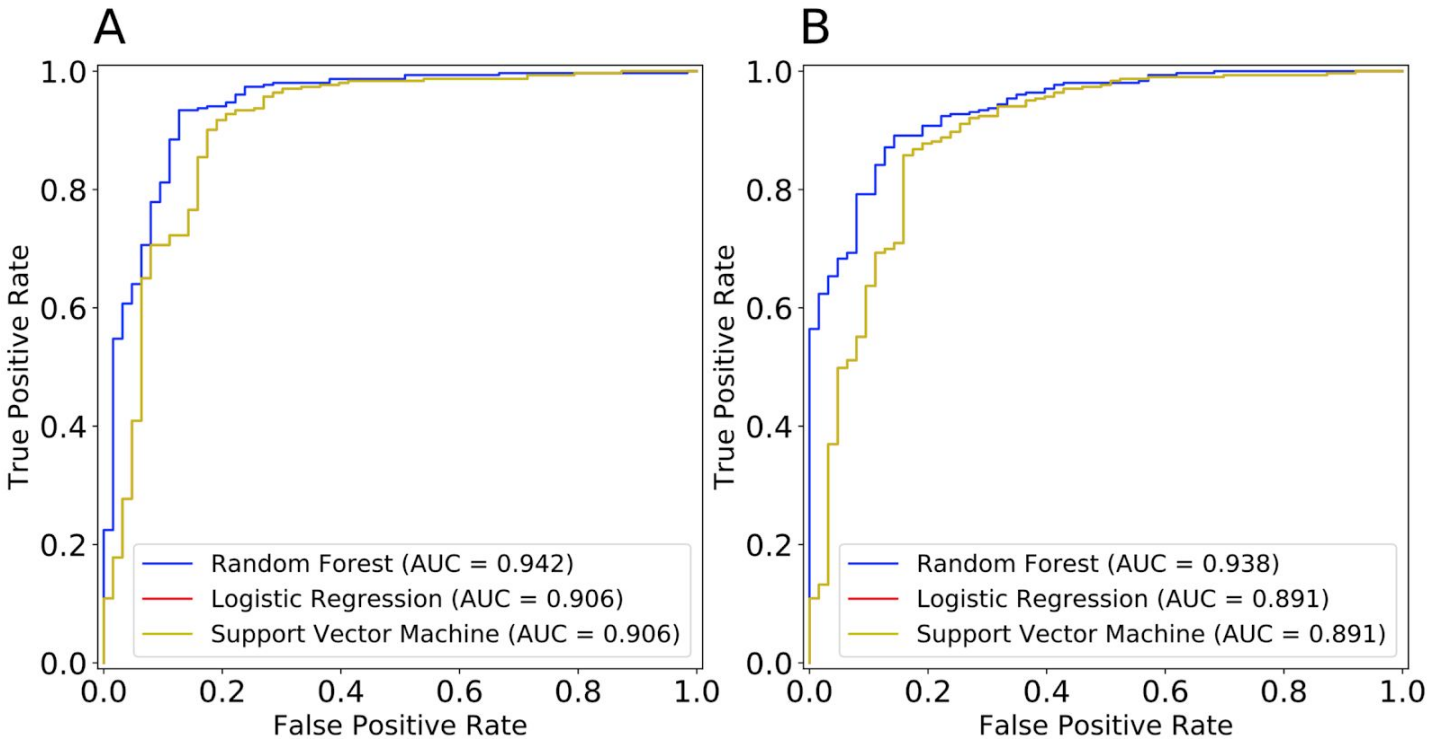


E

Species	Active	Inactive
Homo sapiens	367	64
Pan troglodytes	344	55
Gorilla gorilla gorilla	327	48
Pongo abelii	360	51
Macaca mulatta	336	57
Aotus nancymaae	292	42
Microcebus murinus	288	40
Octodon degus	241	57
Chinchilla lanigera	250	51
Cavia porcellus	278	72
Heterocephalus glaber	238	56
Mus musculus	287	116
Rattus norvegicus	271	110
Microtus ochrogaster	342	128
Jaculus jaculus	244	74
Marmota marmota	253	158
Oryctolagus cuniculus	408	75
Bos taurus	433	377
Capra hircus	228	319
Orcinus orca	333	334
Sus scrofa	328	105
Canis lupus familiaris	293	89
Mustela putorius furo	294	86
Equus caballus	355	64
Pteropus alecto	222	123
Eptesicus fuscus	199	15
Erinaceus europaeus	339	108
Dasypus novemcinctus	389	150
Chrysochloris asiatica	275	97



Supplemental Figure S10: Clades vary in their activity state transition rates. (A-D) Transition probability matrices over 1 million years shown as in Fig. 4 for all descendants of labeled nodes in the phylogenetic tree. Transitions between active and inactive states are in bold. (E) Number of tRNA genes predicted active and inactive including tRNA genes in segmental duplications. Primate species are colored in red and rodent species are colored in blue. For the total number of tRNA genes present upon removal of tRNA genes in segmental duplications, see Fig. 4B.



Supplemental Figure S11: Ten-fold cross-validation achieves slightly more accurate results compared to three-fold cross-validation. Receiver operating characteristic curves for random forest (blue), logistic regression (red) and support vector machine (yellow) upon application to (A) human training data with ten-fold cross-validation and (B) human training data with three-fold cross-validation. The curve corresponding to the logistic regression model is not visible due to overlap with the curve corresponding to the support vector machine model.

SUPPLEMENTAL REFERENCES

Beier H, Grimm M. 2001. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic Acids Res* **29**: 4767–4782.

Bogu GK, Vizán P, Stanton LW, Beato M, Di Croce L, Marti-Renom MA. 2016. Chromatin and RNA maps reveal regulatory long noncoding RNAs in mouse. *Mol Cell Biol* 36: 809–819. doi:10.1128/MCB.00955-15

Chan PP, Lin BY, Mak AJ, Lowe TM. 2019. tRNAscan-SE 2.0: Improved Detection and Functional Classification of Transfer RNA Genes. *bioRxiv* 614032.

Cozen AE, Quartley E, Holmes AD, Hrabeta-Robinson E, Phizicky EM, Lowe TM. 2015. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat Methods* **12**: 879–884.

Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, Muret K, Esquerré D, Zytnicki M, Derrien T, Bardou P et al. 2019. Transcriptome and chromatin structure annotation of liver, CD4+ and CD8+ T cells from four livestock species. *bioRxiv* doi:10.1101/316091

Grosjean H, de Crécy-Lagard V, Marck C. 2010. Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Lett* **584**: 252–264.

Hickey G, Paten B, Earl D, Zerbino D, Haussler D. 2013. HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* **29**: 1341–1342.

Hubisz MJ, Pollard KS, Siepel A. 2011. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief Bioinform* **12**: 41–51.

Johansson L, Gafvelin G, Arnér ESJ. 2005. Selenocysteine in proteins—properties and biotechnological use. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1726**: 1–13.

Jungreis I, Lin MF, Spokony R, Chan CS, Negre N, Victorsen A, White KP, Kellis M. 2011. Evidence of abundant stop codon readthrough in Drosophila and other metazoa. *Genome Res* **21**: 2096–2113.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol* **34**: 1812–1819.

Kutter C, Brown GD, Gonçalves A, Wilson MD, Watt S, Brazma A, White RJ, Odom DT. 2011. Pol III binding in six mammals shows conservation among amino acid isotypes despite divergence among tRNA genes. *Nat Genet* **43**: 948–955.

Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**: 26.

Loughran G, Jungreis I, Tzani I, Power M, Dmitriev RI, Ivanov IP, Kellis M, Atkins JF. 2018. Stop codon readthrough generates a C-terminally extended variant of the human vitamin D receptor with reduced calcitriol response. *J Biol Chem* **293**: 4434–4444.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. 2011. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* **12**: 2825–2830.

Roy B, Leszyk JD, Mangus DA, Jacobson A. 2015. Nonsense suppression by near-cognate tRNAs employs alternative base pairing at codon positions 1 and 3. *Proc Natl Acad Sci U S A* **112**: 3038–3043.

Thornlow B, Hough J, Roger J, Gong H, Lowe T, Corbett-Detig R. 2018. Transfer RNA genes experience

exceptionally elevated mutation rates. *Proceedings of the National Academy of Sciences* **115**: 8996–9001.

Valle RP, Morch MD, Haenni AL. 1987. Novel amber suppressor tRNAs of mammalian origin. *EMBO J* **6**: 3049–3055.

Yacoubi BE, El Yacoubi B, Bailly M, de Crécy-Lagard V. 2012. Biosynthesis and Function of Posttranscriptional Modifications of Transfer RNAs. *Annual Review of Genetics* **46**: 69–95.

Zheng G, Qin Y, Clark WC, Dai Q, Yi C, He C, Lambowitz AM, Pan T. 2015. Efficient and quantitative high-throughput tRNA sequencing. *Nat Methods* **12**: 835–837.