Supplemental Material

Kutter, Brown, Gocalves, Wilson, Watt, Brazma, White, Odom

“Pol III binding in six mammalian genomes reveals high conservation among amino acid isotypes, despite divergence in tRNA gene usage”

SUPPLEMENTAL FIGURES

Figure S1. Distribution of pol III binding in the mouse and human genomes
Figure S2. Pol III occupied tRNA genes are enriched in RNA-seq signals
Figure S3. Pol III occupied tRNA genes occur in clusters
Figure S4. Species-specific occupancy of some loci, despite high evolutionary conservation
Figure S5. Largely the same tRNA genes are occupied by pol III in different mouse tissues
Figure S6. Lineage-specific loss of pol III binding in liver isolated from six mammals aligned to the murine Dnajc5b and Trim55 gene locus
Figure S7. Intersection of all predicted tRNA genes in eutherian genomes
Figure S8. Lineage-specific gain of pol III binding in vivo in liver isolated from six mammals aligned to the murine Olfr750 and Rnase4 gene locus
Figure S9. Codon bias for the six possible triplet codons for leucine is conserved in the six mammals
Figure S10. Triplet codon and amino acid frequency is largely identical in different mouse tissues
Figure S11. Codon usage in mouse tissues is largely identical
Figure S12. Codon usage and pol III binding to tRNA genes correlate in mouse muscle tissue
Figure S13. Codon usage and pol III binding to tRNA genes correlate in mammals
Figure S14. Pol III binding is enriched over tRNA genes in mouse liver

SUPPLEMENTAL TABLES

Table S1. Number of pol III occupied loci (this study and others) in the six mammals
Table S2. Genomic location of pol III-bound mouse tRNA genes
Table S3. Genomic location of pol III-bound rat tRNA genes
Table S4. Genomic location of pol III-bound human tRNA genes
Table S5. Genomic location of pol III-bound macaque tRNA genes
Table S6. Genomic location of pol III-bound dog tRNA genes
Table S7. Genomic location of pol III-bound opossum tRNA genes
Table S8. Number of pol III occupied genes per isotype in the six mammals
Table S9. Spearman correlation between anticodon and isotype gene counts and pol III occupancy in the six mammals
Table S10. Number (percentage) of identifiable tRNA genes within UCSC syntenic blocks (pairwise alignment)
Table S11. Reference genome versions
Figure S1. Distribution of pol III binding in the mouse and human genomes
Pie chart depicts the categories of pol III-occupied loci in per cent and read count (in parentheses) using repeat masker annotation for (A) mouse and (B) human. Transposable elements are further categorized into three groups shown in shades of grey.

Figure S2. Pol III occupied tRNA genes are enriched in RNA-seq signals
The log2 intensities of RNA-seq signals at tRNA genes pol III-bound (black) and -unbound (red) high-confidence tRNA loci (listed as "passed filtering" in Table S1) for two mouse liver replicates are shown. The boxplots show the median as solid horizontal line, the first and third quartiles as the upper and lower bounds of the box, and the whiskers extend to the furthest data point which is no more than 1.5 times the interquartile range. Outliers are shown as circles.
Figure S3. Pol III occupied tRNA genes occur in clusters
The percentage of tRNA genes in clusters of pol III-bound (black) and unbound (grey) loci for the six mammals are illustrated. Numbers above each column show tRNA genes in clusters over the total number of tRNA genes for each category.

Figure S4. Species-specific occupancy of some loci, despite high evolutionary conservation.
This plot shows relative pol III occupancy at three Lysine tRNA genes that are conserved in location across all 6 species in this study. Locus 1 defines a CTT anticodon, while loci 2 and 3 are TTT anticodons. The plot shows that human, macaque and dog have roughly similar occupancy of the three loci, while mouse and especially rat show higher occupancy of locus 2. Opossum has almost completely lost occupancy of locus 3, but compensates with increased occupancy of loci 1 and 2.
Figure S5. Largely the same tRNA genes are occupied by pol III in different mouse tissues

The 3-way VENN diagram intersects the total number of all tRNA genes bound by pol III that passed our filtering criteria (>10 reads on the tRNA gene, and on the 100bp upstream and downstream of the gene). Areas are shaded according to number of tRNA genes (white: low and blue: high).
Figure S6. Lineage-specific loss of pol III binding in liver isolated from six mammals aligned to the murine Dnajc5b and Trim55 gene locus

The genome track of a pol III bound cluster containing four tRNA genes (two arginine and two tyrosine tRNA genes) between the murine Dnajc5b and Trim55 genes and their orthologous in other species shows primate-specific loss of pol III binding to the first tRNA\(^{\text{Arg}}\) gene 3’ downstream of the Dnajc5b gene locus (dashed yellow line). Pol III binding to the two tRNA\(^{\text{Tyr}}\) and second tRNA\(^{\text{Ala}}\) genes are conserved (dashed blue line).
Figure S7. Intersection of all predicted tRNA genes in eutherian genomes
The 5-way VENN diagram includes all the tRNA genes that passed our filtering criteria including both pol III-bound and -unbound tRNA genes. The majority resides in species-specific locations. Species-unique segments contain total number of species-specific tRNA genes (white), which are subdivided into tRNA genes present and absent (dashed outer subdivision) in Ensembl’s 16 amniote alignment. 24 genes are orthologous in all 6 species, including opossum. Areas are shaded according to number of tRNA genes (white: low and blue: high).
Figure S8. Lineage-specific gain of pol III binding in liver isolated from six mammals aligned to the murine Olfr750 and Rnase4 gene locus

The genome track of a pol III bound cluster containing four tRNA genes (one leucine, two tyrosine, and one proline tRNA genes) aligned to the murine Olfr750 and Rnase4 gene locus shows rodent-specific gain of pol III binding to the tRNA\textsubscript{Leu}(TAG) gene 5' upstream of the Olfr750 gene locus (dashed yellow line). This particular leucine tRNA gene could not be found in opossum. The Olfr1637 gene in rat and the OR6S1 gene in human, macaque, dog and opossum are the homologs of the murine Olfr750 gene. Pol III binding to the two tRNA\textsubscript{Tyr} genes is conserved throughout mammalian evolution (dashed blue line), except in dog where these tRNA genes could not be found (dashed grey line) because of gaps in the assembly (black horizontal boxes in this track). Two tyrosine tRNA genes upstream of the OR6S1 gene arose only in the primates and exhibit pol III binding. Additional tRNA genes [tRNA\textsubscript{Pro}, tRNA\textsubscript{Leu}(YAG), tRNA\textsubscript{Pro}, tRNA\textsubscript{Thr} shown in the track and within the 1Mb region not displayed] are present in the opossum genome and lack homologs in the eutherians.
Figure S9. Codon bias for the six possible triplet codons for leucine is conserved in the six mammals
Radar plots represent proportional frequencies of codon usage for leucine. Coloured lines within the radial plot present the data values of each species.

Figure S10. Triplet codon and amino acid frequency is largely identical in different mouse tissues
Triplet codon and amino acid frequency of transcript estimates in mouse liver, brain, and muscle was determined. The intersection of the row-column for each tissue combination in the upper-right triangle shows the correlation for triplet codon frequency (red) and in the lower-left triangle for amino acid frequency (black). The Pearson correlation coefficients are reported in the lower right of each inter-tissue panel.
Figure S11. Codon usage in mouse tissues is largely identical
Radar plots represent proportional frequencies of (A) amino acids weighted by expression, codon usage for (B) leucine and (C) arginine obtained from RNA-seq signals from different mouse tissues. Labels around the plot indicate amino acids (A) and triplet codons (B and C). Labels within the grid of the radial plots describe percentages.

Figure S12. Codon usage and pol III binding to tRNA genes correlate in mouse muscle tissue
Plot correlates the binding of pol III to tRNA isotypes with the amino acid frequency weighted by expression obtained from mRNA-seq data in mouse muscle tissue. Pearson correlation (R) and Spearman’s rank correlation coefficient (rho) is reported in the lower right of the panel.
Figure S13. Codon usage and pol III binding to tRNA genes correlate in mammals
Plots correlate the binding of pol III to tRNA isotypes with the amino acid frequency weighted by expression obtained from mRNA-seq data in liver of (A) human, (B) macaque, (C) mouse, (D) rat, (E) dog, and (F) opossum. The Pearson correlation (R) and Spearman correlation (rho) are reported in the lower right of each panel.

Figure S14. Pol III binding is enriched over tRNA genes in mouse liver
The number of reads was compared in immunoprecipitation experiments with (1900) and without (input) an antibody against pol III in two biological replicates of mouse liver. The distance from the 5’ end of the tRNA gene (position 0) is plotted on the x-axis and normalized read count at each nucleotide is plotted on the y-axis. The density plot shows the enrichment of pol III (red and black line) above the input control (dashed blue and yellow line) on tRNA genes. Shoulders at position +/- 500 bp are caused by neighbouring tRNA genes; if each read is counted only once (i.e. for one of its two flanking genes), the density plot is flat beyond about 150bp from the gene.
tRNAscan-SE marks some predicted loci as probable pseudogenes, and others as probably tRNA genes, but with an undetermined anticodon. In addition, each locus has an associated score. Scores less than 50 are considered “probably not real”. The remainder (meeting none of these criteria) are considered to have “passed filtering” and are candidates for being transcribed genes. Percentages in parentheses are of the “pol III occupied tRNAs”.

(Note that Table S2-S7 are provided as supplemental files in Excel format)

### Table S1. Number of pol III occupied loci (this study and others) in the six mammals

|                           | Mmu | Rno | Hsa | Mml | Cfa | Mdo |
|---------------------------|-----|-----|-----|-----|-----|-----|
| Pol III occupied tRNAs   | 223 | 282 | 224 | 233 | 133 | 216 |
| Pseudogenes               | 20915 | 140962 | 102 | 105 | 70078 | 14184 |
| Passed filtering          | 437 | 515 | 445 | 521 | 553 | 480 |
| Genomic location          |     |     |     |     |     |     |
| Mouse                     | 219 (98%)/ 433 | 216 (77%)/ 444 | 224 (100%)/ 625 | 228 (96%)/ 379 | 130 (98%)/ 906 | n.a. |
| Coughlin et al.           | 209 (94%)/ 492 |       |     |     |     |     |
| Mouse                     | 136 (61%)/ 341 |       |     |     |     |     |
| Oler et al.               | 207 (92%)/ 314 |       |     |     |     |     |
| Raha et al.               | 178 (80%)/ 257 |       |     |     |     |     |

Table S2. Genomic location of pol III-bound mouse tRNA genes
Table S3. Genomic location of pol III-bound rat tRNA genes
Table S4. Genomic location of pol III-bound human tRNA genes
Table S5. Genomic location of pol III-bound macaque tRNA genes
Table S6. Genomic location of pol III-bound dog tRNA genes
Table S7. Genomic location of pol III-bound opossum tRNA genes
Table S8. Number of pol III occupied genes per isotype in the six mammals

| isotype | Mmu | Rno | Hsa | Mml | Cfa | Mdo |
|---------|-----|-----|-----|-----|-----|-----|
| Ala     | 12  | 39  | 13  | 13  | 8   | 11  |
| Arg     | 21  | 21  | 16  | 19  | 14  | 13  |
| Asn     | 6   | 7   | 10  | 11  | 4   | 6   |
| Asp     | 9   | 6   | 10  | 10  | 5   | 9   |
| Cys     | 6   | 5   | 5   | 6   | 5   | 5   |
| Gln     | 8   | 6   | 11  | 10  | 5   | 5   |
| Glu     | 14  | 11  | 15  | 11  | 6   | 13  |
| Gly     | 20  | 16  | 17  | 13  | 9   | 15  |
| His     | 5   | 4   | 6   | 7   | 3   | 4   |
| Ile     | 9   | 8   | 9   | 11  | 7   | 9   |
| Leu     | 18  | 18  | 16  | 19  | 10  | 45  |
| Lys     | 12  | 13  | 13  | 14  | 9   | 11  |
| Met     | 10  | 9   | 12  | 13  | 6   | 10  |
| Phe     | 6   | 5   | 5   | 5   | 3   | 5   |
| Pro     | 13  | 54  | 13  | 14  | 8   | 12  |
| SeC     | 1   | 1   | 1   | 1   | 1   | 1   |
| Ser     | 19  | 27  | 13  | 17  | 9   | 13  |
| Thr     | 12  | 11  | 11  | 11  | 7   | 10  |
| Trp     | 5   | 4   | 4   | 6   | 4   | 4   |
| Tyr     | 8   | 5   | 6   | 8   | 4   | 7   |
| Val     | 9   | 12  | 18  | 14  | 8   | 8   |
### Table S9. Spearman correlation between anticodon and isotype gene counts and pol III occupancy in the six mammals

| species | anticodon | isotype |
|---------|-----------|---------|
| Mmu     | 0.91      | 0.93    |
| Rno     | 0.81      | 0.88    |
| Hsa     | 0.67      | 0.89    |
| Mml     | 0.72      | 0.82    |
| Cfa     | 0.73      | 0.82    |
| Mdo     | 0.79      | 0.91    |

### Table S10. Number (percentage) of identifiable tRNA genes within UCSC syntenic blocks (pairwise alignment)

| species | Mmu | Rno | Hsa | Mml | Cfa | Mdo |
|---------|-----|-----|-----|-----|-----|-----|
| Mmu     | 223 (100%) | 184 (65%) | 148 (66%) | 154 (66%) | 109 (81%) | 95 (44%) |
| Rno     | 184 (83%) | 282 (100%) | 126 (56%) | 140 (60%) | 106 (79%) | 59 (27%) |
| Hsa     | 148 (66%) | 126 (45%) | 224 (100%) | 187 (80%) | 114 (84%) | 96 (44%) |
| Mml     | 154 (69%) | 140 (50%) | 187 (83%) | 233 (100%) | n.a. | 75 (35%) |
| Cfa     | 109 (49%) | 106 (38%) | 114 (51%) | n.a. | 135 (100%) | 75 (35%) |
| Mdo     | 95 (43%) | 59 (21%) | 96 (43%) | 75 (32%) | 75 (56%) | 216 (100%) |

Percentages in parentheses are of the column-heading genome. UCSC does not provide pairwise BlastZ alignments for dog and macaque, so these numbers are unavailable (n.a.).

### Table S11. Reference Genome Versions

| Species                | Abbreviation | Common name | Genome version                  |
|------------------------|--------------|-------------|----------------------------------|
| Mus musculus           | Mmu          | mouse       | MGSCv37/ mm9/ NCBI v37           |
| Rattus norvegicus      | Rno          | rat         | rn4/ NCBI v4                     |
| Homo sapiens           | Hsa          | human       | GRCh37/ hg19/ NCBI v36.3         |
| Macaca mulatta         | Mml          | rhesus macaque | rheMac2/ NCBI v1               |
| Canis familiaris       | Cfa          | dog         | canFam2/ NCBI v2                 |
| Monodelphis domestica  | Mdo          | opossum     | monDom5/ NCBI v1                 |