ORIOLI *et al* – SUPPLEMENTARY DATA

SUPPLEMENTARY MATERIALS AND METHODS

Analysis of viral transcription units

The mouse fibroblast cell line 3T3 (ATCC CRL-1658) was cultured in complete DMEM (Gibco) supplemented with 5% FBS (Hyclone), 2 mM L-glutamine, 10 U/ml penicillin, and 10 μg/ml streptomycin sulfate. γHV68 strain WUMS (ATCC VR-1465) was used for all infections. Virus stocks used for infection were passaged, grown, and titer determined as previously described (Virgin et al., 1997). Infections were carried out at a multiplicity of infection (MOI) of 5 plaque forming units (PFU) per cell. Total RNA was extracted from 3T3 cells 36 hours after infection using the *mir*Vana™ miRNA isolation kit (Applied Biosystems). 5' triphosphate-containing RNA molecules were enriched from the total RNA pool as follows: 3μg of total RNA was incubated with 25U of Antarctic Phosphatase (New England Biolabs) and 40U of Rnasin (Promega) at 37°C for 1 hour. The phosphatase reaction was stopped by heat-inactivation at 70°C for 10 minutes. After ethanol/sodium acetate precipitation, purified RNA was resuspended in 100 μl of RNase-free water. Phosphatased RNA (1μg) was incubated with 40 U of RNasin and 10 U of T4 Polynucleotide Kinase (New England Biolabs) and T4 DNA Ligase Buffer with 10mM ATP (New England Biolabs) at 37°C for 30 min. After heat inactivation of the kinase reaction, RNA was precipitated as described. The entire amount of kinased RNA recovered (~1μg) was used as template in the inverse RT-PCR reaction. To circularize the triphosphate enriched RNA molecules, 1 μg of modified RNA was incubated with 40U of RNasin, 10U of T4 RNA ligase (New England Biolabs) and T4 RNA ligase reaction buffer, in a 50-μl total volume at 37°C for 1 hour. The ligase reaction was heat inactivated at 70°C for 20 minutes. Primer sets specific to the γHV68 pol III transcripts (listed in **Table S2**) were used for inverse RT-PCR amplification of the circularized RNA molecules. RT-PCR amplification reactions were performed in a 25-μl reaction mix using the OneStep RT-PCR Kit (Qiagen) containing each primer at 0.5μM. Circularized RNA (10 μl, ~200 ng) was treated with RQ1 RNase-Free DNase (Promega) prior to addition to the RT-PCR reaction mix. RT-PCR reactions were performed as follows: 30 min at 50°C; 15 min at 95°C; 40 amplification cycles of 30 sec at 94°C followed by 30 sec of 48-56°C (dependent on the Tm of the primer set) followed by 1 min at 72°C and a single final incubation for 10 min at 72°C. Once complete, 18 μl of the 25-μl RT-PCR reaction mixture were resolved on a 3% TAE agarose gel. Products were gel-purified and cloned into the pCR4-TOPO vector using the TOPO-TA Cloning Kit for Sequencing (Invitrogen). Individual colonies were then grown for plasmid DNA recovery and sequencing.

Table S1. tRNA genes and pseudogenes with unusually distant Pol III canonical terminators employed in the *in vitro* transcriptional assays, as annotated in the HUGO Gene Nomenclature Committee database (http://www.genenames.org/) or in the genomic tRNA database (GtRNAdb) (http://lowelab.ucsc.edu/GtRNAdb/Hsapi/).

tRNA gene (HGNC)	tRNA gene name (GtRNAdb)	tRNA gene position[a]	A-box score[b]	B-box score[b]	Sequence[c] (and predicted length of canonically terminated transcript)[d]
TRNAR9	chr6.trna114-ArgCCG	chr6:28710728-28710801 (-)	-14.03	-3.62	ACCCGCGTGC**GGCCGCGTGGCCTAATGGATAAGGCGT CTGATTCCGGATCAGAAGATTGAGGGTTCGAGTCCCT TCGTGGTCGT**GTTCTTACTATTGTCAGGAAATATCTT TACTTTCTGCACATCTCCTTGAGCTTTCTGCAAGCCA GTTAAAAAAAAAAAAACCGGCTTGTTCCCAAGACTTT GGGAACTTGAAATGAAAGAAGACTATTGACAAAAAAA ACAAGAACCCGGTAAGCAACCTGTGTATTGACTATTA AAGTGCACTTTGTCCCTGGCATAGGTACTACCGGAGG CATTCGTAATTTGGAATGGGCTTGGTTTTTCAATTTA TG (321 nt)
TRNAV18	chr6.trna115-ValAAC	chr6:28703206-28703277 (-)	-9.50	-3.21	CATTGTAAGT**GGGGGTGTAGCTCAGTGGTAGAGCGTA TGCTTAACATTCATGAGGCTCTGGGTTCGATCCCCAG CACTTCCA**CAAGTACATTTCCTTATATCTATCGTTGA ACATTTCGTTTCAGTCTTTGAATACTTAGTTCCAGTC TTTTCTTTTTTCTT (147 nt)
TRNAG20	chr16.trna19-GlyGCC	chr16:70823410-70823480 (+)	-11.19	-2.17	CACATCTGGG**GCATTGGTGGTTCAGTGGTAGAATTCT CGCCTGCCACGCGGGAGGCCCGGGTTCGATTCCCGGC CAATGCA**GCAGCAGACCTTTAGTTTAATCGTGAGAGA CTGAATTTGAGATTTCACACTGCCTAAAGAATTTATA AATCCACTCTGAGAACTGAGAATGACTTGTTTCTAGT TGGATTTCAAATGCCTTTGATGAAAACACTTATTTTC CTAAATTCT (216 nt)
TRNAW6	chr7.trna1-TrpCCA	chr7:99067307-99067378 (+)	-15.95	-9.90	AGGTTTCCAA**GACCTCGTGGCGCAACGGCAGCGCGTC TGACTCCAGATCAGAAGGTTGCGTGTTCAAATCACGT CGGGGTCA**AACATGTGATCATCCCTTTCTTTAGCCAA CAGGAGTTATAGCTTGCCGGTAACTCTCAACCTTGGA CACTGTATTTTCTGTTCTTTA (154 nt)
TRNAT18	chr6.trna60-ThrAGT	chr6:27694473-27694546 (+)	-12.71	-1.92	ACACAGGTGA**GGCTTCGTGGCTTAGCTGGTTAAAGCG CCTGTCTAGTAAACAGGAGATCCTGGGTTCGAATCCC AGCGAGGCCT**CTTTATTTCTTCCCCTAAACTTAGGTA AATTCTTGTCACTAGTTAAACCTGACTTAGATGTATA CCTAATAAAGCAGCAAACCTTGAATCTCGATGTCTGA GGCACTTTGTAGAGGAACGTGGTGATGGGTTGGAAAT ATTTACGATGAATTAGCTAAATCTATGCCAGAGACCA ACCGTCTAGAACCATCTAGAATTGAGCTGCAATTTTC TACGTTTAA (290 nt)
TRNAH11	chr1.trna106-HisGTG	chr1:149155828-149155899 (-)	-16.94	-1.92	CCTGTGGCTC**GCCATGATCGTATAGTGGTTAGTACTC TGCGCTGTGGCCGCAGCAACCTCGGTTCGAATCCGAG TCACGGCA**ATGTCGTTCTTCCAGGCCGTCAGCTATCC ACTTTCGCTCCCTGCAACCAGGCGCCTCAGGGAAAGG AAAAGAGACCTCACAGCCCACTAACCTGGCGAGGAAT CCAAGGGAAGCGCTGGACGTGCTCCGTCCATTTCCGA GCCTCACACAGCAAATAACCCAGTCAGCCCTGTTGGT GGCTTACACGAAGGCAATGGTGCCCTCAGAACCTGTT TTCCAGTCACTG (293 nt)
TRNAH3	chr6.trna33-HisGTG	chr6:27125906-27125977 (+)	-16.94	-1.92	CTGGTGGCGT**GCCGTGATCGTATAGTGGTTAGTACTC TGCGTTGTGGCCGCAGCAACCTCGGTTCGAATCCGAG TCACGGCA**GTACCTTGATGTCGCCTCAATTTCTCAAC GTACTGAGCAGTACCTTGACGTCGCCTCAATTTTTCA ACATACT (141 nt)

TRNAL12	chr6.trna126-LeuAAG	chr6:28446400-28446481 (-)	-15.68	-7.35	GGACTAGCGT**GGTAGCGTGGCCGAGTGGTCTAAGACG CTGGATTAAGGCTCCAGTCTCTTCGGGGGCGTGGGTT TGAATCCCACCGCTGCCA**GGTTTATTGCAAAATTAGT AGGGATATCTAAAGCTCATTGAGATTTGCAAGAAAAA CATTTTCCACTGCGAG (149 nt)

[a]Coordinates are the boundaries of the gene (Ensemble release 58, human assembly GRCh37).

[b] A-box and B-box scores were calculated with Pol3scan (Pavesi et al. 1994, Nucleic Acids Res 22: 1247; Percudani et al. 1997, J Mol Biol 268:322).

[c] The sequence include 10 bp upstream and 10 bp downstream of the predicted full-length pre-tRNA.

Bold: tRNA coding sequence. Underlined: predicted Pol III non-canonical termination signals.

Underlined/red: predicted Pol III canonical termination signals ($T_{\geq 4}$).

[d] The lengths of canonically terminated pre-tRNAs are calculated including all the Ts of the termination signal and 5 nt of 5' leader sequence

Table S2. Primers and Northern probes list

tRNA gene (HGNC) primers	Primer sequence (5' to 3')
TRNAR9_for	ATCTCTAATTACGAAGTTAAAGGT
TRNAR9_rev	ATTCATGAAACAAGTAGGACAAGA
TRNAV18_for	CTTGGAACTTCGCTCCAGGTGGA
TRNAV18_rev	GGAGTGTTAGGACCCATAGTACA
TRNAG20_for	AGGTCTGAAGGCAAATCAACTGGAC
TRNAG20_rev	CTGCTTTAGCGGGCCTAGGTGCAT
TRNAW6_for	ATTATTCCCGAACCCCGTGGTGG
TRNAW6_rev	AGCTAAGGTGCCGGTCCACTGC
TRNAT18_for	TTCAGCACCTATTCTTATTCTACAG
TRNAT18_rev	CACAACACTCAATTTATACCATATG
TRNAH11_for	TATGAAAGAATAGCGCGATCCCCCT
TRNAH11_rev	ATACAAACTTGGCGACGAGCTGGG
TRNAH3_for	AGCAAAATATTGGGGACGGG
TRNAH3_rev	GCTTGGGTAGGGTTGGGAAG
TRNAL12_for	CAGAGTCCTCAGAGTCCCCG
TRNAL12_rev	TACCTTGTGCTCGCAGTGGA
RNA target	**Northern probe sequence (5' to 3')**
γHV68 miR-M1-5	GAGACGACCCGATCTCAACTCT
γHV68 tRNA4	CCGCTCTACCAATTGAGCTACC
U6 snRNA	CGAATTTGCGTGTCATCCTTGC
γHV68 pol III transcripts	**Primer sets for inverse RT-PCR amplification (5' to 3')**
γHV68 pol III-1	sense: AGAAATGGCCGTACTTCC antisense: CCAGAGCTCGGACTTG
γHV68 pol III-4	sense: GCAAACCCGAGCTCCTC antisense: CGGGACCCGGGATTGACC
γHV68 pol III-5	sense: GGGATATCGCGCCCACC antisense: CCGGAACCGACAGGATAC

Table S3. gHV68 tRNA4-miR-M1-5-like transcriptional unit

GCGGCCGCAACCGCGCCGATAAGCTTCCGGGGTCCCAACCCTGTAAGGCAGAACGGCTGCCCAGCT
CGGATGACGCCACACTAACGTAGCCTCCAGACCGCCCAGTGTGGGTGTGTCCAAGCTCACGTCGCC
CGGCGTGGCCCCCCCGCTCCCCAATGACGTAACTGCCCTGCAGCTTCTAGTAGCTTTTCGCAGCGT
CTCCGACCGTCGGGGTAGCTCAATTGGTAGAGCGGCAGGCTCATCCCCTGCAGGTTCTCGGTTCAA
TCCCGGGTCCCGACG(*)CAGAGTTGAGATCGGGTCGTCTCCCCCTGGCGGAAGGAGGCAAACC
CGAGCTCCTCCTTCTTTTTCTCGAG

(*) indicates the termination signals as described in the text: T_5, T_2AT_3, T_3GT_2, TCT_3, T_3CT, T_2AT_2, T_2GT_2 and T_3, respectively. Underlined is the UAS of the 7SL RNA gene (Hs7SL-1, Englert et al. 2004, Biochimie 86:867).

Figure S1

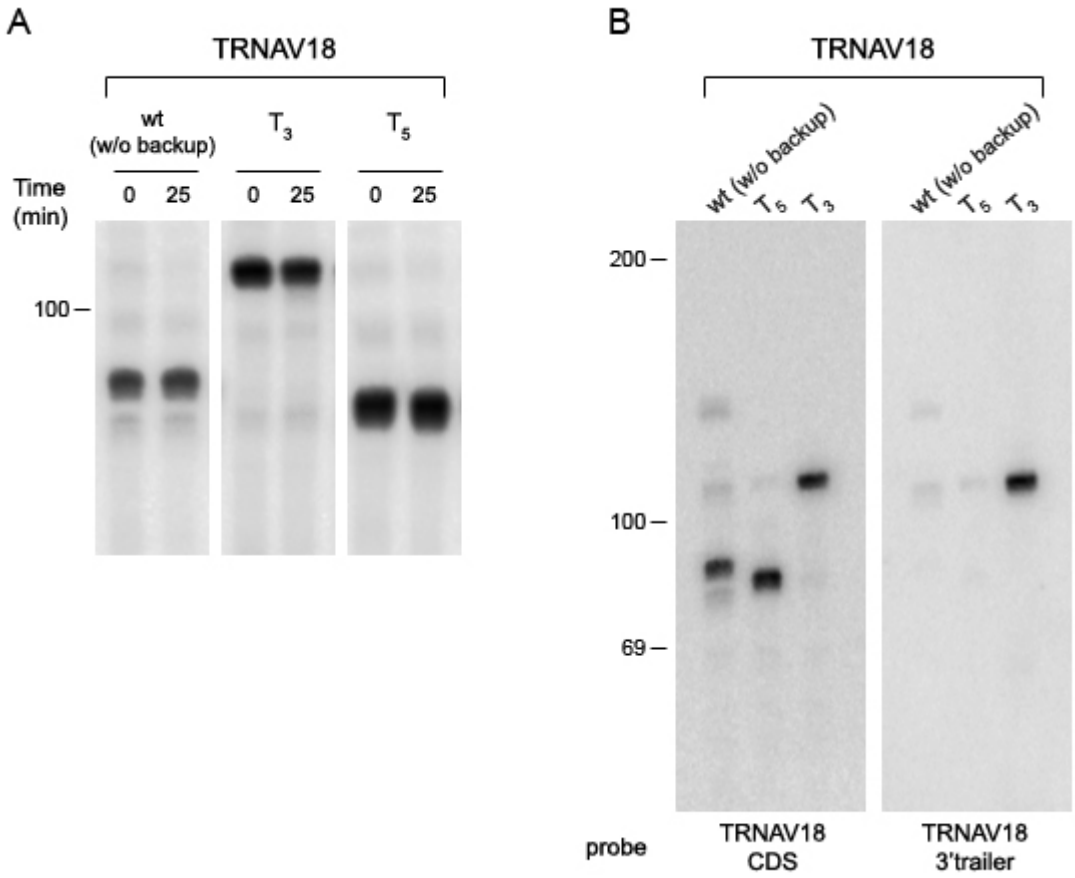
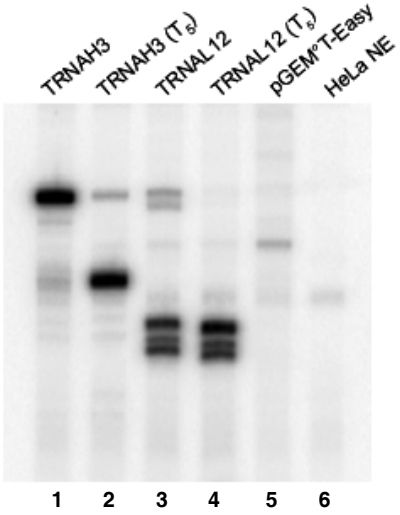


Figure S1. Absence of pre-tRNA processing or degradation under *in vitro* transcription conditions. (*A*) Pulse-chase experiment. *In vitro* transcription of wt TRNAV18 and of its modified variants containing either a T_5 or a T_3 as the first encountered termination sequence, followed by a backup terminator (cf. Figure 3), was carried out in a HeLa nuclear extract under standard conditions, then the reaction was either stopped (time 0) or chased with 2 mM UTP for 25 minutes. Processing of all RNA species produced during the first incubation period resulted to be absent or negligible within a further 25-min time window (low levels of non-specific degradation, more marked for the transcript containing the long 3' trailer, were observed at longer chase times). (*B*) Northern analysis of cold transcripts produced by *in vitro* transcription under standard conditions with the wt, T_3-containing and T_5-containing TRNAV18 templates. No RNA species, except the ones expected as primary transcripts with the different templates, were observed.

Figure S2



TRNAH3 (*chr.6 27233885-27233956)* 3'-flanking sequence

…AATTTTTCA…

ACCTTGATGTCGCCTCAATTTCTCAACGTACTGAGCAGTACC
TTGACGTCGCCTCAATTTT

TRNAL12 (*chr.6 28554460-28554379)* 3'-flanking sequence

…GGTTTTTGC…

GGTTTATTGCAAAATTAGTAGGGATATCTAAAGCTCATTGAGA
TTTGCAAGAAAAACATTTT

Figure S2. Analysis of transcription termination at TRNAH3 and TRNAL12 genes.
The TRNAH3 and TRNAL12 genes were transcribed *in vitro* using a HeLa cell nuclear extract. TRNAH3 has a T_5 located 58 bp downstream of the end of the coding sequence, preceded by a T_3CT motif located ~35 bp upstream; TRNAL12 has a T_4 located 58 bp downstream of the end of the coding sequence, preceded by a T_3AT_2 motif located ~50 bp upstream. The sequences of the 3'-flanking regions of both tDNAs are reported on the right, together with the mutations introduced to improve the non-canonical terminator.
Note the very poor termination at the T_3CT motif of TRNAH3 (lane 1), to be compared with 8% termination produced by T_3CT in the TRNAV18 context (Figure 3); and, in contrast, the high termination efficiency at the T_3AT_2 motif of TRNAL12 (lane 3), to be compared with 62% termination in the TRNAV18 context (Figure 3). As expected, replacement of the two non-canonical terminators with T_5 restored almost complete termination at the upstream position for both tRNA genes (lanes 2 and 4). Transcript size heterogeneity observed with TRNAL12 is most likely due to multiple transcription start sites, rather than to heterogeneous termination, because it did not change when the native terminator was mutated to T_5.

Figure S3

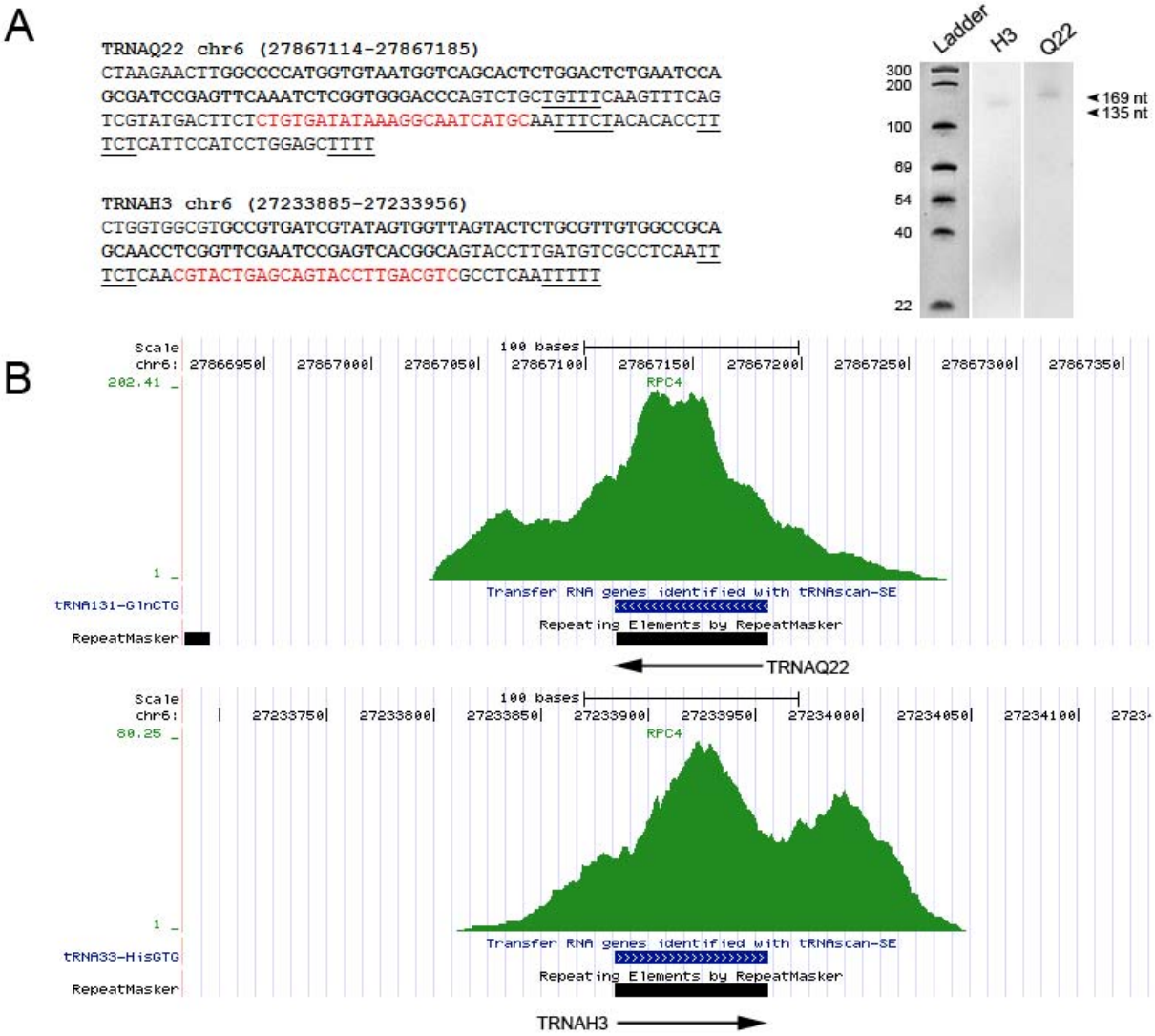


Figure S3. ***In vivo*** **analysis of Pol III read-through at individual tRNA gene terminators**
(*A*) Left: Sequence of TRNAQ22 and TRNAH3, including 10 bp of 5'-flanking sequence, the tRNA coding sequence (in bold characters) and the 3'-flanking sequence up to the closest T_5 termination signal. Canonical and non-canonical terminators are underlined; the sequences complementary to the oligonucleotide DNA probes are in red characters. Right: Northern blot analysis of total RNAs extracted from HeLa cells with probes complementary to the 3'-flanking sequences. The approximate lengths of hybridizing pre-tRNA transcripts are indicated on the right. (*B*) Trailing of RNA polymerase III downstream of two tRNA genes. UCSC genome browser view of the indicated region of chromosome 6 with the TRNAQ22 and TRNAH3 loci. The visual peak tracks were generated as described previously for POLR3D (Canella et al, 2010). The Y axis indicates cumulated tag weights. The arrows indicate the direction of tRNA gene transcription.

Figure S4

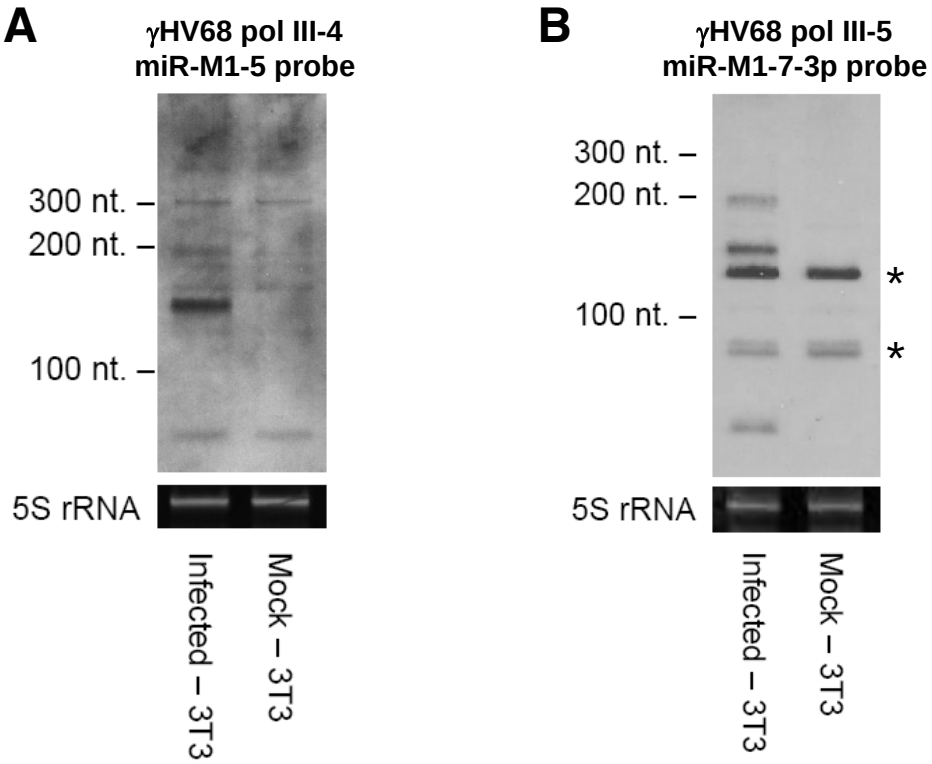


Figure S4. Analysis of the pol III-4 and pol III-5 transcription units of γHV68
(*A*) Northern blot of γHV68 pol III-4 from total RNA isolated from infected 3T3 cells collected at 24 hours post-infection, using a probe specific for the miR-M1-5 portion of the transcript. Ethidium bromide-stained 5S rRNA shown below serves as a loading control. The size of the predominant band (~135 nt) corresponds to the one predicted for a transcript terminated at T_4 between miR-M1-5 and miR-M1-6. The fainter 200-nt band corresponds to a transcript generated by Pol III reading through T_4 and terminating at the downstream located T_6 (see Figure 6). (*B*) Northern blot of γHV68 pol III-5 from total RNA isolated from infected 3T3 cells collected at 36 hours post-infection, using a probe specific for the miR-M1-7-3p portion of the transcript. Ethidium bromide-stained 5S rRNA shown below serves as a loading control. Three clearly detectable transcripts are produced: one of about 200 nt in length, likely generated by termination at the most downstream located T_4; one of about 135 nt, likely generated by termination at the non-canonical TTATT signal; and one of about 60 nt, likely corresponding to pre-miR-M1-7 (see Figure 6). Asterisks indicate non-specific transcripts detected by the probe in both infected and control samples.

Figure S5

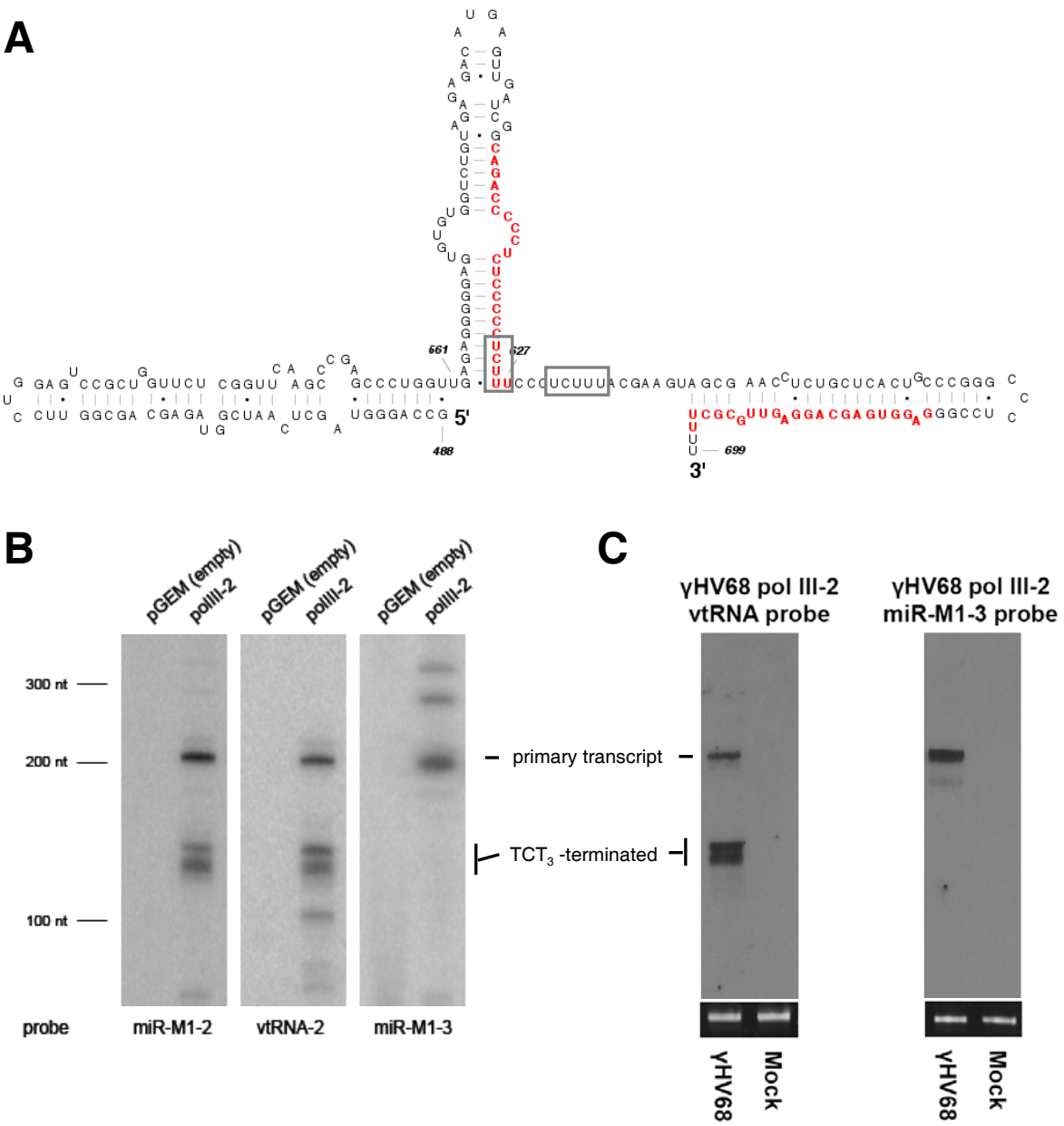


Figure S5. Analysis of the Pol III-2 transcription unit of γHV68

(*A*) Predicted secondary structure of the γHV68 pol III-2 primary transcript containing the lowest ΔG°37 value using the RNAstructure 4.6 software (Mathews et al. 2004, Proc Natl Acad Sci USA 101:7287-92). Nucleotides comprising miR-M1-2 and miR-M1-3 are shown in red. The two TCT_3 non-canonical termination signals are enclosed in rectangles. (*B*) Plasmid-borne Pol III-2 transcription unit (or empty pGEM-T Easy plasmid) was *in vitro* transcribed using a HeLa cell nuclear extract in the absence of any radiolabeled nucleotides. Transcription products were then analyzed by Northern hybridization, using the specific probes indicated below the gel image. (*C*) Northern blot of γHV68 pol III-2 from total RNA isolated from infected 3T3 cells collected at 24 hours post-infection, using probes specific for either the tRNA-like portion of the transcript (left panel) or miR-M1-3 (right panel). Ethidium bromide-stained 5S rRNA shown below serves as a loading control.

The faster migrating doublet in both Northern blots was interpreted as corresponding to transcripts terminated at each of the two TCT_3 signals; the slower migrating RNA corresponds to the primary transcript generated by polymerases that read through the TCT_3 signals and ends at the T_4 element located just downstream of the miR-M1-3 sequence. The low-abundance miR-M1-3 was not detectable by Northern blot, as expected (Diebel et al 2010, *RNA* **16**:170-185)