Prediction of conserved microRNA targets, microRNA suppression of immediate-early viral genes, and implications for herpesvirus latency

Jiří Vaníček[1]*, Ray Marin[1], Eian Murphy[2], Thomas Shenk[2], Harlan Robins[3], Arnold Levine[3]

[1] Laboratory of Theoretical Physical Chemistry, Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland, [2] Princeton University, Princeton, NJ, [3] Institute for Advanced Study, Princeton, NJ

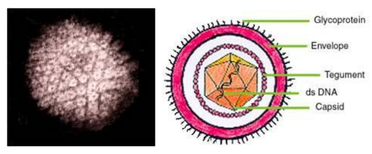
*jiri.vanicek@epfl.ch

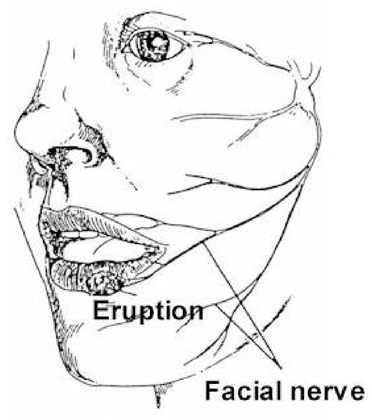
Abstract

A target prediction algorithm is introduced and applied to predict target genes of microRNAs encoded by herpesviruses [1] and by the human [2]. The algorithm is based on the over-representation of complementary sites conserved among related species or viral strains. While there is little conservation among microRNAs of different herpesvirus subfamilies, a common pattern of regulation emerged: the algorithm predicts that human cytomegalovirus, Epstein-Barr virus, and Kaposi's sarcoma-associated herpesvirus all employ microRNAs to suppress expression of their own genes, including their immediate-early genes. In human cytomegalovirus, a virus-coded microRNA, miR-112-1, was predicted to target the viral immediate-early protein 1 mRNA. To test this prediction, mutant viruses were generated that were unable to express the microRNA, or encoded an immediate-early 1 mRNA lacking its target site. Analysis of RNA and protein within infected cells demonstrated that miR-UL112-1 inhibits expression of the major immediate-early protein. We propose that herpesviruses use microRNA-mediated suppression of immediate-early genes as part of their strategy to enter and maintain latency. In the case of human genome, a dataset of 15'806 experimentally verified miRNA-mRNA interactions allowed a detailed characterization of the conservation filter, confirming that strengthening the filter increases the precision of the algorithm [2,3]. In comparison with other algorithms, ours was the most precise while maintaining a high sensitivity.

What are herpesviruses?

- Example: Herpes simplex virus type 1:
  - LYTIC INFECTION: cold sores
  - LATENT INFECTION: virus dormant, no proteins made, "invisible" to the immune system
- REACTIVATION
- Regulation of latency = one of the poorly understood processes in viral biology





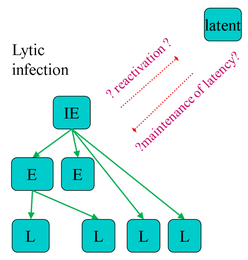
Human herpesviruses

- **Herpes simplex virus type 1 (HSV-1)**: cold sores
- **Herpes simplex virus type 2 (HSV-2)**: genital herpes
- **Varicella zoster virus (VZV)**: primary = chicken pox, reactivation = shingles
- **Epstein-Bar virus (EBV)**: infectious mononucleosis, Burkitt's lymphoma, nosopharyngeal carcinoma
- **Human cytomegalovirus (HCMV)**: mononucleosis-like syndrome, retinitis, neonatal infection ⇒ birth defects
- **Kaposi's sarcoma-associated herpesvirus (KSHV)**: Kaposi's sarcoma, primary effusion lymphoma
- **HHV6 and HHV7**: roseola

Latent and lytic expression cascades

4 types of genes:

- **Latent genes**
- **Immediate early genes (IE)**:
  - transcribed in presence of inhibitors of translation
  - activate E and L genes
- **Early genes (E):** expressed in presence of inhibitors of DNA replication
- **Late genes (L):** depend on DNA replication



Discovery of microRNAs in human herpesviruses: HSV-1, EBV, HCMV, KSHV

MicroRNA target prediction algorithm

Assumptions

- 3'UTR sequence **co-evolves** with miRNA sequence ⇒ expect **over-representation** of seed oligomer in regulated 3'UTR compared to the background sequence [4]
- **Background** 3'UTR sequence based on 1st order local **Markov model** (preserve mono- and di-nucleotide content of each 3'UTR separately)

Reasons:

- Avoid CpG under-representation
- Sufficient statistics
- Avoid local variation of sequence composition

Predicted seed frequency p in the background:

$$p = p(X_n X_{n-1} \dots X_1) = p(X_n|X_{n-1}) \dots p(X_2|X_1) f(X_1) = \frac{f(X_n X_{n-1}) \dots f(X_2 X_1)}{f(X_{n-1}) \dots f(X_2)}$$

Ranking and significance

- For each 3'UTR and miRNA compute:
  l= 3'UTR length
  p = **expected** seed frequency
  c = **actual** seed count
- For each 3'UTR-miR combination compute $\mathbf{PV_{SH}}$ that the match would occur by chance
- **Rank** predictions according to PV_{SH}

$$\mathrm{PV}_{\mathrm{bin}}(l, c, p) = \sum_{i=c}^{l} \binom{l}{i} p^i (1-p)^{l-i}$$

Immediate early genes predicted to be microRNA targets. Model of latency

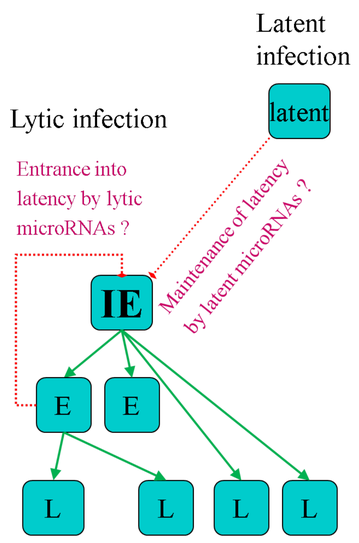
Prediction of the algorithm: miRNAs inhibit major immediate early genes in 3 human herpesviruses!

Virus	3'UTR	Length	miRNA	Seed	Count	Rank	Percentile
EBV	BZLF1, BRLF1	53	ebv-miR-BART15	2-8	1	3 of 2720	99.89
EBV	BZLF1, BRLF1	53	ebv-miR-BHRF1-3	2-8	1	4 of 2720	99.85
HCMV	IE1	92	hcmv-miR-UL112-1	2-8	1	10 of 4896	99.80
KSHV	Zta, Rta	1144	kshv-miR-K12-6-3p	3-8	4	1 of 1394	99.93

We propose that **LATENCY ENTERED INTO and/or MAINTAINED** by repression of immediate early genes by miRNAs:

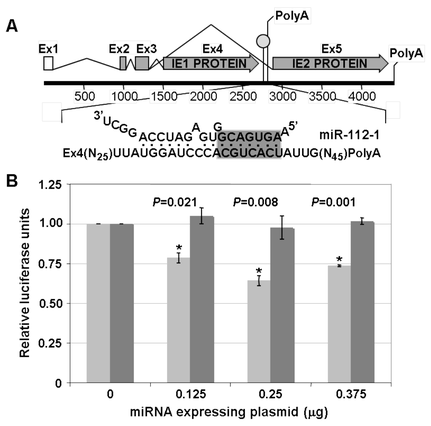
- Explains mysterious absence of protein expression in latency
- MicroRNAs expressed during latency
- Targeting IE genes the most efficient way to maintain latency
- Simple

Proposed model of latency:



Experimental confirmation of the prediction in the human cytomegalovirus

miR-UL112-1 inhibits expression from a reporter mRNA containing the IE1 3'UTR:



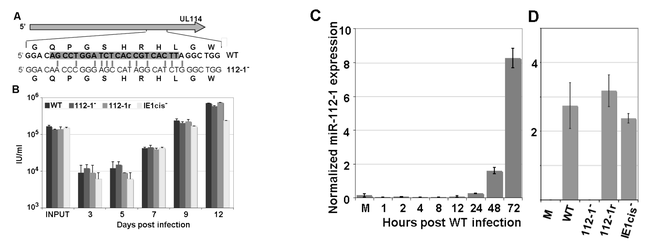
(A) The predicted miR-UL112-1 binding site within the HCMV major IE locus

(B) Reporter assay for miR-UL112-1 function

Wild type IE 3'UTR (light gray)
Mutant IE 3'UTR (dark gray)
Details: 293T cells were co-transfected with:

1. firefly luciferase expression plasmids containing either the wild-type or the mutant IE 3'UTR as well as a Renilla luciferase internal control
2. indicated amounts of a miR-UL112-1 expressing plasmid

UL112-1- and IE1 cis- mutants grow with wild type kinetics:



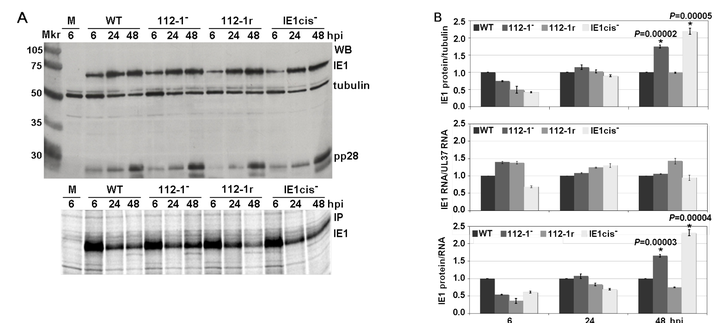
(A) Nucleotides mutated in the 112-1 mutant relative to WT

(B) Growth of virus in MRC5 fibroblasts. IU = infectious units

(C) Accumulation of the miRNA in WT-infected MRC5 fibroblasts over time

(D) Accumulation of miR-112-1 at 48 hpi

Viruses that lack mir-UL112-1 or its binding site synthesize more IE1 protein:



MRC5 fibroblasts were infected with the specified virus. Cells were 35S-labeled for 1h before harvesting at the indicated times after infection

(A) -upper: Western blot of IE1, the late virus-coded pp28 and tibulin

(A) -lower: Immunoprecipitation followed by electrophoresis for 35S-labeled IE1.

(B) -upper: Quantification of 35S-labeled IE1 relative to tubulin

(B) -middle: Quantification of IE1 RNA relative to UL37 RNA by quantitative RT-PCR

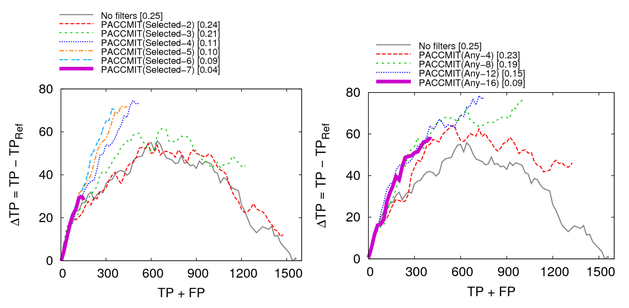
(B) -lower: Ratio of IE1 protein to IE1 mRNA

Notation:
M = mock infected cells
WT = wild type virus
112-1- = microRNA mutant
112-1r = revertant of 112-1

IE1cis- = mutant lacking the 7-nucleotide seed sequence in IE1 mRNA

Prediction of conserved targets in the human. Comparison with other methods

Conservation filter: more species ⇒ more strict ⇒ higher ΔTP ⇒ higher precision!

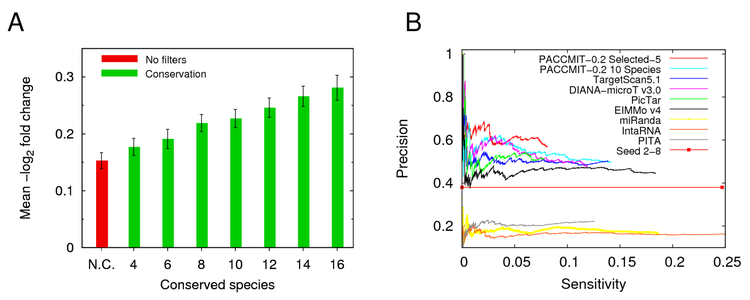


TP = true positive
FP = false positive
PR = precision = TP/(TP + FP)

$\mathbf{TP_{Ref}}$ = true positives expected from a reference method:

$$\mathrm{TP}_{\mathrm{Ref}} = \mathrm{PR}_{\mathrm{Ref}}(\mathrm{TP} + \mathrm{FP})$$

$\mathbf{PR_{Ref}}$ = precision of the reference method



(A) Targets predicted using a more strict conservation filter are on average more repressed ($p < 0.05$).

(B) Comparison with widely used methods shows that our method (using a conservation and an accessibility filter) is the most precise.

References

1. Murphy, E., Vanicek, J., Robins, H., Shenk, T., Levine, A. J. *Proc. Nat. Acad. Sci. USA*, **2008**, *105*, 5453.
2. Marin, R.M. and Vanicek, J. *Nucleic Acids Res.*, **2011**, *39*, 19.
3. Marin, R.M. and Vanicek, J. *in preparation*.
4. Robins, H. and Press, W. H. *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 15557.