Methods 91 (2015) 57–68

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Identification, validation, and characterization of noncanonical miRNAs

James M. Burke [1], Chad V. Kuny [1], Rodney P. Kincaid, Christopher S. Sullivan *

The University of Texas at Austin, Institute for Cellular and Molecular Biology, Center for Synthetic and Systems Biology, Center for Infectious Disease and Dept. Molecular Biosciences, 1 University Station A5000, Austin, TX 78712-0162, United States

ARTICLE INFO

Article history:
Received 15 May 2015
Received in revised form 17 July 2015
Accepted 21 July 2015
Available online 22 July 2015

Keywords:
miRNA
miRNA biogenesis
RNAP III
Drosha
Dicer
Virus
hsa-pri-miR-5683

ABSTRACT

Many eukaryotes and some viruses encode microRNAs (miRNAs), small RNAs that post-transcriptionally regulate gene expression. While most miRNAs are generated through the activity of RNA Polymerase II (RNAP II) and subsequent processing by Drosha and Dicer, some viral miRNAs utilize alternative pathways of biogenesis. Some members of the herpesvirus and retrovirus families can direct synthesis of miRNAs through RNAP III transcription rather than RNAP II and can utilize atypical enzymes to generate miRNAs. Though the advantages of alternative miRNA biogenesis remain unclear for herpesviruses, the retroviral miRNA biogenesis routes allow the RNAP II transcribed retroviral genome to escape Drosha cleavage while still expressing abundant, biologically-active miRNAs. These RNAP III-derived miRNAs have unique characteristics that allow for their identification and characterization. In this article, we describe procedures to predict, validate, and characterize RNAP III-transcribed miRNAs and other small RNAs, while providing resources that are also useful for canonical miRNAs.

1. Introduction

MicroRNAs (miRNAs) are small RNAs (~22 nt) that direct post-transcriptional repression of eukaryotic gene expression [1,2]. Most miRNAs are generated through an established biogenesis pathway. This pathway begins with RNA Polymerase II (RNAP II)-meditated transcription of longer primary miRNA (pri-miRNA) transcripts containing one or more stem–loop structures (~3 helices) in which miRNAs are embedded. The microprocessor, composed of the RNase III enzyme Drosha and the RNA-binding protein DGCR8, cleaves these stem–loop structures to release precursor miRNA hairpins (pre-miRNA) [3–7]. Pre-miRNAs are subsequently exported to the cytosol where they are processed by Dicer [8–10] to yield ~22 nt duplex RNAs [3]. Typically, one strand is then incorporated into the RNA-induced silencing complex (RISC) to direct mRNA transcript association and subsequent repression of gene expression [11–13].

In contrast to the canonical pathway of miRNA biogenesis, multiple viruses from the herpesvirus [14–18] and retrovirus [19–21] families are known to utilize alternative pathways to generate miRNAs. The herpesvirus saimiri (HVS) pre-miRNAs are generated by the integrator complex, which is better known for its role in small nuclear RNA processing [22]. The murid gammaherpesvirus 68 (MHV-68) pri-miRNAs are transcribed by RNAP III [16,23] and processed by tRNase Z rather than Drosha to generate pre-miRNAs [14]. The foamy viruses and bovine leukemia virus (BLV), both retroviruses, express pri-miRNAs via RNAP III [18,19,21]. While some retroviral pri-miRNAs bypass Drosha processing and are directly processed by Dicer, others are processed by Drosha only in the context of RNAP III transcripts [19–21]. Ultimately, this miRNA biogenesis strategy, whereby sub-genomic RNAP III-transcribed pri-miRNAs are expressed from proviral genomes, allows for the production of retroviral miRNAs without Drosha-mediated cleavage of the RNA genome intermediate, which could potentially decrease viral fitness. The methodology involved in the identification and validation of this alternative pathway of miRNA synthesis will be the primary focus of this article.

The genes and associated transcripts transcribed by RNAP III differ from their RNAP II counterparts in multiple aspects. Transcripts generated by RNAP III are generally much shorter (~100–300 nts) than RNAP II transcripts, and are initiated from one of the three known types of RNAP III promoters [24,25]. In Section 2.1 we describe a bioinformatic strategy to predict RNAP III transcribed pri-miRNAs through the identification of these promoter sequences as well as predicted secondary structure in the transcript.

Once a potential RNAP III-transcribed miRNA has been predicted using the algorithm, its expression and dependence on RNAP III for expression must be verified experimentally. In Section 2.2, we describe two assays to verify that predicted pri-miRNAs are synthesized by RNAP III rather than the canonical

* Corresponding author.
E-mail address: Chris_sullivan@mail.utexas.edu (C.S. Sullivan).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ymeth.2015.07.013
1046-2023/

RNAP II. The first assay utilizes α-amanitin under conditions specific for RNAP II inhibition [14]. Insensitivity to α-amanitin indicates that other RNA polymerases are responsible for the transcription of the given pri-miRNA. The second assay takes advantage of a 5′-monphosphate dependent exoribonuclease and a 5′ RNA polyphosphatase to determine if the transcript of interest has a triphosphate group at its 5′ end, which is a characteristic of RNAP III primary transcripts [24].

Once a small RNA (~22 nt) has been identified, whether transcribed by RNAP II or RNAP III, several criteria must be met to confirm that the small RNA is indeed a miRNA. In Section 2.3, we discuss the parameters and the established assays that are needed to classify a small RNA as a miRNA. These assays include those that determine whether a small RNA uses Drosha (Section 2.3 and Fig. 3) and/or Dicer (Fig. 3D) for maturation and demonstrate that the small RNA associates with RNAi machinery. The Drosha cleavage assay described in Section 3 utilizes overexpression of Drosha to enhance processing of structures embedded within RNAP II-transcribed reporter transcripts. This assay can be used to rapidly and quantitatively screen sequences for susceptibility to Drosha cleavage.

In sum, the methods described in this paper allow for the prediction, identification, and validation of RNAP III-transcribed miRNAs and other small RNAs. While these miRNAs were first identified as encoded in viral genomes, in principle, host equivalents can also be analyzed via these methods.

2. Prediction and validation of noncanonical RNAP III-transcribed viral miRNAs

2.1. Prediction algorithm

The ability to identify non-coding RNAs (ncRNAs) from genome sequence data is a challenging problem. However, some specialized algorithms have been developed to identify specific types of ncRNAs. Perhaps the most successful examples are algorithms to identify transfer RNA (tRNA) genes [24–28]. Eukaryotic tRNAs, 7SL RNA, Alu SINEs, adenovirus VA RNAs, and the murine gammaherpesvirus 68 microRNAs (miRNAs) all share a similar RNA Polymerase III (RNAP III) type 2 gene structure. Algorithms devised for tRNA prediction have taken advantage of both the intragenic promoter elements in the DNA sequence as well as small structural motifs in the predicted RNA secondary structure. The strategy that we developed for the prediction of novel miRNAs transcribed by RNAP III can be broken into two distinct tasks: identification of primary sequences with similarity to RNAP III control regions (promoter and terminator) and the identification of pre-miRNA like structures in the predicted RNA secondary structure (Fig. 1). The first step identifies the internal promoter sequences (Box A and Box B) and the terminator sequence. We allow for deviation from the canonical organization by allowing for the possibility of the terminator sequence to precede the B Box sequence as in the bovine leukemia virus (BLV) miRNAs [19]. This approach could also be extended to other promoter architectures including the RNAP III, type 1 promoter of the 5s ribosomal RNA or the RNAP III, type 3 promoter of U6 snRNA [22]. In our implementation of the algorithm we also optionally filter out coding sequences to reduce the total search space as none of the reported RNAP III transcribed microRNAs overlap with coding sequences. In the second step of the algorithm, the RNA secondary structure of the predicted candidate transcript is calculated using the RNAfold software [29]. The RNA secondary structure is then searched for a pre-microRNA like structure. We allow for shorter pre-microRNA stem–loops than is typically found in other microRNA prediction strategies as many of the currently identified viral RNAP III transcribed microRNAs are not processed by Drosha and appear to have shorter stem–loop structures. The data output that results is in a tab delimited text format that is viewable in spreadsheet programs.

2.1.1. Materials and reagents

(1) Computer with a Linux based operating system (Fedora 20 x86_64 or Ubuntu 14.04 LTS 64-bit).
(2) Python 2.7 (installed by default on above operating systems).

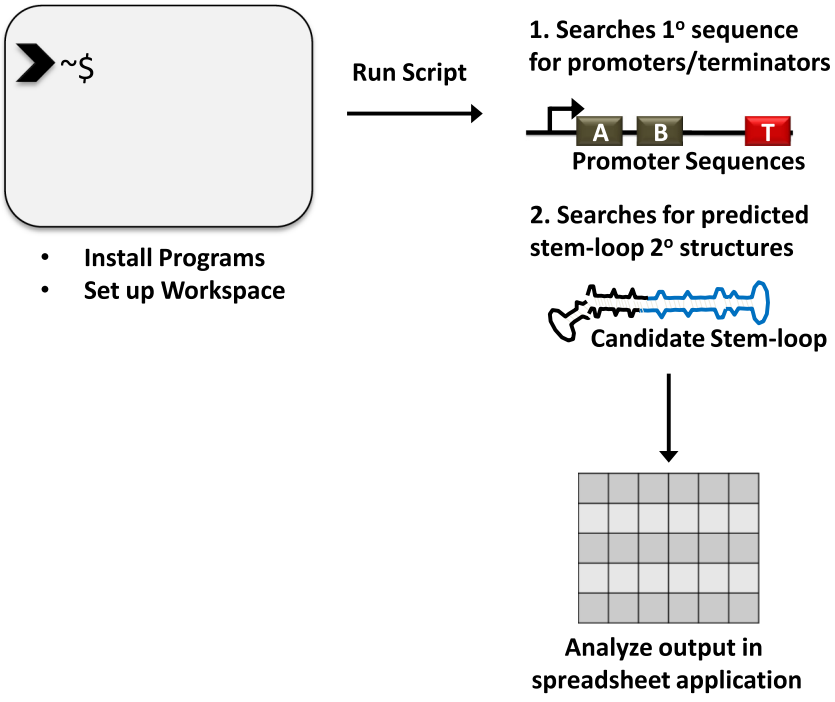

Fig. 1. Flowchart for the process of RNAP III transcribed small RNA bioinformatics prediction.