Gene expression

miRNAmeConverter: an R/bioconductor package for translating mature miRNA names to different miRBase versions

Stefan J. Haunsberger, Niamh M. C. Connolly and Jochen H. M. Prehn*

Department of Physiology & Medical Physics, Royal College of Surgeons in Ireland, Centre for Systems Medicine, Dublin 2, Ireland

*To whom correspondence should be addressed.
Associate editor: Cenk Sahinalp

Received on April 5, 2016; revised on August 19, 2016; editorial decision October 12, 2016; accepted on October 18, 2016

Abstract

Summary: The miRBase database is the central and official repository for miRNAs and the current release is miRBase version 21.0. Name changes in different miRBase releases cause inconsistencies in miRNA names from version to version. When working with only a small number of miRNAs the translation can be done manually. However, with large sets of miRNAs, the necessary correction of such inconsistencies becomes burdensome and error-prone. We developed miRNAmeConverter, available as a Bioconductor R package and web interface that addresses the challenges associated with mature miRNA name inconsistencies. The main algorithm implemented enables high-throughput automatic translation of species-independent mature miRNA names to user selected miRBase versions. The web interface enables users less familiar with R to translate miRNA names given in form of a list or embedded in text and download of the results.

Availability and Implementation: The miRNAmeConverter R package is open source under the Artistic-2.0 license. It is freely available from Bioconductor (http://bioconductor.org/packages/miRNAmeConverter). The web interface is based on R Shiny and can be accessed under the URL http://www.systemsmedicineireland.ie/tools/mirna-name-converter/. The database that miRNAmeConverter depends on is provided by the annotation package miRBaseVersions.db and can be downloaded from Bioconductor (http://bioconductor.org/packages/miRBaseVersions.db). Minimum R version 3.3.0 is required.

Contact: stefanhaunsberger@rcsi.ie

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

miRNAs are involved in post-transcriptional gene regulation (Bartel, 2004) and play key roles in a broad spectrum of biological processes such as differentiation (Naguibneva et al., 2006), proliferation and apoptosis (Cheng, 2005; Wang et al., 2015). Information about miRNAs is stored in the miRBase database and follows a miRNA naming convention (Ambros et al., 2003; Griffiths-Jones, 2004; Kozomara and Griffiths-Jones, 2014). The current miRBase release version 21 contains 28 645 stem loop sequences, resulting in 35 828 mature miRNAs from 223 organisms (Supplementary Figure S1). During its development, miRNAs have been added to or deleted from the database, while some miRNA names have been changed. In miRBase version 19, for example, around 4800 mature miRNAs were renamed (Supplementary Figure S2). As a result, many miRNA names across different miRBase versions are inconsistent. For example the mature miRNA hsa-miR-190a-5p was originally named hsa-miR-190 in miRBase release 2.0. This name was changed to hsa-miR-190a in release 18.0, and was again renamed to hsa-miR-190a-5p in version 19.0. These naming inconsistencies can cause various difficulties, such as when utilizing online miRNA-
associated tools such as miRTarBase (Hsu et al., 2014), which currently requires miRBase version 21.0 as input, or miRecords (Xiao et al., 2009), which requires version 17.0. To the best of our knowledge, miRConverter, as part of miRSystem (Lu et al., 2012), and miRBase Tracker (Vandesompele, 2014) are the only tools publicly available to convert miRNA names to different versions. miRBase Tracker supports single input values. However, neither of these tools can be used for high throughput translation in workflows.

We have developed miRNAMEConverter, an open source Bioconductor R package available in R and via web interface that translates mature miRNA names to any chosen miRBase version. In addition to the miRBase version specific miRNA name the user can also verify the sequence of the corresponding miRNA (available in Bioconductor v. 3.4). The package has already been utilised to translate more than 2000 mature miRNA names in the EpimiRBase database, a manually curated database listing miRNAs identified as dysregulated in epilepsy-related publications (Mooney et al., 2016). In addition to the translation function, our miRNAMEConverter package also includes functions for identifying/verifying valid miRNA names and detecting the most likely miRBase version of a given set of miRNAs.

2 Using the miRNAMEConverter package

2.1 Using the package in R

The package automatically processes sets of miRNA names in R. The documentation consists of manual pages for each function and a vignette containing use cases and executable code chunks.

A typical use case would be where a set of mature miRNA names needs to be translated to different miRBase versions. In the example below the set of mature miRNA names, mmu-miRNAs, is translated by the translateMiRNAName-function to the user-selected miRBase versions 8, 8.1, 10, 18 and 21.

```r
> library(miRNAMEConverter)
> miRNAs = c("mmu-miR-210", "mmu-miR-213", "mmu-miR-517")
> nc = MiRNAMEConverter()
> miRNAs.translated = translateMiRNAName(nc, miRNAs, versions = c(8, 8.1, 10, 18, 21))
> Following miRNA will be neglected (not listed in miRBase): mmu-miR-517
```

In this case, the algorithm detects that hsa-miR-517 is not a valid miRNA name. The variable miRNAs.translated (shown below) now contains the result data frame and two attached attributes, description and sequence. Description contains more detailed information about why hsa-miR-517 is not listed in the result whereas sequence contains miRNA sequences (e.g. use attr(miRNAs.translated, 'sequence') to access attribute sequence).

```r
> miRNAs.translated

mimat input v8.0 v8.1
MMAT0000658 mmu-miR-210 mmu-miR-210 mmu-miR-210
MMAT0000660 mmu-miR-213 mmu-miR-213 mmu-miR-181a*
v10.0 v18.0
MMAT0000670 mmu-miR-210 mmu-miR-210-3p mmu-miR-210-3p
mmu-miR-181a-1* mmu-miR-181a-1-3p mmu-miR-181a-1-3p

As shown, the name of miRNA mmu-miR-210 was changed to mmu-miR-210-3p in v18.0, whereas mmu-miR-213 was renamed multiple times to its most recent name in v21.0, mmu-miR-181a-1-3p.
```

2.2 The web interface

The layout of the miRNAMEConverter web interface is clean and straightforward (Supplementary Figure S3). Here, the user pastes the input values into the input text field, chooses the desired output miRBase versions and presses the ‘Translate’ button. The input values can consist of a list of mature miRNA names or a text paragraph, as the algorithm detects and filters text strings matching likely miRNA names. Following a successful translation, results are available for download in three different formats.

3 Conclusion

Here, we described the miRNAMEConverter Bioconductor R package and its corresponding web interface that we developed to address mature miRNA naming inconsistencies across different miRBase versions. The miRNAMEConverter is fast, high-throughput and user-friendly. The corresponding web interface is straightforward to use and is therefore accessible to biologists and other users less familiar with the programming language R.

Acknowledgements

Many thanks to Dr R. Conlon, Dr M. Salvucci, Dr N. Monsefi and A. Browne for motivation, beta testing and fruitful discussions.

Funding

This work was supported by the Science Foundation Ireland research grants 12/COEN/18—Joint Programming in Neurodegenerative Disorders, COEN-Phase II, ‘Neuro-miR’ and 13/IA/1881 to J.H.M.P.

Conflict of Interest: none declared.

References