Package: microRNA (via r-universe)

October 29, 2026

Version 1.71.0

Author R. Gentleman, S. Falcon

Title Data and functions for dealing with microRNAs

Description Different data resources for microRNAs and some functions for manipulating them.

License Artistic-2.0

Maintainer ``Michael Lawrence" <lawremi@gmail.com>

Imports Biostrings (>= 2.11.32)

Depends R (>= 2.10)

biocViews Infrastructure, GenomeAnnotation, SequenceMatching

Config/pak/sysreqs zlib1g-dev

Repository https://bioc.r-universe.dev

Date/Publication 2026-04-28 12:33:13 UTC

RemoteUrl https://github.com/bioc/microRNA

RemoteRef HEAD

RemoteSha 0fdb9ce9b9ed00b1dcc834aec2b973258b96c5fa

Contents

get_selfhyb_subseq 2
hsSeqs 3
hsTargets 4
matchSeeds 5
mmSeqs 6
mmTargets 6
RNA2DNA 7
s3utr 8
seedRegions 9

Index **10**

get_selfhyb_subseq *Get Self-Hybridizing Subsequences*

Description

This function finds the longest self-hybridizing subsequences present in RNA or DNA sequences.

Usage

```
get_selfhyb_subseq(seq, minlen, type = c("RNA", "DNA"))
show_selfhyb_counts(L)
show_selfhyb_lengths(L)
```

Arguments

seq	character vector of RNA or DNA sequences
minlen	an integer specifying the minimum length in bases of the self-hybridizing subsequences. Subsequences with length less than minlen will be ignored.
type	one of "RNA" or "DNA" depending on the type of sequences provided in seq. Note that you cannot mix RNA and DNA sequences.
L	The output of get_selfhyp_subseq.

Details

get_selfhyb_subseq finds the longest self-hybridizing subsequences of the specified minimum length.

These are defined to be the longest string that is found in both the input sequence, seq, and in its reverse complement.

Value

A list with an element for each sequence in seq. The list will be named using names(seq).

Each element is itself a list with an element for each longest self-hybridizing subsequence (there can be more than one). Each such element is yet another list with components:

starts	integer vector giving the character start positions for the self-hybridizing subsequence in the sequence.
rcstarts	integer vector giving the character start positions for the reverse complement of the self-hybridizing subsequence in the sequence.

Author(s)

Seth Falcon

Examples

```
seqs = c(a="UGAGGUAGUAGGUUGUAUAGUU", b="UGAGGUAGUAGGUUGUGUGGUU",
         c="UGAGGUAGUAGGUUGUAUGGUU")

ans = get_selfhyb_subseq(seqs, minlen=3, type="RNA")
length(ans)

ans[["a"]]

show_selfhyb_counts(ans)
show_selfhyb_lengths(ans)
```

---

hsSeqs *Human Mature microRNA Sequences*

Description

A set of human microRNA sequences.

Usage

```
data(hsSeqs)
```

Format

A character vector.

Details

Each sequence represents a different mature human microRNA.

Source

http://microrna.sanger.ac.uk/sequences/index.shtml

References

miRBase: microRNA sequences, targets and gene nomenclature. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. NAR, 2006, 34, Database Issue, D140-D144

The microRNA Registry. Griffiths-Jones S. NAR, 2004, 32, Database Issue, D109-D111

Examples

```
data(hsSeqs)
```

hsTargets *Human microRNAs and their target IDs*

Description

A set of human microRNA names and their corresponding known targets given as ensembl Transcript IDs.

Usage

```
data(hsTargets)
```

Format

A data frame of microRNAs and their target ensembl IDs as recovered from miRBase. Additional columns are also provided to give the Chromosome as well as the start and end position of the microRNA binding site, and the strand orientation (plus or minus).

Details

Each mapping represents a different human microRNA, paired with one viable target. Other information about where the microRNA binds is also included. Some microRNAs have multiple targets and so some microRNAs may be represented more than once.

Source

http://microrna.sanger.ac.uk/sequences/index.shtml

References

miRBase: microRNA sequences, targets and gene nomenclature. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. NAR, 2006, 34, Database Issue, D140-D144

The microRNA Registry. Griffiths-Jones S. NAR, 2004, 32, Database Issue, D109-D111

Examples

```
data(hsTargets)
```

matchSeeds *A function to match seed regions to sequences.*

Description

Given an input set of seed regions and a set of sequences all locations of the seed regions (exact matches) within the sequences are found.

Usage

```
matchSeeds(seeds, seqs)
```

Arguments

seeds	The seeds, or short sequences, to match.
seqs	The sequences to find matches in.

Details

We presume that the problem is an exact matching problem and that all sequences are in the correct orientation for that. If, for example, you start with seed regions from a microRNA (for seeds) and 3'UTR sequences (for seqs), then you would want to reverse complement one of the two sequences. And make sure all sequences are either DNA or RNA.

Names from either seeds or seqs are propogated, as much as is possible.

Value

A list containing one entry for each element of seeds that had at least one match in one entry of seqs. Each element of this list is a named vector containing the elements of seqs that the corresponding seed has an exact match in.

Author(s)

R. Gentleman

See Also

seedRegions

Examples

```
library(Biostrings)
data(hsSeqs)
data(s3utr)
hSeedReg = seedRegions(hsSeqs)
comphSeed = as.character(reverseComplement(RNAStringSet(hSeedReg)))
comph = RNA2DNA(comphSeed)
mx = matchSeeds(comph, s3utr)
```

mmSeqs *Mouse Mature microRNA Sequences*

Description

A set of mouse microRNA sequences.

Usage

```
data(mmSeqs)
```

Format

A character vector.

Details

Each sequence represents a different mature mouse microRNA.

Source

http://microrna.sanger.ac.uk/sequences/index.shtml

References

miRBase: microRNA sequences, targets and gene nomenclature. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. NAR, 2006, 34, Database Issue, D140-D144

The microRNA Registry. Griffiths-Jones S. NAR, 2004, 32, Database Issue, D109-D111

Examples

```
data(mmSeqs)
```

mmTargets *Mouse microRNAs and their target IDs*

Description

A set of mouse microRNA names and their corresponding known targets given as ensembl Transcript IDs.

Usage

```
data(mmTargets)
```

Format

A data frame of microRNAs and their target ensembl IDs as recovered from miRBase. Additional columns are also provided to give the Chromosome as well as the start and end position of the microRNA binding site, and the strand orientation (plus or minus).

Details

Each mapping represents a different mouse microRNA, paired with one viable target. Other information about where the microRNA binds is also included. Some microRNAs have multiple targets and so some microRNAs may be represented more than once.

Source

http://microrna.sanger.ac.uk/sequences/index.shtml

References

miRBase: microRNA sequences, targets and gene nomenclature. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. NAR, 2006, 34, Database Issue, D140-D144

The microRNA Registry. Griffiths-Jones S. NAR, 2004, 32, Database Issue, D109-D111

Examples

```
data(mmTargets)
```

RNA2DNA *A Function to translate RNA sequences into DNA sequences.*

Description

RNA and DNA differ in that RNA uses uracil (U) and DNA uses thiamine (T), this function translates an RNA sequence into a DNA sequence by translating the characters.

Usage

```
RNA2DNA(x)
```

Arguments

x	A valid RNA sequence.

Details

No checking for validity of sequence is made, and the input sequence is translated to upper case.

Value

A character vector, of the same length as x where all characters are in upper case, and any instance of U in x is replaced by a T.

Author(s)

R. Gentleman

See Also

chartr

Examples

```
input = c("AUCG", "uuac")
RNA2DNA(input)
```

---

s3utr *Test sequence data*

Description

A vector of 3' UTR sequence data, the names correspond to Entrez Gene IDs and the data were extracted using biomaRt.

Usage

```
data(s3utr)
```

Format

A character vector, the values are the 3' UTR for a set of genes, the names are Entrez Gene Identifiers.

Details

The data were downloaded using the getSequence function in the biomaRt package and duplicate strings removed. There remain some duplicated Entrez IDs but the reported 3' UTRs are different.

Examples

```
data(s3utr)
```

seedRegions *A function to retrieve the seed regions from microRNA sequences*

Description

The seed region of a microRNA consists of a set of nucleotides at the 5' end of the microRNA, typically bases 2 through 7, although some times 8 is used.

Usage

```
seedRegions(x, start = 2, stop = 7)
```

Arguments

x	A vector of microRNA sequences.
start	The start locations, can be a vector.
stop	The stop locations, can be a vector.

Details

We use substr to extract these sequences.

Value

A vector of the same length as x with the substrings.

Author(s)

R. Gentleman

See Also

substr

Examples

```
data(hsSeqs)
seedRegions(hsSeqs[1:5])
seedRegions(hsSeqs[1:3], start=c(2,1,2), stop=c(8,7,9))
```

Index

∗ **datasets**
hsSeqs, 3
hsTargets, 4
mmSeqs, 6
mmTargets, 6
s3utr, 8
∗ **manip**
get_selfhyb_subseq, 2
matchSeeds, 5
RNA2DNA, 7
seedRegions, 9

chartr, *8*

get_selfhyb_subseq, 2

hsSeqs, 3
hsTargets, 4

matchSeeds, 5
mmSeqs, 6
mmTargets, 6

RNA2DNA, 7

s3utr, 8
seedRegions, *5*, 9
show_selfhyb_counts
*(*get_selfhyb_subseq*)*, 2
show_selfhyb_lengths
*(*get_selfhyb_subseq*)*, 2
substr, *9*