qMAP

(quantitative mapping for RNA differential fragmentation)

Version 1.0.0

User Manual

By:

Rawal HC, *et al.*

October, 2025

Table of Contents

1. Introduction

qMAP (quantitative mapping for RNA differential fragmentation) is a an R package consisting of modules designed to identify differentially fragmented parental RNAs between groups through quantitative mapping of sncRNAs using a linear model and random permutations; and to identify the small non-coding RNAs (sncRNA) species contributing to the identified differential fragmentation of the parental RNA.

qMAP modules:

qMAP module, for the identification of differentially fragmented parental RNAs between groups, have two models or methods:

qMAP Model 1 identify differentially fragmented parental RNAs between groups by comparing the *t*-statistic values from the real sample grouping against absolute values of the *t*-statistic based on the permutated grouping by shuffling the samples between groups.

qMAP Model 2 identify differentially fragmented parental RNAs between groups by comparing the overall difference in normalized coverage between the two groups against the overall difference in normalized coverage based on the permutated grouping by shuffling the samples between groups.

qMAP_mh module identify the sncRNA species that are involved in the observed differential fragmentation in a parental RNA by constructing a 2×2 contingency matrix, and then combine all these contingency matrices using *Mantel-Haenszel* procedure with continuity correction. Finally calculating the *Mantel-Haenszel* statistic on a chi-square distribution.

2. Installation and usage

qMAP is an R package and runs within the R computing environment:

- R, and
- RStudio (https://rstudio.com).

2.1. Pre-requisites:

Rtools, R packages – seqinr, matrixStats, Biostrings, dplyr, stringdist.

2.2. Installation:

To install *qMAP* directly from GitHub, run the following command in R:

```
install.packages(“remotes”)
remotes::install_github("cozyrna/qMAP")
```

Alternatively, if you have downloaded the package source file (qMAP_*.tar.gz), install it by opening R or RStudio and running:

```
install.packages("~/qMAP_*.tar.gz", repos = NULL, type = "source")
```

Replace ~/qMAP_*.tar.gz with the actual path and filename of the downloaded package.

2.3. Running *qMAP* modules:

After installing the *qMAP* package, load it in R with:

```
library(qMAP)
```

To access the description and help documentation for the package and its functions, use:

```
?qMAP.single_family
```

```
?qMAP.MH.single_family
```

2.3.1. *qMAP*:

In *qMAP* module, the identification of differentially fragmented parental RNAs between groups is performed through quantitative mapping of sncRNAs. The analysis uses:

- A raw read-count matrix of sncRNA reads across samples
- A FASTA file containing the sequences of parental RNAs

For each queried parental RNA (sncrna_family), the output includes:

- mean_diff - mean difference in coverage between two groups
- p – The *P*-value indicating statistical significance
- species_num - number of contributing sncRNA species mapped to the parental RNA.

2.3.1.1. Inputs:

method: Specify either “1” or “2” to select which model to run for identifying differentially fragmented parental RNAs.

- Use “2” to perform analysis with Model 2 (method = 2).
- Use “1” (default) to run Model 1 (method = 1).

a: A user-provided read-count matrix where the first and second columns are labeled as "Sequence" and "Annotation", respectively. The remaining columns contain read counts from different samples under study. Typically, the first half of these columns correspond to the samples from Condition1 (e.g., control or healthy tissue), and the second half of the columns correspond to the samples from Condition2 (e.g., treated or infected tissue). Each sequence in the “Sequence” column must be unique.

cl: A sample classification vector allowing only two groups:

- 1 for controls
- 2 for cases
- -1 for the samples to be excluded from the analysis

For example, consider a count matrix with:

- 3 number of samples from Condition 1 (say, control)

- 4 number of samples from Condition 2 (say, case1),
- 5 number of samples from Condition 3 (say, case2)

Now, if you want to compare to compare Condition 1 against Condition 2, the sample classification vector (cl) would be:

cl = c(1, 1, 1, 2, 2, 2, 2, -1, -1, -1, -1, -1)

For comparing Condition 1 against Condition 3, use:

cl = c(1, 1, 1, -1, -1, -1, -1, 2, 2, 2, 2, 2).

If there are only 2 conditions of samples, such as Condition1 with 4 samples and Condition 2 with 5 samples, we will use:

cl = c(1,1,1,1, 2,2,2,2,2).

Multiple example files are available in the folder **"./..../R/.../qMAP/extdata/"** that can be references for the proper formatting and structure of the input matrix (**a**):

sample_matrix1.txt : This matrix file contains a total of 9 samples, comprising 5 samples (Sample_1, Sample_2, Sample_3, Sample_4, Sample_5) from Condition 1, and 4 samples (Sample_6, Sample_7, Sample_8, Sample_9) from Condition 2. [Note: the sample classification vector (cl) for this will be define as: cl = c(1, 1, 1, 1, 1, 2, 2, 2, 2)]

sample_matrix2.txt : a matrix file with total 10 samples comprising 3 samples (Healthy_1, Healthy_2, Healthy_3) from Condition 1, 3 samples from Condition 2 (Patient_1, Patient_2, Patient_3), and 3 samples (Infected_1, Infected_2, Infected_3) from Condition 3. [Note: For comparing Condition 1 against Condition 3, the sample classification vector (cl) for this will be define as: cl = c(1, 1, 1, -1, -1, -1, 2, 2, 2). For comparing Condition 2 and Condition 3: cl = c(-1, -1, -1, 1, 1, 1, 2, 2, 2)].

sample_matrix3.txt : a matrix file with total 16 samples comprising 4 samples from each of the 4 conditions. [Note: For comparing Condition 1 against Condition 2, use cl = c(1, 1, 1, 1, 2, 2, 2, 2, -1,-1,-1,-1, -1,-1,-1,-1), or one can simply use a sub-matrix after trimming out the last 8 columns and then for matrix with remaining 8 columns, use cl = c(1, 1, 1, 1, 2, 2, 2, 2).

parental_rna_filename: A FASTA file containing sequences of parental RNA families. [Note: parental RNA FASTA files for human and mouse are provided in the ./data folder. Available files include:

- hg19-mt_tRNAs_CCA.fa
- hg19-tRNAs_CCA.fa
- human_rRNA.fa
- mm10-mt_tRNAs_CCA.fa
- mm10-tRNAs_CCA.fa
- mouse_rRNA.fa

sncrna_family: This parameter specifies a single sncRNA family name to search for in the user-provided parental RNA FASTA file. If a match is found, differential fragmentation analysis will be performed for that family.
Note: use any single family name, such as :

- “mature-tRNA-Ala-AGC”
- “tRNA-Ala-AGC”
- “mature-mt_tRNA-His-GTG”
- “mt_tRNA-His-GTG”
- “5S-rRNA”
- “16S-rRNA”

max.mismatch: This parameter specifies the maximum number of allowed mismatches (default is 1) when mapping sequence reads onto the FASTA sequence of the parental RNA.

Note: The function searches for sequence reads from the “Sequence” column of the input read-count matrix, mapping them onto the FASTA sequence (from parental_rna_filename) of the specified parental RNA family (sncrna_family).

min_count: This parameter sets the minimum allowed mean read-count across samples for the sncRNA family to be considered in the analysis. The default value is 10, meaning only families with mean read-counts greater than or equal to 10 across samples will be included.
Note: For families with a large number of species (e.g., rRNA families) , higher values such as 100, 1000, etc., can be used to reduce run-time.]

shuffling_round: This parameter specifies the minimum number of sample shuffling rounds performed during the statistical model to identify differential fragmentation. The default value is 100 for Model 1 (method =1), and 1000 for Model 2 (method = 2)).

output: This parameter specifies the name of the output file, By default, if no name is provided, the output file will be saved in the working directory with a name like “qMAP_*_*_*_*_output.txt", where * represents various parameters used in the run.

2.3.1.2. Example runs:

Note: After installation, the sample matrix files used in the example runs can be found in the "../qMAP/extdata/" directory.

To run the examples exactly as shown in the help documentation, you must first set the working directory to the base folder of the installed qMAP package. For example: setwd("./..../R/.../qMAP/").

Alternatively, you can directly provide the full path to the appropriate example matrix files without changing the working directory.

There are two variants of *qMAP* function call for a single sncRNA family:

1. Function call with read-count matrix file path as input:

    qMAP.single_family.1(input_file, ...)

2. Function call with read-count matrix as a data frame as input:

    qMAP.single_family.2(input_dataframe, ...)

This second variant is useful when one has to call *qMAP* function for multiple sncRNA families.

Example 1:

qMAP.single_family.1("~/sample_matrix1.txt", cl=c(1,1,1,1,1,2,2,2,2), sncrna_family="5S-rRNA", parental_rna_filename="~/mouse_rRNAs.fa")

Run differential fragmentation analysis for 5S-rRNA family using corresponding FASTA sequence in the user provided parental RNA file (mouse_rRNAs.fa) by mapping sequence reads in the query read-count matrix file (sample_matrix1.txt) comprising read counts for 5 control (cl = 1) and 4 case (cl = 2) samples. It will run this for method 1 (default) with allowed maximum mismatch 1 (default) and with 100 shuffling rounds (default, for method 1), only if the mean total read-counts across the samples for the mapped reads (against query sncRNA family) > 10 (default).

Example 2:

qMAP.single_family.1("~/sample_matrix1.txt", cl=c(1,1,1,1,1,2,2,2,2), sncrna_family="5S-rRNA", parental_rna_filename="~/mouse_rRNAs.fa", method = "1", max.mismatch =1, shuffling_round = 100, min_count = 10)

Run differential fragmentation analysis in a similar manner as for example 1 above.

Example 3:

qMAP.single_family.1("~/sample_matrix2.txt", cl=c(1,1,1,-1,-1,-1,2,2,2), sncrna_family="16S-rRNA", parental_rna_filename="~/mouse_rRNAs.fa", output = “qmap_results.txt”)

Run differential fragmentation analysis for 16S-rRNA family using corresponding FASTA sequence in the user provided parental RNA file (mouse_rRNAs.fa) by mapping sequence reads in the query read-count matrix file (sample_matrix2.txt) comprising read counts for 3 control (first three sample columns in the matrix) (cl = 1) and 3 case (last three sample columns in the matrix) (cl = 2) samples. Here, 3 samples excluded (cl = -1). It will run this for method 1 (default) with allowed maximum mismatch 1 (default) and with 100 shuffling rounds (default, for method 1), only if the mean total read-counts across the samples for the mapped reads (against query sncRNA family) > 10 (default). The result output will be saved to the specified file, qMAP_results.txt.

Example 4:

qMAP.single_family.1("~/sample_matrix1.txt", cl=c(1,1,1,1,1,2,2,2,2), sncrna_family="tRNA-Gly-CCC", parental_rna_filename="~/mm10-tRNAs_CCA.fa", method = "2")

Run differential fragmentation analysis for tRNA-Gly-CCC family by applying method 2 using corresponding FASTA sequence in the user provided parental RNA file (mm10-tRNAs_CCA.fa) by mapping sequence reads in the query read-count matrix file (sample_matrix1.txt) comprising read counts for 5 control (cl = 1) and 4 case (cl = 2) samples. It will run this with allowed maximum mismatch 1 (default) and with 1000 shuffling rounds (default, for method 2), only if the mean total read-counts across the

samples for the mapped reads (against query sncRNA family) > 10 (default).

Example 5:

qMAP.single_family.1("~/sample_matrix1.txt", cl=c(1,1,1,1,1,2,2,2,2), sncrna_family="tRNA-Gly-CCC", parental_rna_filename="~/mm10-tRNAs_CCA.fa", min_count = 2, method = "2")

Run differential fragmentation analysis for tRNA-Gly-CCC family by applying method 2 using corresponding FASTA sequence in the user provided parental RNA file (mm10-tRNAs_CCA.fa) by mapping sequence reads in the query read-count matrix file (sample_matrix1.txt) comprising read counts for 5 control (cl = 1) and 4 case (cl = 2) samples. It will run this with allowed maximum mismatch 1 (default) and with 1000 shuffling rounds (default, for method 2), only if the mean total read-counts across the samples for the mapped reads (against query sncRNA family) > 2.

Example 6:

qMAP.single_family.1("~/sample_matrix3.txt", cl=c(1,1,1,1,2,2,2,2,-1,-1,-1,-1,-1,-1,-1,-1), sncrna_family="tRNA-Val-AAC", parental_rna_filename="~/mm10-tRNAs_CCA.fa", max.mismatch = 0, method = "2")

Run differential fragmentation analysis for tRNA-Val-AAC family by applying method 2 using corresponding FASTA sequence in the user provided parental rna file (mm10-tRNAs_CCA.fa) by mapping sequence reads (with mismatch zero) in the query read-count matrix file (sample_matrix3.txt) comprising read counts for 4 Contioin1 (cl = 1), 4 Condition2 (cl = 2) samples, and rest 8 for other conditions. It will run this with 1000 shuffling rounds (default, for method 2), only if the mean total read-counts across the samples for the mapped reads (against query sncRNA family) > 10 (default).

Example 7:

Running with query read-count matrix as a data frame as input

First read and save the read-count matrix file as a dataframe

```
source("qMAP.R")
a <- read.delim("sample_matrix1.txt", header = TRUE)
```

Then call *qMAP* function

```
qMAP.single_family.2(a, cl=c(1,1,1,1,1,2,2,2,2), sncrna_family="tRNA-His-GTG", parental_rna_filename="~/mm10-tRNAs_CCA.fa", max.mismatch = 1, method = "2")
```

Run differential fragmentation analysis for tRNA-His-GTG family by applying method 2 using corresponding FASTA sequence in the user provided parental RNA file (mm10-tRNAs_CCA.fa) by mapping sequence reads (with mismatch 1) in the query read-count matrix file (a, sample_matrix1.txt) comprising read counts for 5 control (cl = 1) and 4 case (cl = 2) samples. It will run this with 1000 shuffling rounds (default, for method 2), only if the mean total read-counts across the samples for the mapped reads (against query sncRNA family) > 10 (default).

Example 8:

##To run *qMAP* for multiple families, it is suggested to run with the second variant (qMAP.single_family.2) in a for loop as follows:

First read and save the read-count matrix file as a dataframe

```
source("qMAP.R")
a = read.delim("sample_matrix2.txt")
cl = c(1,1,1,2,2,2,-1,-1,-1)
result = c()
```

Read sncRNA families from list in the file

```
sncrna_family_list = read.table("mmu_gtsrna.txt")[,1]
```

Then for each family in the list, call *qMAP* function

```
for (i in 1:length(sncrna_family_list))
```

```
{
result = rbind(result, qMAP.single_family.2(a, cl=cl, sncrna_family=sncrna_family_list[i], parental_rna_filename ="~/mm10-tRNAs_CCA.fa", method=1, shuffling_round=100, output="tmp.txt"))
}
write.table(result, "qmap_multifamily_results.txt", quote=F, row.names=F, sep="\t")
```

Run differential fragmentation analysis for each of the RNA families (listed in a file; here it is mmu_gtsrna.txt) using corresponding FASTA sequences in the user provided parental RNA file (mm10-tRNAs_CCA.fa) by mapping sequence reads in the query read-count matrix file (sample_matrix2.txt) comprising read counts for 3 control (first three sample columns in the matrix) (cl = 1) and 3 case (next three sample columns in the matrix) (cl = 2) samples. Here, last 3 samples excluded (cl = -1). It will run this for method 1 with allowed maximum mismatch 1 (default) and with 100 shuffling rounds, only if the mean total read-counts across the samples for the mapped reads (against query sncRNA family) > 10 (default). The result output will be saved to the specified file, qmap_multifamily_results.txt.

2.3.1.3. *qMAP* output:

Sample output files are available in the “extdata” folder:

- qmap_results1.txt
- qmap_results2.txt
- qmap_multifamily_results1.txt
- qmap_multifamily_results2.txt

Each output file contains four columns:

1. parental_rna – name of the query sncRNA_family
2. mean_diff – mean difference in coverage between the two groups of samples

3. p – calculated *P* value for the observed difference

4. species_num - number of sncRNA species mapped to the parental RNA family

In this output:

- mean_diff reflects the magnitude of differential fragmentation for the specified sncRNA family (parental_rna) between the two sample groups.
- p (*P*-value) indicates the statistical significance of the observed difference. In general, a *P*-value < 0.05 is considered statistically significant.
- species_num represents the number of contributing sncRNA species mapped to the parental RNA, providing insight into the depth and robustness of the observed difference.

2.3.2. *qMAP_mh*:

The *qMAP_mh* module is designed to identify individual sncRNA species that contribute to the observed differential fragmentation of a given parental RNA across groups. It require two inputs:

- A raw read-count matrix of sncRNA reads across samples
- A FASTA file containing the sequences of parental RNAs

For each queried parental RNA (sncrna_family), the output includes:

- sequence - the sncRNA species sequence
- or - the odds ratio
- p - the *P*-value inferred from the *Mantel-Haenszel* statistic
- start - the start position on the parental RNA where the sncRNA species is mapped.

2.3.2.1. Inputs:

a: A user-provided read-count matrix where the first and second columns are labeled as "Sequence" and "Annotation", respectively. The remaining columns contain read counts from different samples under study. Typically, the first half of these columns correspond to the samples from Condition1 (e.g., control or healthy tissue), and the second half of the columns correspond to the samples from Condition2 (e.g., treated or infected tissue). Each sequence in the “Sequence” column must be unique.

cl: A sample classification vector allowing only two groups:

- 1 for controls
- 2 for cases
- -1 for the samples to be excluded from the analysis

For example, consider a count matrix with:

- 3 number of samples from Condition 1 (say, control)
- 4 number of samples from Condition 2 (say, case1),
- 5 number of samples from Condition 3 (say, case2)

Now, if you want to compare to compare Condition 1 against Condition 2, the sample classification vector (cl) would be:

cl = c(1, 1, 1, 2, 2, 2, 2, -1, -1, -1, -1, -1)

For comparing Condition 1 against Condition 3, use:

cl = c(1, 1, 1, -1, -1, -1, -1, 2, 2, 2, 2, 2).

If there are only 2 conditions of samples, such as Condition1 with 4 samples and Condition 2 with 5 samples, we will use:

cl = c(1,1,1,1, 2,2,2,2,2).

Multiple example files are available in the folder **"./..../R/.../qMAP/extdata/"** that can be references for the proper formatting and structure of the input matrix (**a**):

sample_matrix1.txt : This matrix file contains a total of 9 samples, comprising 5 samples (Sample_1, Sample_2, Sample_3, Sample_4, Sample_5) from Condition1, and 4 samples (Sample_6, Sample_7, Sample_8, Sample_9) from Condition2. [Note: sample classification vector (cl) for this will be define as cl = c(1, 1, 1, 1, 1, 2, 2, 2, 2)]

sample_matrix2.txt : a matrix file with total 10 samples comprising 3 samples (Healthy_1, Healthy_2, Healthy_3) from Condition1, 3 samples from Condition2 (Patient_1, Patient_2, Patient_3), and 3 samples (Infected_1, Infected_2, Infected_3) from Condition3. [Note: For comparing Condition1 against Condition3, sample classification vector (cl) for this will be define as cl = c(1, 1, 1, -1, -1, -1, 2, 2, 2). For comparing Condition2 and Condition3, cl = c(-1, -1, -1, 1, 1, 1, 2, 2, 2)].

sample_matrix3.txt : a matrix file with total 16 samples comprising 4 samples from each of the 4 conditions. [Note: For comparing Condition1 against Condition2, use cl = c(1, 1, 1, 1, 2, 2, 2, 2, -1,-1,-1,-1, -1,-1,-1,-1), or one can simply use a sub-matrix after trimming out the last 8 columns and then for matrix with remaining 8 columns, use cl = c(1, 1, 1, 1, 2, 2, 2, 2).

parental_rna_filename: A FASTA file containing sequences of parental RNA families. [Note: parental RNA FASTA files for human and mouse are provided in the ./data folder. Available files include:

- hg19-mt_tRNAs_CCA.fa
- hg19-tRNAs_CCA.fa

- human_rRNA.fa
- mm10-mt_tRNAs_CCA.fa
- mm10-tRNAs_CCA.fa
- mouse_rRNA.fa

sncrna_family: This parameter specifies a single sncRNA family name to search for in the user-provided parental RNA FASTA file. If a match is found, differential fragmentation analysis will be performed for that family.
Note: use any single family name, such as :

- “mature-tRNA-Ala-AGC”
- “tRNA-Ala-AGC”
- “mature-mt_tRNA-His-GTG”
- “mt_tRNA-His-GTG”
- “5S-rRNA”
- “16S-rRNA”

max.mismatch: This parameter specifies the maximum number of allowed mismatches (default is 1) when mapping sequence reads onto the FASTA sequence of the parental RNA.
Note: The function searches for sequence reads from the “Sequence” column of the input read-count matrix, mapping them onto the FASTA sequence (from parental_rna_filename) of the specified parental RNA family (sncrna_family).

minimum_mean_count: This parameter sets the minimum allowed mean read count (default: 5) across samples for the individual sncRNA species. Only sncRNA species (i.e., rows in the read-count matrix) with mean read counts greater than or equal to minimum_mean_count will be included in the analysis.

output: This parameter specifies the name of the output file, By default, if no name is provided, the output file will be saved in the working directory with a name like “qMAP_MH_output_X.txt", where X represents query ‘sncrna_family’ parameter used in the run.

2.3.2.2. Example runs:

Note: After installation, the sample matrix files used in the example runs can be found in the "../qMAP/extdata/" directory.
To run the examples exactly as shown in the help documentation, you must first set the working directory to the base folder of the installed *qMAP* package. For example: setwd("./..../R/.../qMAP/").

Alternatively, you can directly provide the full path to the appropriate example matrix files without changing the working directory.

There are two variants of *qMAP_mh* function call for a single sncRNA family:

1. Function call with read-count matrix file path as input:
   qMAP.MH.single_family.1(input_file, ...)

2. Function call with read-count matrix as a data frame as input:
   qMAP.MH.single_family.2(input_dataframe, ...)

This second variant is useful when one has to call *qMAP_mh* function for multiple sncRNA families.

Example 1:

```
qMAP.MH.single_family.1("~/sample_matrix1.txt", cl=c(1,1,1,1,1,2,2,2,2),
sncrna_family="5S-rRNA", parental_rna_filename="~/mouse_rRNAs.fa")
```

Run the identification of sncRNA species contributing to the differential fragmentation of 5S-rRNA family using corresponding FASTA sequence in

the user provided parental RNA file (mouse_rRNAs.fa) by mapping sequence reads in the query read-count matrix file (sample_matrix1.txt) comprising read counts for 5 control (cl = 1) and 4 case (cl = 2) samples, with allowed maximum mismatch 1 (default) and considering only the reads with the row mean read-counts across the samples > 5 (default).

Example 2:

qMAP.MH.single_family.1("~/sample_matrix2.txt", cl=c(1,1,1,-1,-1,-1,2,2,2), sncrna_family="tRNA-Gly-CCC", minimum_mean_count = 10, parental_rna_filename="~/mm10-tRNAs_CCA.fa")

Run the identification of sncRNA species contributing to the differential fragmentation of tRNA-Gly-CCC family using corresponding FASTA sequence in the user provided parental RNA file (mm10-tRNAs_CCA.fa) by mapping sequencing reads in the query read-count matrix file (sample_matrix1.txt) comprising read counts for 3 control (first three sample columns in the matrix) (cl = 1) and 3 case (last three sample columns in the matrix) (cl = 2) samples, with allowed maximum mismatch 1 (default) and considering only the reads with the row mean read-counts across the samples > 10. Here, 3 samples are excluded (cl = -1).

Example 3:

qMAP.MH.single_family.1("~/sample_matrix3.txt", cl=c(1,1,1,1,2,2,2,2,-1,-1,-1,-1,-1,-1,-1,-1), sncrna_family="tRNA-Val-AAC", minimum_mean_count = 10, max.mismatch = 0, parental_rna_filename="~/mm10-tRNAs_CCA.fa", output = “qmap.mh_results.txt”)

Run the identification of sncRNA species contributing to the differential fragmentation of tRNA-Val-AAC family using corresponding FASTA sequence in the user provided parental RNA file (mm10-tRNAs_CCA.fa) by mapping sequencing reads in the query read-count matrix file (sample_matrix3.txt) comprising read counts for 4 Contioin1 (cl = 1), 4 Condition2 (cl = 2) samples, and rest 8 for other conditions, with allowed

maximum mismatch 0 and considering only the reads with the row mean read-counts across the samples > 10. The result output will be saved to the specified file, qmap.mh_results.txt.

Example 4:

Running with query read-count matrix as a data frame as input

First read and save the read-count matrix file as a dataframe

```
a <- read.delim("sample_matrix1.txt", header = TRUE)
```

Then call *qMAP_mh* function

```
qMAP.MH.single_family.2(a, cl=c(1,1,1,1,1,2,2,2,2), sncrna_family="tRNA-His-GTG", parental_rna_filename="~/mm10-tRNAs_CCA.fa")
```

Run the identification of sncRNA species contributing to the differential fragmentation of tRNA-His-GTG family using corresponding FASTA sequence in the user provided parental RNA file (mm10-tRNAs_CCA.fa) by mapping sequencing reads in the query read-count matrix file (sample_matrix1.txt) comprising read counts for 5 control (cl = 1) and 4 case (cl = 2) samples, with allowed maximum mismatch 1 (default) and considering only the reads with the row mean read-counts across the samples > 5 (default).

Example 5:

##To run *qMAP_mh* for multiple families, it is suggested to run with the second variant (qMAP.MH.single_family.2) in a for loop as follows:

First read and save the read-count matrix file as a dataframe

```
source("qMAP.R")
a = read.delim("sample_matrix2.txt")
cl = c(1,1,1,2,2,2,-1,-1,-1)
result = c()
```

Read sncRNA families from list in the file

```
sncrna_family_list = read.table("mmu_gtsrna.txt")[,1]
```

```
## Then for each family in the list, call qMAP_mh function
for (i in 1:length(sncrna_family_list))
{
result = rbind(result, qMAP.MH.single_family.2(a, cl=cl, sncrna_family=sncrna_family_list[i], parental_rna_filename ="~/mm10-tRNAs_CCA.fa", output="tmp.txt"))
}
write.table(result, "qmap.mh_multifamily_results.txt", quote=F, row.names=F, sep="\t")
```

Run the identification of sncRNA species contributing to the differential fragmentation of each of the RNA families (listed in a file; here it is mmu_gtsrna.txt) using corresponding FASTA sequences in the user provided parental RNA file (mm10-tRNAs_CCA.fa) by mapping sequencing reads in the query read-count matrix file (sample_matrix2.txt) comprising read counts for 3 control (first three sample columns in the matrix) (cl = 1) and 3 case (next three sample columns in the matrix) (cl = 2) samples. Here, last 3 samples excluded (cl = -1). It will run with allowed maximum mismatch 1 (default) and considering only the reads with the row mean read-counts across the samples > 5 (default). The result output will be saved to the specified file, qmap.mh_multifamily_results.txt.

2.3.2.3. *qMAP_mh* output

Sample output files are available in the “extdata” folder:

- qmap.mh_results1.txt
- qmap.mh_results2.txt
- qmap.mh_multifamily_results1.txt
- qmap.mh_multifamily_results2.txt

Each output file contains four columns:

1. parental_rna – name of the query parental RNA (sncRNA_family)

2. sequence – sequence of the sncRNA species mapped over parental RNA family with allowed mismatch

3. or – the odds ratio

4. p – *P* value inferred from the *Mantel-Haenszel* statistic

5. start – start position on the parental RNA sequence where the sncRNA species sequence is mapped

In this output:

- The odds ratio (or) > 1 suggests an over-representation of the sncRNA in the cases, while or < 1 suggests an over-representation of the sncRNA in the controls.
- *P*-value indicates the statistical significance of this over or under representation. In general, a *P*-value < 0.05 is considered statistically significant.

3. Troubleshooting:

There is a possibility that users may encounter errors if the input matrix does not adhere to the specified format. Additionally, errors may arise even when there are no sncRNA species mapped to the mentioned sncrna_family with the specified threshold (mismatch, mean count, etc.). Below, we discuss some common errors or issues users may face, along with their potential reasons and solutions for correction.

(i) Error message: Error in file(file, "rt") : cannot open the connection
In addition: Warning message:
In file(file, "rt") :
cannot open file '...': No such file or directory
Called from: file(file, "rt")

Solution: Ensure that the correct file path is provided for the input matrix file. Check that the file exists at the specified location and that you have provided the correct file name.

(ii) Error message: Error: Number of samples (x) in the input matrix is not equal to the mentioned samples (y) in the cl option."

Solution: This error may occur when the sample details (y) in cl option does not match with number of samples (x) in the input read-count matrix. Ensure that the number of elements in cl option is equal to the number of columns having read counts of the samples in the read-counts matrix file.

(iii) Error message: Error in `.rowNamesDF<-`(x, value = value) :
duplicate 'row.names' are not allowed
In addition: Warning message:
non-unique value when setting 'row.names': ‘ROW_NAME’
Called from: `.rowNamesDF<-`(x, value = value)

Solution: This error occurs when the first column (Sequence) in the input read-count matrix file contains duplicate values. The Sequence must be unique for each entry. Check the first column of your input matrix and remove any duplicate values to resolve this issue.

(iv) Error message: Error in rowMeans(e) : 'x' must be numeric
Called from: rowMeans(e)

Solution: This error occurs when the input read-count matrix contains non-numeric values in the columns meant to have read counts for the samples. To resolve this, you can either:

- Remove the rows that contain non-numeric values, or
- Replace the non-numeric values with suitable numbers, such as 0, depending on the context of your analysis.

(v) Warning: There are some non-integer values in the read-count matrix (e.g., 0.5, 1.01). Only whole numbers should be used.

Solution: This warning message appears when the input read-count matrix contains non-integer values (such as 0.5, 1.5, etc.) in the columns meant to have read counts for the samples. In general, it is recommended to use only whole numbers as read counts in the input matrix.

(vi) Error message: Error in check_whole_numbers(e) :
The read-count matrix contains non-finite values (NA, Inf, or NaN).
Called from: check_whole_numbers(e)

Solution: This error message appears when the input read-count matrix contains non-finite values (NA, Inf, or NaN) in the columns meant to have read counts for the samples. To resolve this issue, you can either:

- Remove the rows containing these non-finite values, or
- Replace these non-finite values with suitable numbers, such as 0, depending on your analysis needs.

(vii) Error message: Error in check_whole_numbers(e) :
All columns in the read-count matrix must be numeric. There are alphabets/characters in the matrix.
Called from: check_whole_numbers(e)

Solution: This error message appears when the input read-count matrix contains non-numeric values (characters or alphabets) in the columns meant to have read counts for the samples. To resolve this issue, you can either:

- Remove the rows containing these non-numeric values, or
- Replace these non-non-numeric values with suitable numbers, such as 0, depending on your analysis needs.

(viii) Error message: Error in if (mean(colSums(e_tmp)) < min_count) { :
missing value where TRUE/FALSE needed

Called from: qMAP.single_family(function()
read.delim(input_file), ...)

Solution: This error occurs, while running *qMAP* module (qMAP.single_family), when there are "NA" values in the columns containing raw read-counts for the samples in the input read-count matrix file. To resolve this issue, you can either:

- Remove the rows containing "NA" values, or
- Replace the "NA" values with suitable numbers, such as 0, depending on your analysis needs.

(ix) Error message: Error in mantelhaen.test(mh_data) : NAs are not allowed
Called from: mantelhaen.test(mh_data)

Solution: This error occurs, while running *qMAP_mh* module (qMAP.MH.single_family), when there are "NA" values in the columns containing raw read-counts for the samples in the input read-count matrix file. To resolve this issue, you can either:

- Remove the rows containing "NA" values, or
- Replace the "NA" values with suitable numbers, such as 0, depending on your analysis needs.

(x) Message: "Mentioned sncRNA family – XYZ, does not exist in the provided parental RNA file."

Comment: The XYZ sncRNA family name you are looking for differential fragmentation analysis, must be present in the provided parental RNA file.
Check the provided FASTA files in the ./data folder. For example, if using human_rRNA.fa as the parental RNA file, then you can provide the query sncRNA family name as anyone from the following:
"4.5S-rRNA", "5S-rRNA", "5.8S-rRNA", "12S-rRNA", "16S-rRNA", "18S-rRNA", "28S-rRNA", "45S-rRNA".

While using genomic tRNAs FASTA file (eg. hg19-tRNAs_CCA.fa), one should mention the sncRNA family search name as:
“mature-tRNA-Ala-AGC”, “mature-tRNA-His-GTG”, etc.

While using mitochondrial tRNAs FASTA file (eg. hg19-mt_tRNAs_CCA.fa), one should mention the sncRNA family search name as:
“mature-mt_tRNA-Ala-TGC”, “mature-mt_tRNA-His-GTG”, etc.

5. Generate RNA coverage along the length

The R script rna_coverage.R can be used to generate sncRNA species coverage along the length of the parental RNA sequence.

To run the script, you will need:

- matrix_file: An read count matrix file containing sRNA sequences, annotation, and raw count values in different samples (e.g., sample_matrix1.txt, sample_matrix2.txt, sample_matrix3.txt)
- sncrna_family: a single sncRNA family name to search for in the annotation column of query read-count matrix (e.g., 5S-rRNA, tRNA-His-GTG)
- parental_rna: DNA sequence of the parental RNA (e.g., “GTCTACGGCCATACCACCCTGAACGCGCCCGATCTCGTCTGATCTCGGAAGCTAAGCAGGGTCGGGCCTGGTTAGTACTTGGATGGGAGACCGCCTGGGAATACCGGGTGCTGTAGGCTTT”)

How to run:

Access the rna_coverage.R script from the installed package folder (scripts/), open it in your R environment, edit the input options (matrix_file, sncrna_family, parental_rna) as needed, and run the code.