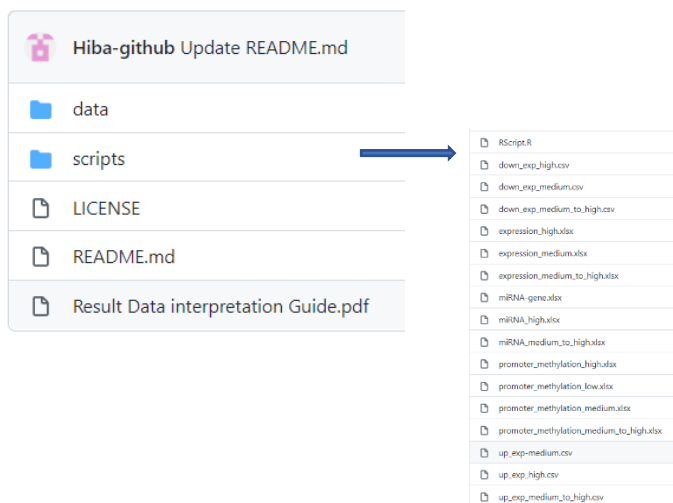


# User guide

1. Download data and setting up the R script
2. Analyze the data: Run R script
3. Visualize and export the result
4. Interpretate result

1. Download and Analyze data:



## Step1 :

Create a folder in your desktop and name it "Script".

## Step2 :

access the "Script" directory in GitHub page and download all the files and the R script.

## Step3 :

Open the R script with R studio software.

## Step4:

In line 3 , replace the pathway to your own.

## Step5 :

Run lines 3 to 7

## Step7 :

Replace the gene symbol "STAT3" to your own target gene symbol (in all the Script)

## Step 8 :

Select and run the target prevalence sections from the name of the section to the line containing the function "write.table"

```
1 ##### Run line 3:7
2 # set directory
3 setwd("C:/Users/LENOVO/Desktop/Script")
4 #cal libraries
5 library(survival)
6 library(tidyverse)
7 library(readxl)

10 #####
11 # Gene expression in High prevalence countries -----
12 expression_high <- read_excel("expression_high.xlsx")
13
14 e<- expression_high %>%
15   filter(gene== "STAT3")
16 e
17 write.table(e,"expression_profile_high.txt")
18
```

```
10 #####
11 # Gene expression in High prevalence countries -----
12 expression_high <- read_excel("expression_high.xlsx")
47 #####
48 # Gene promoter methylation in high prevalence countries -----
49
19 #####
20 # Is the Gene Differentially expressed in high prevalence countries ? -
21
```

```

238 # The miRNA regulating the gene -----
239 miRNA_gene <- read_excel("miRNA-gene.xlsx")
240
241 miRNA <- miRNA_gene %>%
242   filter(gene == "STAT3")
243 miRNA
244 #export result
245 write.table( miRNA , "miRNA_gene.txt" , row.names=F , sep = ",")
246
247
248
251 # Is the miRNAs expressed in MS in high prevalence countries ?
252
253
254 miRNA_high <- read_excel("miRNA_high.xlsx")
255
256
257 miRNA_ms1 <- miRNA_high %>%
258   filter(ID == "hsa-miR-130a-3p")
259 miRNA_ms1
260 #export result
261 write.table( miRNA_ms1 , "miRNA_expressed.txt" , row.names=F , sep = ",")
262
263
264

```

Step 9 :

Run this section

Step10 :

Run the miRNA section  
corresponding to your target  
prevalence.

**Note :** only high and medium to high prevalence miRNA data are available .

Maybe a future version will contain more data .

## 2. Result example

### • Gene expression

```

> e
# A tibble: 1 x 3
  Gene.symbol P.Value logFC
  <chr>      <chr>    <chr>
1 STAT3      0.510874 -1.39e-02

```

### • DGE

```

> exp_high
[1] "The gene is not differentially expressed"
> |

```

### • Methylation :

	gene	cpg	P.Value	B	logFC	UCSC_CpG_Islands_Name	Relation_to
	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
1	STAT3	cg06378498	7.23e-01	-4.83962	-2.42e-02	chr17:40539837-40540775	Island
2	STAT3	cg20716209	1.29e-01	-4.44458	-4.18e-01	chr17:40539837-40540775	S_Shore
3	STAT3	cg04517036	1.26e-01	-4.40942	-2.44e-01	chr17:40539837-40540775	Island
4	STAT3	cg07158701	5.60e-01	-4.63081	1.70e-01	chr17:40539837-40540775	Island
5	STAT3	cg09476841	7.54e-02	-4.28616	-2.18e-01	chr17:40539837-40540775	S_Shore
6	STAT3	cg13655938	9.90e-01	-4.73448	2.42e-03	chr17:40539837-40540775	S_Shore
7	STAT3	cg19557623	2.53e-01	-4.56735	-1.78e-01	chr17:40539837-40540775	Island
8	STAT3	cg09804439	1.05e-01	-4.47415	-2.58e-01	chr17:40539837-40540775	Island
9	STAT3	cg17411949	3.46e-01	-4.63537	-1.34e-01	chr17:40539837-40540775	Island
10	STAT3	cg01337508	2.18e-01	-4.53393	-1.86e-01	chr17:40539837-40540775	Island
11	STAT3	cg15704988	6.64e-02	-4.30271	-1.93e-01	chr17:40539837-40540775	Island

Step 9 :

Run this section

Step10 :

Run the miRNA section  
corresponding to your target  
prevalence.

- miRNA (healthy condition)

```
> miRNA
# A tibble: 14 x 2
  gene      'miRNA family'
  <chr>    <chr>
1 STAT3 hsa-miR-130a-3p/hsa-miR-130b-3p/hsa-miR-301a-3p/hsa-miR-301b-3p/hsa-miR-3666/hsa-miR-4295/hsa-miR-454-3p
2 STAT3 hsa-miR-106a-5p/hsa-miR-106b-5p/hsa-miR-17-5p/hsa-miR-20a-5p/hsa-miR-20b-5p/hsa-miR-519d-3p/hsa-miR-526b-3p/hsa-miR-93-5p
3 STAT3 hsa-miR-124-3p.1
4 STAT3 hsa-miR-181a-5p/hsa-miR-181b-5p/hsa-miR-181c-5p/hsa-miR-181d-5p/hsa-miR-4262
5 STAT3 hsa-let-7a-5p/hsa-let-7b-5p/hsa-let-7c-5p/hsa-let-7d-5p/hsa-let-7e-5p/hsa-let-7f-5p/hsa-let-7g-5p/hsa-let-7i-5p/hsa-miR-4458/hsa-miR-4500/hsa-miR-4501
6 STAT3 hsa-miR-124-3p.2/hsa-miR-506-3p
7 STAT3 hsa-miR-29a-3p/hsa-miR-29b-3p/hsa-miR-29c-3p
8 STAT3 hsa-miR-21-5p/hsa-miR-590-5p
9 STAT3 hsa-miR-410-3p
10 STAT3 hsa-miR-125a-5p/hsa-miR-125b-5p/hsa-miR-4319
11 STAT3 hsa-miR-6835-3p
12 STAT3 hsa-miR-371a-5p
13 STAT3 hsa-miR-296-5p
14 STAT3 hsa-miR-665
```

\* miRNA expression in MS

If the miRNA is expressed:

```
> miRNA_ms1
# A tibble: 1 x 3
  ID          P.Value logFC
  <chr>      <chr>    <chr>
1 hsa-miR-665 4.51e-01 -0.082918
```

If the miRNA is NOT expressed:

```
> miRNA_ms1
# A tibble: 0 x 3
```

Export data:

Results will be automatically exported in the working directory.

### 3. Definitions:

- **P value** : is the probability for the experimental outcome as observed or more extreme, if there is no difference in expression between the experimental conditions.  
A small P-value indicates evidence of differential expression, either overexpression or underexpression.
- **logFC** : represents the variation of mRNA abundance across different biological conditions. A positive value mean that the overexpressed and a negative value means that's the gene is underexpressed
- **B value** : the estimate of methylation level using the ratio of intensities between methylated and unmethylated alleles.  
 $\beta$  are between 0 and 1 with 0 being unmethylated and 1 fully methylated
- **UCSC CpG Islands Name**: Chromosomal coordinates of the CpG Island from UCSC.
- **Relation to UCSC CpG Island**: The location of the CpG relative to the CpG island.
- **DMR**: Differentially methylated regions  
If p value <0.05 the cpG is considered differentially methylated
  - CDMR = Cancer-specific Differentially Methylated Region.
  - RDMR = Reprogramming-specific Differentially Methylated Region.
  - NA = No data