

遇到下载的包有问题或者报错怎么办

关于SummarizedExperiment这样一个类对象的了解

## 遇到下载的包有问题或者报错怎么办

比如我想安装一个包：airway 用install.packages()报错了怎么办

### 1) 用BiocManager安装

```
1 if (!require("BiocManager", quietly = TRUE))
2   install.packages("BiocManager")
3 BiocManager::install("airway")
```

### 2) 直接下载压缩文件到R包的路径下

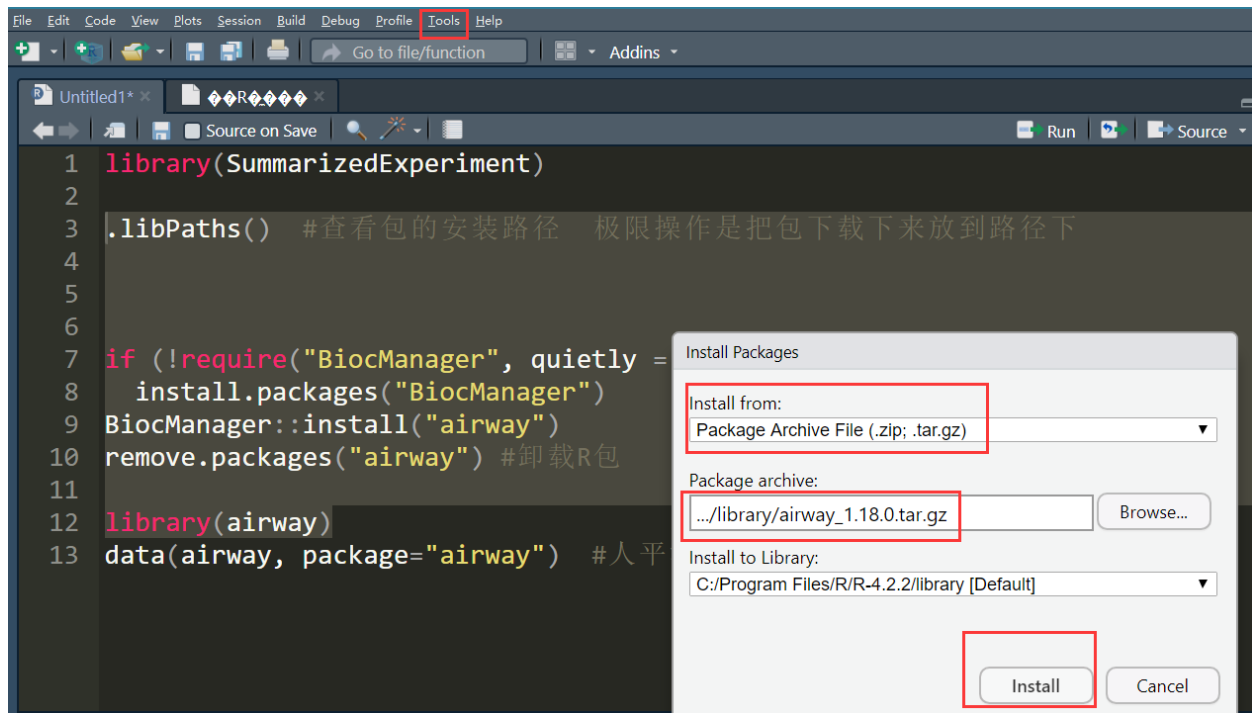
#### Package Archives

Follow [Installation](#) instructions to use this package in your R session.

Source Package	<a href="#">airway_1.18.0.tar.gz</a>
Windows Binary	
macOS Binary (x86_64)	
macOS Binary (arm64)	
Source Repository	git clone https://git.bioconductor.org/packages/airway
Source Repository (Developer Access)	git clone git@git.bioconductor.org:packages/airway
Package Short Url	<a href="https://bioconductor.org/packages/airway/">https://bioconductor.org/packages/airway/</a>
Package Downloads Report	<a href="#">Download Stats</a>

查看路径用.libPaths()

下一步点击tools的install packages

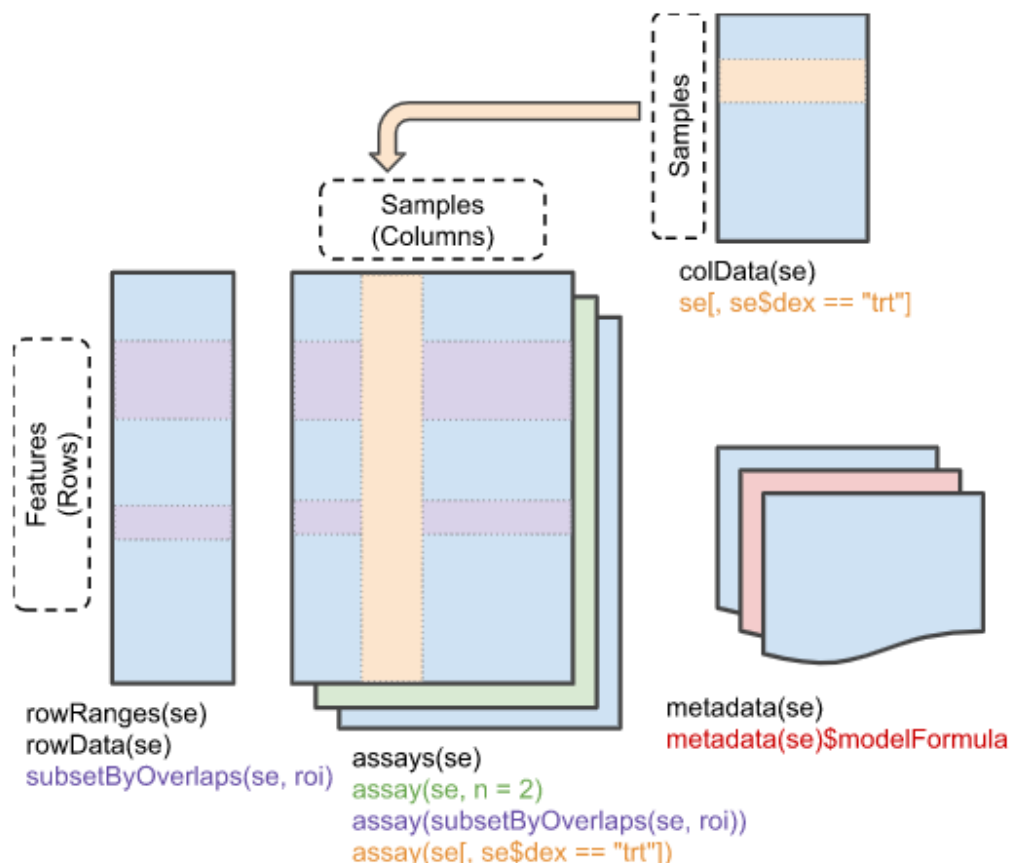


3) 下载压缩文件到R包的路径下直接解压就可以了

4) 用github下载

## 关于SummarizedExperiment这样一个类对象的了解

基因表达量的很多数据都存储在这样的类对象下，注意理解下面这个图，我一步一步运行你就能慢慢明白了



```
1 library(airway)
2 data(airway, package="airway") #人平滑肌细胞的RNA-seq测序的原始counts数据
```

Data

airway

Large

RangedSummarizedExperiment

我们看一下这个数据：64102行，8列 行是基因名列是样本名

```
> airway
class: RangedSummarizedExperiment
dim: 64102 8
metadata(1): ''
assays(1): counts
rownames(64102): ENSG000000000003 ENSG000000000005 ... LRG_98 LRG_99
rowData names(0):
colnames(8): SRR1039508 SRR1039509 ... SRR1039520 SRR1039521
colData names(9): SampleName cell ... Sample BioSample
```

如何获取count数据：用assays

```
> assays(airway)$counts
```

	SRR1039508	SRR1039509	SRR1039512	SRR1039513	SRR1039516
ENSG000000000003	679	448	873	408	1138
ENSG000000000005	0	0	0	0	0
ENSG000000000419	467	515	621	365	587
ENSG000000000457	260	211	263	164	245
ENSG000000000460	60	55	40	35	78
ENSG000000000938	0	0	2	0	1
ENSG000000000971	3251	3679	6177	4252	6721
ENSG00000001036	1433	1062	1733	881	1424
ENSG00000001084	519	380	595	493	820
ENSG00000001167	394	236	464	175	658
ENSG00000001460	172	168	264	118	241

看一下对于行的操作：用的rowRanges()这个命令：

包含了基因的起始位置和终止位置，可以计算基因长度

```
> rowRanges(airway)
GRangesList object of length 64102:
$ENSG000000000003
GRanges object with 17 ranges and 2 metadata columns:
      seqnames      ranges strand | exon_id exon_name
      <Rle>        <IRanges> <Rle> | <integer> <character>
[1]      X 99883667-99884983     - |    667145 ENSE00001459322
[2]      X 99885756-99885863     - |    667146 ENSE00000868868
[3]      X 99887482-99887565     - |    667147 ENSE00000401072
[4]      X 99887538-99887565     - |    667148 ENSE00001849132
[5]      X 99888402-99888536     - |    667149 ENSE00003554016
...      ...      ...      ... |    ...      ...
[13]     X 99890555-99890743     - |    667156 ENSE00003512331
[14]     X 99891188-99891686     - |    667158 ENSE00001886883
[15]     X 99891605-99891803     - |    667159 ENSE00001855382
[16]     X 99891790-99892101     - |    667160 ENSE00001863395
[17]     X 99894942-99894988     - |    667161 ENSE00001828996
```

看一下对于列的操作：用的colData()这个命令：实际上就是metadata

```
> colData(airway)
DataFrame with 8 rows and 9 columns
      SampleName      cell      dex      albut      Run      avgLength
      <factor> <factor> <factor> <factor> <factor> <integer>
SRR1039508 GSM1275862 N61311      untrt      untrt SRR1039508      126
SRR1039509 GSM1275863 N61311      trt        untrt SRR1039509      126
SRR1039512 GSM1275866 N052611     untrt      untrt SRR1039512      126
SRR1039513 GSM1275867 N052611     trt        untrt SRR1039513       87
SRR1039516 GSM1275870 N080611     untrt      untrt SRR1039516     120
SRR1039517 GSM1275871 N080611     trt        untrt SRR1039517     126
SRR1039520 GSM1275874 N061011     untrt      untrt SRR1039520     101
SRR1039521 GSM1275875 N061011     trt        untrt SRR1039521      98
```

可以根据设置metadata的筛选条件选取数据

```
> airway[, airway$dex == "trt"]
class: RangedSummarizedExperiment
dim: 64102 4
metadata(1): ''
assays(1): counts
rownames(64102): ENSG000000000003 ENSG000000000005 ... LRG_98 LRG_99
rowData names(0):
colnames(4): SRR1039509 SRR1039513 SRR1039517 SRR1039521
colData names(9): SampleName cell ... Sample BioSample
```

```
> airway_trt=airway[, airway$dex == "trt"]
> assays(airway_trt)$counts
```

	SRR1039509	SRR1039513	SRR1039517	SRR1039521
ENSG00000000003	448	408	1047	572
ENSG00000000005	0	0	0	0
ENSG00000000419	515	365	799	508
ENSG00000000457	211	164	331	229
ENSG00000000460	55	35	63	60
ENSG00000000938	0	0	0	0
ENSG00000000971	3679	4252	11027	7995

如果你想构造一个SummarizedExperiment

也就是三部分 一个matrix、一个colData、一个rowRanges

```
1 nrows <- 200
2 ncols <- 6
3 counts <- matrix(runif(nrows * ncols, 1, 1e4), nrows)
4 rowRanges <- GRanges(rep(c("chr1", "chr2"), c(50, 150)),
5                       IRanges(floor(runif(200, 1e5, 1e6)), width=100),
6                       strand=sample(c("+", "-"), 200, TRUE),
7                       feature_id=sprintf("ID%03d", 1:200))
8 colData <- DataFrame(Treatment=rep(c("ChIP", "Input"), 3),
9                      row.names=LETTERS[1:6])
10
11 SummarizedExperiment(assays=list(counts=counts),
12                      rowRanges=rowRanges, colData=colData)
```

```
class: RangedSummarizedExperiment
dim: 200 6
metadata(0):
assays(1): counts
rownames: NULL
rowData names(1): feature_id
colnames(6): A B ... E F
colData names(1): Treatment
```

其余的操作：可以直接取二维子集

```
1 airway[1:5, 1:3]
```

```
> airway[1:5, 1:3]
class: RangedSummarizedExperiment
dim: 5 3
metadata(1): ''
assays(1): counts
rownames(5): ENSG00000000003 ENSG00000000005 ENSG00000000419
            ENSG00000000457 ENSG00000000460
rowData names(0):
colnames(3): SRR1039508 SRR1039509 SRR1039512
colData names(9): SampleName cell ... Sample BioSample
```

```
1 counts <- matrix(1:15, 5, 3, dimnames=list(LETTERS[1:5],
      LETTERS[1:3]))#LETTERS
2 dates <- SummarizedExperiment(assays=list(counts=counts),
3                               rowData=DataFrame(month=month.name[1:5],
4                               day=1:5))
5 # Subset all January assays
6 dates[rowData(dates)$month == "January", ]
```

```
> dates[rowData(dates)$month == "January", ]
class: SummarizedExperiment
dim: 1 3
metadata(0):
assays(1): counts
rownames(1): A
rowData names(2): month day
colnames(3): A B C
colData names(0):
```