

The R package *trioClasses* for definition of the class *SNPTrioExperiment*, an extension of *SummarizedExperiment*, for use in trio based analyses of genetic data.

Samuel G. Younkin

March 12, 2013

1 Packages & Data

```
> library("trioClasses")
> data("sample")
> data("8q24-european-all.sm")
```

2 SummarizedExperiment

```
> se <- SummarizedExperiment(assays = SimpleList(geno = t(sm)),
  colData = col.DF, rowData = gr)
```

3 Pedigree

```
> ped <- PedClass(ped.DF)
```

4 SNPTrioExperiment

```
> (ste <- SNPTrioExperiment(se, pedigree = ped))
```

```
class: SNPTrioExperiment
dim: 8951 960
exptData(0):
assays(1): geno
rownames(8951): chr8:129296000 chr8:129296113 ...
  chr8:130354703 chr8:130354790
rowData metadata column names(0):
colnames(960): H_ME-DS10776_2-DS10776_2
  H_ME-DS10776_3-DS10776_3 ... H_ME-DS11313_3-DS11313_3
  H_ME-DS11313_1-DS11313_1
colData names(1): id
pedigree(4139): famid id fid mid sex dx
complete trios(320):
```

5 Methods

5.1 ScanTrio

```
> (ste.rare <- ste[!(MAF(ste) >= 0.01 | is.na(MAF(ste))))])

class: SNP Trio Experiment
dim: 6397 960
exptData(0):
assays(1): geno
rownames(6397): chr8:129296113 chr8:129296185 ...
               chr8:130354703 chr8:130354790
rowData metadata column names(0):
colnames(960): H_ME-DS10776_2-DS10776_2
               H_ME-DS10776_3-DS10776_3 ... H_ME-DS11313_3-DS11313_3
               H_ME-DS11313_1-DS11313_1
colData names(1): id
pedigree(4139): famid id fid mid sex dx
complete trios(320):

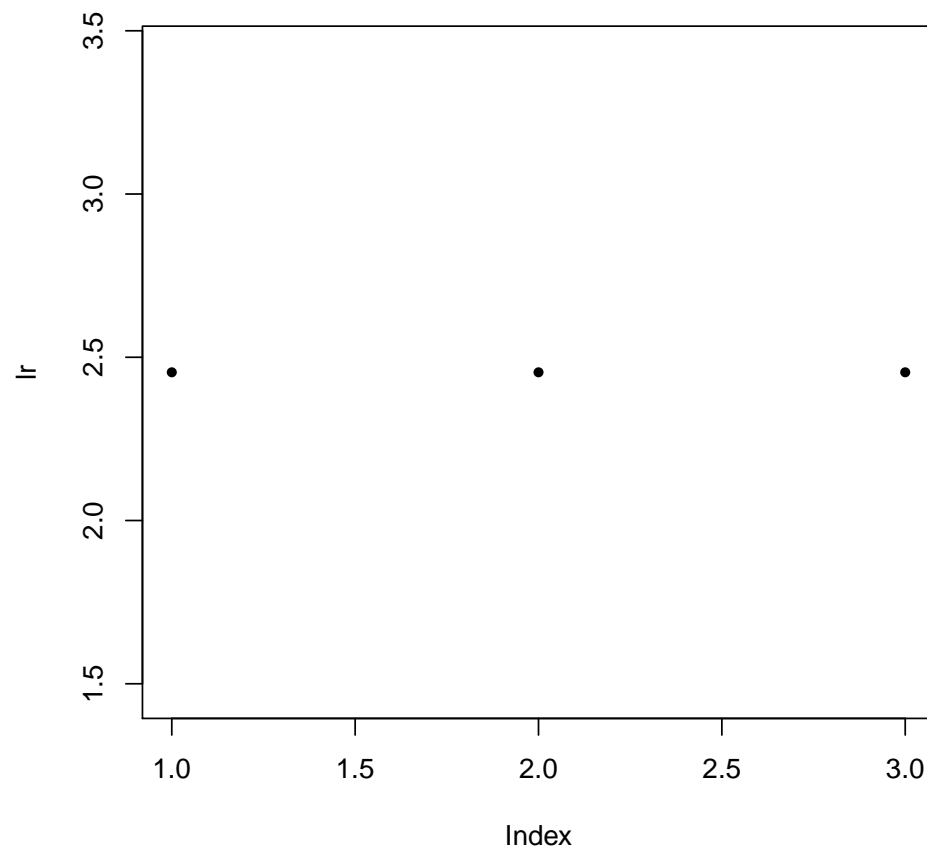
> summary(MAF(ste.rare))

      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
0.0000000 0.0007812 0.0008065 0.0474000 0.0164100 0.5000000

> (scan.trio <- ScanTrio(object = ste.rare, window = rowData(ste.rare[1:3]) +
  250000, block = range(rowData(ste.rare))))

DataFrame with 3 rows and 7 columns
      lr  minor.in  major.in  minor.out  major.out  mendel.in
<numeric> <numeric> <numeric> <numeric> <numeric> <numeric>
1  2.453997    1478    1699    3970    4315      0
2  2.453997    1478    1699    3970    4315      0
3  2.453997    1478    1699    3970    4315      0
mendel.out
<numeric>
1      0
2      0
3      0
```

```
> with(as(scan.trio, "data.frame"), plot(lr, pch = 20))
```



5.2 Holger-style Genotype Matrix

Coercion to matrix for *trio*

```
> geno <- as(ste, "matrix")
> aTDT(geno[, 1:5])
```

```
$stat
[1] 16.13333 26.47059 21.55172 20.16129 35.00000
```

```
$pval
[1] 5.903578e-05 2.675813e-07 3.444129e-06 7.117887e-06 3.297053e-09
```

```
$transMinor
[1] 4 2 2 3 0
```

```
$transMajor
[1] 26 32 27 28 35
```

Or apply the aTDT method to the SNP Trio Experiment directly.

```
> aTDT(ste[1:5])
```

```
$stat
[1] 23.405063 3.368421 26.888889 27.000000 5.254545
```

```
$pval
[1] 1.312327e-06 6.645742e-02 2.154942e-07 2.034555e-07 2.188952e-02
```

```
$transMinor
[1] 18 30 14 15 19
```

```
$transMajor
[1] 61 46 58 60 36
```

A Count of Transmission of Variants (*TransCount()*)

A.1 Window

```
> window <- gr[100] + 10000
> window2 <- gr[100] + 1000
> block <- range(rowData(ste))

> TransCount(ste, window)
```

```
$minor
[1] 4007
```

```
$major
[1] 4249
```

```
$mendel
[1] 0
```

A.2 Not in window, but in block

```
> TransCount(ste, setdiff(block, window))
```

```
$minor
[1] 158538
```

```
$major
[1] 152271
```

```
$mendel
[1] 1
```

A.3 In the whole block

```
> TransCount(ste, block)
```

```
$minor
[1] 162545
```

```
$major
[1] 156520
```

```
$mendel
[1] 1
```

A.4 In both windows

```
> TransCount(ste, GRangesList(window, window2))
```

```
$minor
[1] 4007 468
```

```
$major
[1] 4249 521
```

```
$mendel
[1] 0 0
```

A.5 Not “In both windows,” but in block

```
> TransCount(ste, GRangesList(setdiff(block, window), setdiff(block,
  window2)))
```

```
$minor
[1] 158538 162077
```

```
$major
[1] 152271 155999
```

```
$mendel
[1] 1 1
```

B Count of Transmission of Rare Variants (*TransCount()*)

B.1 In both windows

B.2 Not “In both windows,” but in block

C Classes

C.1 SnpMatrix, DataFrame, GRanges

The four key ingredients are the SNP matrix, the pedigree information as a `DataFrame`, position of the SNPs given by a `GRanges` object, and covariate data given as a `DataFrame`.

```
> sm
```

```
A SnpMatrix with 960 rows and 8951 columns
Row names: H_ME-DS10776_2-DS10776_2 ... H_ME-DS11313_1-DS11313_1
Col names: chr8:129296000 ... chr8:130354790
```

```
> ped.DF
```

```
DataFrame with 4139 rows and 6 columns
```

	famid	id
	<character>	<character>
1	4778	H_ME-4778_1-4778_1.2
2	4778	H_ME-4778_2-4778_2.2
3	4778	H_ME-4778_3-4778_3.2
4	4783	H_ME-4783_1-4783_1.1
5	4783	H_ME-4783_2-4783_2.1

6	4783	H_ME-4783_3-4783_3.1	
7	20000492	H_ME-20000492_1-20000492_1.1	
8	20000492	H_ME-20000492_2-20000492_2_a.1	
9	20000492	H_ME-20000492_3-20000492_3.1	
...	
4131	DS12332	H_ME-DS12332_1-DS12332_1	
4132	DS12332	H_ME-DS12332_2-DS12332_2	
4133	DS12332	H_ME-DS12332_3-DS12332_3	
4134	DS99998	H_ME-DS10193_5-DS10193_5.2	
4135	DS99998	H_ME-DS10193_6-DS10193_6	
4136	DS99998	H_ME-DS10193_7-DS10193_7	
4137	DS99999	H_ME-DS10707_4-DS10707_4.2	
4138	DS99999	H_ME-DS10707_5-DS10707_5	
4139	DS99999	H_ME-DS10707_6-DS10707_6	
		fid	mid
		<character>	<character>
1		H_ME-4778_2-4778_2.2	H_ME-4778_3-4778_3.2
2		NA	NA
3		NA	NA
4		H_ME-4783_2-4783_2.1	H_ME-4783_3-4783_3.1
5		NA	NA
6		NA	NA
7	H_ME-20000492_2-20000492_2_a.1	H_ME-20000492_3-20000492_3.1	
8		NA	NA
9		NA	NA
...	
4131		H_ME-DS12332_2-DS12332_2	H_ME-DS12332_3-DS12332_3
4132		NA	NA
4133		NA	NA
4134		H_ME-DS10193_6-DS10193_6	H_ME-DS10193_7-DS10193_7
4135		NA	NA
4136		NA	NA
4137		NA	NA
4138		H_ME-DS10707_4-DS10707_4.2	H_ME-DS10707_6-DS10707_6
4139		NA	NA
	sex	dx	
	<numeric>	<numeric>	
1	1	1	
2	1	0	
3	2	0	
4	2	1	
5	1	0	
6	2	1	
7	2	1	
8	1	0	
9	2	0	
...	
4131	1	1	
4132	1	0	
4133	2	0	
4134	2	1	

```

4135      1      0
4136      2      0
4137      1      0
4138      1      1
4139      2      0

```

```
> gr
```

GRanges with 8951 ranges and 0 metadata columns:

```

      seqnames      ranges strand
      <Rle>         <IRanges> <Rle>
chr8:129296000 chr8 [129296000, 129296000] *
chr8:129296113 chr8 [129296113, 129296113] *
chr8:129296185 chr8 [129296185, 129296185] *
chr8:129296191 chr8 [129296191, 129296191] *
chr8:129296198 chr8 [129296198, 129296198] *
chr8:129296209 chr8 [129296209, 129296209] *
chr8:129296289 chr8 [129296289, 129296289] *
chr8:129296343 chr8 [129296343, 129296343] *
chr8:129296434 chr8 [129296434, 129296434] *
...
chr8:130353593 chr8 [130353593, 130353593] *
chr8:130353671 chr8 [130353671, 130353671] *
chr8:130354142 chr8 [130354142, 130354142] *
chr8:130354182 chr8 [130354182, 130354182] *
chr8:130354239 chr8 [130354239, 130354239] *
chr8:130354240 chr8 [130354240, 130354240] *
chr8:130354296 chr8 [130354296, 130354296] *
chr8:130354703 chr8 [130354703, 130354703] *
chr8:130354790 chr8 [130354790, 130354790] *
---
seqlengths:
chr8
NA

```

C.2 SummarizedExperiment

We combine three of the key ingredients when we create the SummarizedExperiment object.

```

> se <- SummarizedExperiment(assays = SimpleList(geno = t(sm)),
  colData = col.DF, rowData = gr)

```

C.3 SNP TrioExperiment

Now, we include the pedigree information as an object of class PedClass. We keep PedClass independent of SNP TrioExperiment for flexibility.

```

> ped <- PedClass(ped.DF)
> ste <- SNP TrioExperiment(se, pedigree = ped)

```

Here is the show method.


```

> ste

class: SNP Trio Experiment
dim: 8951 960
exptData(0):
assays(1): geno
rownames(8951): chr8:129296000 chr8:129296113 ...
               chr8:130354703 chr8:130354790
rowData metadata column names(0):
colnames(960): H_ME-DS10776_2-DS10776_2
               H_ME-DS10776_3-DS10776_3 ... H_ME-DS11313_3-DS11313_3
               H_ME-DS11313_1-DS11313_1
colData names(1): id
pedigree(4139): famid id fid mid sex dx
complete trios(320):

```

And now we verify that it is indeed an extension of SummarizedExperiment.

```

> getClass("SNPTrioExperiment")

Class "SNPTrioExperiment" [package "trioClasses"]

Slots:

Name:                pedigree                exptData
Class:                PedClass                SimpleList

Name:                rowData                colData
Class: GenomicRangesORGRangesList          DataFrame

Name:                assays
Class:                Assays

Extends: "SummarizedExperiment"

```

C.4 PedClass

Now we investigate the pedigree slot of the SNP Trio Experiment object.

```

> class(pedigree(ste))

[1] "PedClass"
attr(,"package")
[1] "trioClasses"

> getClass("PedClass")

Class "PedClass" [package "trioClasses"]

Slots:

Name:                rownames                nrows                listData

```

```

Class: characterORNULL      integer      list

Name:      elementType elementMetadata  metadata
Class:      character DataTableORNULL   list

```

Extends:

```

Class "DataFrame", directly
Class "DataTable", by class "DataFrame", distance 2
Class "SimpleList", by class "DataFrame", distance 2
Class "DataTableORNULL", by class "DataFrame", distance 3
Class "List", by class "DataFrame", distance 3
Class "Vector", by class "DataFrame", distance 4
Class "Annotated", by class "DataFrame", distance 5

```

```
> pedigree(ste)
```

PedClass with 4139 rows and 6 columns

	famid		id
	<factor>		<factor>
1	4778	H_ME-4778_1-4778_1.2	
2	4778	H_ME-4778_2-4778_2.2	
3	4778	H_ME-4778_3-4778_3.2	
4	4783	H_ME-4783_1-4783_1.1	
5	4783	H_ME-4783_2-4783_2.1	
6	4783	H_ME-4783_3-4783_3.1	
7	20000492	H_ME-20000492_1-20000492_1.1	
8	20000492	H_ME-20000492_2-20000492_2_a.1	
9	20000492	H_ME-20000492_3-20000492_3.1	
...
4131	DS12332	H_ME-DS12332_1-DS12332_1	
4132	DS12332	H_ME-DS12332_2-DS12332_2	
4133	DS12332	H_ME-DS12332_3-DS12332_3	
4134	DS99998	H_ME-DS10193_5-DS10193_5.2	
4135	DS99998	H_ME-DS10193_6-DS10193_6	
4136	DS99998	H_ME-DS10193_7-DS10193_7	
4137	DS99999	H_ME-DS10707_4-DS10707_4.2	
4138	DS99999	H_ME-DS10707_5-DS10707_5	
4139	DS99999	H_ME-DS10707_6-DS10707_6	
		fid	mid
		<factor>	<factor>
1		H_ME-4778_2-4778_2.2	H_ME-4778_3-4778_3.2
2		NA	NA
3		NA	NA
4		H_ME-4783_2-4783_2.1	H_ME-4783_3-4783_3.1
5		NA	NA
6		NA	NA
7	H_ME-20000492_2-20000492_2_a.1	H_ME-20000492_3-20000492_3.1	
8		NA	NA
9		NA	NA
...	
4131	H_ME-DS12332_2-DS12332_2	H_ME-DS12332_3-DS12332_3	
4132		NA	NA

4133		NA		NA
4134	H_ME-DS10193_6-DS10193_6		H_ME-DS10193_7-DS10193_7	
4135		NA		NA
4136		NA		NA
4137		NA		NA
4138	H_ME-DS10707_4-DS10707_4.2		H_ME-DS10707_6-DS10707_6	
4139		NA		NA

	sex	dx
	<factor>	<factor>
1	1	1
2	1	0
3	2	0
4	2	1
5	1	0
6	2	1
7	2	1
8	1	0
9	2	0
...
4131	1	1
4132	1	0
4133	2	0
4134	2	1
4135	1	0
4136	2	0
4137	1	0
4138	1	1
4139	2	0

C.5 geno accessor

```
> class(geno(ste))

[1] "SnpMatrix"
attr(,"package")
[1] "snpStats"

> getClass("SnpMatrix")

Class "SnpMatrix" [package "snpStats"]

Slots:

Name:    .Data
Class: matrix

Extends:
Class "matrix", from data part
Class "array", by class "matrix", distance 2
Class "mMatrix", by class "matrix", distance 2
Class "structure", by class "matrix", distance 3
Class "vector", by class "matrix", distance 4, with explicit coerce
```

Known Subclasses: "XSnpmatrix"

```
> geno(ste)
```

A SnpMatrix with 960 rows and 8951 columns

Row names: H_ME-DS10776_2-DS10776_2 ... H_ME-DS11313_1-DS11313_1

Col names: chr8:129296000 ... chr8:130354790

C.6 RowData

```
> rowData(ste)
```

GRanges with 8951 ranges and 0 metadata columns:

	seqnames	ranges	strand
	<Rle>	<IRanges>	<Rle>
chr8:129296000	chr8 [129296000, 129296000]		*
chr8:129296113	chr8 [129296113, 129296113]		*
chr8:129296185	chr8 [129296185, 129296185]		*
chr8:129296191	chr8 [129296191, 129296191]		*
chr8:129296198	chr8 [129296198, 129296198]		*
chr8:129296209	chr8 [129296209, 129296209]		*
chr8:129296289	chr8 [129296289, 129296289]		*
chr8:129296343	chr8 [129296343, 129296343]		*
chr8:129296434	chr8 [129296434, 129296434]		*
...
chr8:130353593	chr8 [130353593, 130353593]		*
chr8:130353671	chr8 [130353671, 130353671]		*
chr8:130354142	chr8 [130354142, 130354142]		*
chr8:130354182	chr8 [130354182, 130354182]		*
chr8:130354239	chr8 [130354239, 130354239]		*
chr8:130354240	chr8 [130354240, 130354240]		*
chr8:130354296	chr8 [130354296, 130354296]		*
chr8:130354703	chr8 [130354703, 130354703]		*
chr8:130354790	chr8 [130354790, 130354790]		*

seqlengths:
chr8
NA

C.7 ColData

D Validity