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 13, 2013

1 Packages & Data

```
> library("trioClasses")
> data("sample")
> data("BMP4-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: 511 968
exptData(0):
assays(1): geno
rownames(511): chr14:54383433 chr14:54383470 ...
  chr14:54444385 chr14:54444750
rowData metadata column names(0):
colnames(968): 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):
```

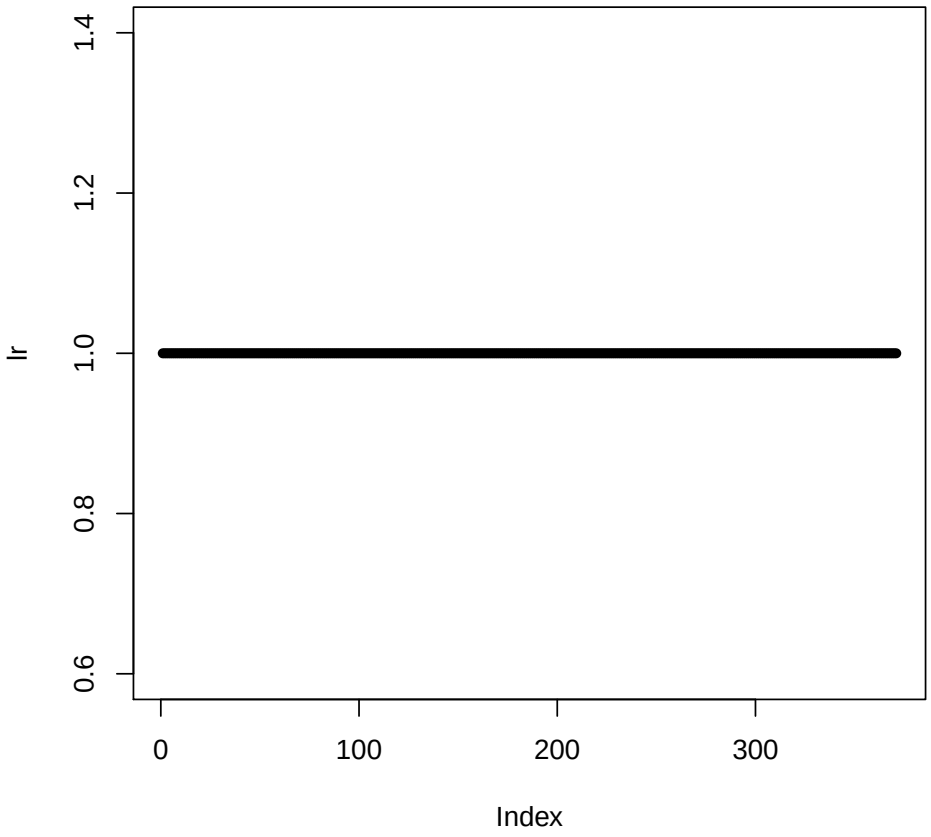
5 Methods

5.1 ScanTrio

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

class: SNPTrioExperiment
dim: 371 968
exptData(0):
assays(1): geno
rownames(371): chr14:54383433 chr14:54383819 ...
  chr14:54444016 chr14:54444385
rowData metadata column names(0):
colnames(968): 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 0.0007825 0.0015940 0.0015630 0.0098250

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

DataFrame with 371 rows and 7 columns
           lr  minor.in  major.in minor.out major.out mendel.in
    <numeric> <integer> <integer> <integer> <integer> <integer>
1           1       330       361         0         0         0
2           1       330       361         0         0         0
3           1       330       361         0         0         0
4           1       330       361         0         0         0
5           1       330       361         0         0         0
6           1       330       361         0         0         0
7           1       330       361         0         0         0
8           1       330       361         0         0         0
9           1       330       361         0         0         0
...       ...       ...       ...       ...       ...       ...
363         1       330       361         0         0         0
364         1       330       361         0         0         0
365         1       330       361         0         0         0
366         1       330       361         0         0         0
367         1       330       361         0         0         0
368         1       330       361         0         0         0
369         1       330       361         0         0         0
370         1       330       361         0         0         0
371         1       330       361         0         0         0
    mendel.out
     <integer>
```

1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
...	...
363	0
364	0
365	0
366	0
367	0
368	0
369	0
370	0
371	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] 24.50000 15.11429 27.12903 11.30769 10.31429

$pval
[1] 7.430984e-07 1.011946e-04 1.903182e-07 7.718664e-04 1.320044e-03

$transMinor
[1] 2 6 1 9 8

$transMajor
[1] 30 29 30 30 27
```

Or apply the aTDT method to the SNPTrioExperiment directly.

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

$stat
[1] 18.000000 16.780822  3.813559  7.736842  6.391304

$pval
[1] 2.209050e-05 4.195519e-05 5.083931e-02 5.410534e-03 1.146807e-02

$transMinor
[1] 18 19 22 18 24

$transMajor
[1] 54 54 37 39 45
```

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] 3286

$major
[1] 3513

$mendel
[1] 0
```

A.2 Not in window, but in block

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

$minor
[1] 6522

$major
[1] 6712

$mendel
[1] 0
```

A.3 In the whole block

```
> TransCount(ste, block)

$minor
[1] 9808

$major
[1] 10225

$mendel
[1] 0
```

A.4 In both windows

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

$minor
[1] 3286  188
```

```
$major
[1] 3513  195

$mendel
[1] 0 0
```

A.5 Not "In both windows," but in block

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

$minor
[1] 6522 9620

$major
[1]  6712 10030

$mendel
[1] 0 0
```

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  968 rows and  511 columns
Row names:  H_ME-DS10776_2-DS10776_2 ... H_ME-DS11313_1-DS11313_1
Col names:  chr14:54383433 ... chr14:54444750

> 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 511 ranges and 0 metadata columns:
                 seqnames              ranges strand
                    <Rle>           <IRanges>  <Rle>
  chr14:54383433    chr14 [54383433, 54383433]      *
  chr14:54383470    chr14 [54383470, 54383470]      *
  chr14:54383819    chr14 [54383819, 54383819]      *
  chr14:54383945    chr14 [54383945, 54383945]      *
  chr14:54384030    chr14 [54384030, 54384030]      *
  chr14:54384232    chr14 [54384232, 54384232]      *
  chr14:54384251    chr14 [54384251, 54384251]      *
  chr14:54384291    chr14 [54384291, 54384291]      *
  chr14:54384382    chr14 [54384382, 54384382]      *
             ...      ...                 ...    ...
  chr14:54443495    chr14 [54443495, 54443495]      *
  chr14:54443499    chr14 [54443499, 54443499]      *
  chr14:54443514    chr14 [54443514, 54443514]      *
  chr14:54443831    chr14 [54443831, 54443831]      *
  chr14:54443880    chr14 [54443880, 54443880]      *
  chr14:54444016    chr14 [54444016, 54444016]      *
  chr14:54444124    chr14 [54444124, 54444124]      *
  chr14:54444385    chr14 [54444385, 54444385]      *
  chr14:54444750    chr14 [54444750, 54444750]      *
  ---
  seqlengths:
   chr14
      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 SNPTrioExperiment

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

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

Here is the show method.

```
> ste

class: SNPTrioExperiment
dim: 511 968
exptData(0):
assays(1): geno
rownames(511): chr14:54383433 chr14:54383470 ...
  chr14:54444385 chr14:54444750
rowData metadata column names(0):
colnames(968): 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 SNPTrioExperiment 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  968 rows and  511 columns
Row names:  H_ME-DS10776_2-DS10776_2 ... H_ME-DS11313_1-DS11313_1
Col names:  chr14:54383433 ... chr14:54444750
```

C.6 RowData

```
> rowData(ste)
```

```
GRanges with 511 ranges and 0 metadata columns:
                 seqnames              ranges strand
                    <Rle>           <IRanges>  <Rle>
  chr14:54383433    chr14 [54383433, 54383433]      *
  chr14:54383470    chr14 [54383470, 54383470]      *
  chr14:54383819    chr14 [54383819, 54383819]      *
  chr14:54383945    chr14 [54383945, 54383945]      *
  chr14:54384030    chr14 [54384030, 54384030]      *
  chr14:54384232    chr14 [54384232, 54384232]      *
  chr14:54384251    chr14 [54384251, 54384251]      *
  chr14:54384291    chr14 [54384291, 54384291]      *
  chr14:54384382    chr14 [54384382, 54384382]      *
             ...      ...                 ...    ...
  chr14:54443495    chr14 [54443495, 54443495]      *
  chr14:54443499    chr14 [54443499, 54443499]      *
  chr14:54443514    chr14 [54443514, 54443514]      *
  chr14:54443831    chr14 [54443831, 54443831]      *
  chr14:54443880    chr14 [54443880, 54443880]      *
  chr14:54444016    chr14 [54444016, 54444016]      *
  chr14:54444124    chr14 [54444124, 54444124]      *
  chr14:54444385    chr14 [54444385, 54444385]      *
  chr14:54444750    chr14 [54444750, 54444750]      *
  ---
  seqlengths:
   chr14
      NA
```

C.7 ColData

D Validity