

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, Robert Scharpf, Holger Schwender, Ingo Ruczinski

March 6, 2013

## 1 Packages

```
> library("trioClasses")
```

## 2 Sample Data

```
> data("sample")
> se <- SummarizedExperiment(assays = SimpleList(geno = t(sm)),
  colData = col.DF, rowData = gr)
> ped <- PedClass(ped.DF)
> ste <- SNPTrioExperiment(se, pedigree = ped)
```

```
> ste
```

```
class: SNPTrioExperiment
dim: 511 968
exptData(0):
assays(1): geno
rownames(511): chr14:54383433 chr14:54383470 ...
  chr14:54444385 chr14:54444750
rowData values names(0):
colnames(968): H_ME-CSc175_1-CSc175_1_a.1
  H_ME-CSc175_2-CSc175_2 ... H_ME-DS11418_2-DS11418_2
  H_ME-DS11418_3-DS11418_3
colData names(28): vcfid jfb_famid ... id ped
pedigree(4139): famid id fid mid sex dx
```

## 3 Methods

Now create a list of three SnpMatrices for trio analyses using `GenoTrio()`.

```
> (gtrio <- GenoTrio(ste))
```

```
$0
```

```
A SnpMatrix with 320 rows and 511 columns
```

```

Row names: H_ME-CSc175_1-CSc175_1_a.1 ... H_ME-DS11418_1-DS11418_1
Col names: chr14:54383433 ... chr14:54444750

```

```
$F
```

```

A SnpMatrix with 320 rows and 511 columns
Row names: H_ME-CSc175_2-CSc175_2 ... H_ME-DS11418_2-DS11418_2
Col names: chr14:54383433 ... chr14:54444750

```

```
$M
```

```

A SnpMatrix with 320 rows and 511 columns
Row names: H_ME-CSc175_3-CSc175_3 ... H_ME-DS11418_3-DS11418_3
Col names: chr14:54383433 ... chr14:54444750

```

Now we can do things easily, such as count transmissions.

```
> with(gtrio, sum(F == 0 & M == 1 & O == 1, na.rm = TRUE))
```

```
[1] 411
```

Or run Holger's functions.

```

> geno <- ctcbind(gtrio)
> table(aTDT(geno)$transMinor)

```

	0	1	2	3	4	5	7	8	9	10	11	12	14	15	17	18	19
297	72	18	8	8	2	1	5	7	3	1	1	3	1	2	1	2	
20	21	22	23	24	26	27	28	31	33	44	45	46	48	56	58	68	
3	1	1	1	1	2	2	2	1	1	1	1	1	1	1	1	1	1
76	85	90	91	92	94	96	98	99	100	103	105	106	110	111	112	113	
1	1	2	1	1	1	1	2	1	2	1	1	1	2	2	3	2	
114	125	128	129	130	132	133	134	135	138	141	142	143	144	145	151	152	
1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	
157	158	159	164	165	166	167	168	169	175	176							
1	1	1	1	2	1	1	1	2	1	1							

I haven't confirmed yet that TransCount is working properly. The counts seem high, but I'm not sure if the comparison is correct.

```
> TransCount(ste, gr[1])
```

```
[1] 319
```

## 4 Classes

### 4.1 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	
	famid		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 elementMetadata cols:

	seqnames	ranges	strand
	<Rle>	<IRanges>	<Rle>
chr14:54383433	chr14	[1, 1]	*
chr14:54383470	chr14	[2, 2]	*
chr14:54383819	chr14	[3, 3]	*
chr14:54383945	chr14	[4, 4]	*
chr14:54384030	chr14	[5, 5]	*
chr14:54384232	chr14	[6, 6]	*
chr14:54384251	chr14	[7, 7]	*
chr14:54384291	chr14	[8, 8]	*
chr14:54384382	chr14	[9, 9]	*
...	...	...	...
chr14:54443495	chr14	[503, 503]	*
chr14:54443499	chr14	[504, 504]	*
chr14:54443514	chr14	[505, 505]	*
chr14:54443831	chr14	[506, 506]	*
chr14:54443880	chr14	[507, 507]	*
chr14:54444016	chr14	[508, 508]	*
chr14:54444124	chr14	[509, 509]	*
chr14:54444385	chr14	[510, 510]	*
chr14:54444750	chr14	[511, 511]	*

```

---
seqlengths:
  chr14
    NA

```

## 4.2 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)

```

## 4.3 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 TrioExperiment
dim: 511 968
exptData(0):
assays(1): geno
rownames(511): chr14:54383433 chr14:54383470 ...
  chr14:54444385 chr14:54444750
rowData values names(0):
colnames(968): H_ME-CSc175_1-CSc175_1_a.1
  H_ME-CSc175_2-CSc175_2 ... H_ME-DS11418_2-DS11418_2
  H_ME-DS11418_3-DS11418_3
colData names(28): vcfid jb_famid ... id ped
pedigree(4139): famid id fid mid sex dx

```

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

```

> getClass("SNP TrioExperiment")

```

```

Class "SNP TrioExperiment" [package "trioClasses"]

```

Slots:

Name:	pedigree	exptData
Class:	PedClass	SimpleList
Name:	rowData	colData
Class:	GenomicRangesORGRangesList	DataFrame
Name:	assays	

```
Class: SimpleList
```

```
Extends: "SummarizedExperiment"
```

## 4.4 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	

## 4.5 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 "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-CSc175_1-CSc175_1_a.1 ... H_ME-DS11418_3-DS11418_3
Col names:  chr14:54383433 ... chr14:54444750
```

## 4.6 RowData

```
> rowData(ste)

GRanges with 511 ranges and 0 elementMetadata cols:
      seqnames      ranges strand
      <Rle>      <IRanges> <Rle>
chr14:54383433 chr14      [1, 1]      *
chr14:54383470 chr14      [2, 2]      *
chr14:54383819 chr14      [3, 3]      *
chr14:54383945 chr14      [4, 4]      *
chr14:54384030 chr14      [5, 5]      *
chr14:54384232 chr14      [6, 6]      *
chr14:54384251 chr14      [7, 7]      *
chr14:54384291 chr14      [8, 8]      *
chr14:54384382 chr14      [9, 9]      *
      ...      ...      ...
chr14:54443495 chr14 [503, 503]      *
chr14:54443499 chr14 [504, 504]      *
chr14:54443514 chr14 [505, 505]      *
chr14:54443831 chr14 [506, 506]      *
```



```

chr14:54443880    chr14 [507, 507]    *
chr14:54444016    chr14 [508, 508]    *
chr14:54444124    chr14 [509, 509]    *
chr14:54444385    chr14 [510, 510]    *
chr14:54444750    chr14 [511, 511]    *
---
seqlengths:
chr14
NA

```

## 4.7 ColData

## 5 Validity