

GGBase: infrastructure for genetics of gene expression

VJ Carey

October 28, 2009

1 Introduction

Data representations for genetics of gene expression in humans must be as efficient as possible. Up to Bioconductor 2.1, integrative genomics experiments could be represented by `racExSet` instances as defined in package `GGtools`. Genotypes were represented as integers.

For Bioconductor 2.2, we have access to the `snpMatrix` package of D. Clayton, and we have introduced the `smlSet` class to represent genetics of gene expression studies with lists of `snp.matrix` instances, along with ordinary representations of expression data.

2 Primary class structure, and associated methods

```
> library(GGBase)
> getClass("smlSet")
```

```
Class "smlSet" [package "GGBase"]
```

```
Slots:
```

```
Name:          smlEnv          annotation          organism
Class:         environment     character          character
```

```
Name:          assayData      phenoData          featureData
Class:         AssayData      AnnotatedDataFrame AnnotatedDataFrame
```

```
Name:          experimentData  protocolData      .__classVersion__
Class:         MIAME          AnnotatedDataFrame Versions
```

```
Extends:
```

```
Class "eSet", directly
```

Class "VersionedBiobase", by class "eSet", distance 2

Class "Versioned", by class "eSet", distance 3

```
> library(GGtools)
```

```
> showMethods(class = "smlSet", where = "package:GGBase")
```

Function: [(package base)

x="smlSet", i="ANY", j="ANY", drop="ANY"

Function: coerce (package methods)

from="smlSet", to="ExpressionSet"

Function: combine (package Biobase)

x="smlSet", y="smlSet"

Function: exprs (package Biobase)

object="smlSet"

Function: getAlleles (package GGBase)

x="smlSet", rs="rsid"

Function: getSnpLocs (package GGBase)

x="smlSet", c="chrnum"

x="smlSet", c="missing"

Function: plot_EvG (package GGBase)

gsym="genesym", rsid="rsid", sms="smlSet"

gsym="probeId", rsid="rsid", sms="smlSet"

Function: plot_EvG2 (package GGBase)

gsym="genesym", rsid1="rsid", rsid2="rsid", sms="smlSet"

gsym="probeId", rsid1="rsid", rsid2="rsid", sms="smlSet"

Function: show (package methods)

object="smlSet"

Function: smlEnv (package GGBase)

x="smlSet"

Function: smList (package GGBase)

x="smlSet"

Function: snpNames (package GGBase)

```
x="smlSet", c="chrnum"  
x="smlSet", c="missing"
```

```
Function: snps (package GGBase)  
x="smlSet", chr="chrnum"
```

```
> showMethods(class = "smlSet", where = "package:GGtools")
```

```
Function: [ (package base)  
x="smlSet", i="ANY", j="ANY", drop="ANY"
```

```
Function: coerce (package methods)  
from="smlSet", to="ExpressionSet"
```

```
Function: gwSnpTests (package GGtools)  
sym="formula", sms="smlSet", cnum="cnumOrMissing", cs="missing"  
sym="formula", sms="smlSet", cnum="snpdepth", cs="chunksize"  
sym="formula", sms="smlSet", cnum="snpdepth", cs="missing"
```

```
Function: hbTests (package GGtools)  
fmla="genesym", sms="smlSet", cnum="chrnum", rsid="numeric", rad="numeric"
```

```
Function: residTests (package GGtools)  
fit="cwSnpScreenResult", sms="smlSet", litfmla="formula", rsnum="missing"
```

3 Example data structure

```
> data(hmceuB36.2021)  
> hmceuB36.2021
```

```
snp.matrix-based genotype set:  
number of samples: 90  
number of chromosomes present: 2  
annotation: illuminaHumanv1.db  
Expression data dims: 47293 x 90
```

```
Phenodata: An object of class "AnnotatedDataFrame"
```

```
sampleNames: NA06985, NA06991, ..., NA12892 (90 total)
```

```
varLabels and varMetadata description:
```

```
famid: hapmap family id  
persid: hapmap person id  
...: ...  
male: logical TRUE if male  
(7 total)
```

4 SNP location infrastructure

see the `newSNPloc` vignette

5 Conclusion

The concepts discussed above define the bulk of the tools supporting the `gwSnpTests` method of `GGtools`. Consult the vignette of `GGtools` for more information.