

# GGBase: infrastructure for genetics of gene expression

VJ Carey

April 14, 2011

## 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 2.8, we have access to the `snpStats` package of D. Clayton, and we have introduced the `smlSet` class to represent genetics of gene expression studies with lists of `Snpmatrix` 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:	MIxEx	AnnotatedDataFrame	Versions

```
Extends:
```

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

```

Class "VersionedBiobase", by class "eSet", distance 2
Class "Versioned", by class "eSet", distance 3

> 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: plot_EvG2 (package GGBase)
gsym="genesym", rsid1="rsid", rsid2="rsid", sms="smlSet"
gsym="probeId", rsid1="rsid", rsid2="rsid", sms="smlSet"

Function: plot_EvG (package GGBase)
gsym="genesym", rsid="rsid", sms="smlSet"
gsym="probeId", rsid="rsid", sms="smlSet"

Function: show (package methods)
object="smlSet"

Function: smList (package GGBase)
x="smlSet"

Function: smlEnv (package GGBase)
x="smlSet"

Function: snpNames (package GGBase)
x="smlSet", c="chrnum"
x="smlSet", c="missing"

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

```

```
Function: updateObject (package Biobase)
object="smlSet"
```

### 3 Example data structure

```
> data(smlSet.example)
> smlSet.example
```

```
SnpMatrix-based genotype set:
number of samples: 10
number of chromosomes present: 1
annotation: illuminaHumanv1.db
Expression data dims: 100 x 10
Phenodata: An object of class "AnnotatedDataFrame"
  sampleNames: NA06985 NA06991 ... NA07048 (10 total)
  varLabels: famid persid ... male (7 total)
  varMetadata: labelDescription
```

### 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.