

# GGBase: infrastructure for genetics of gene expression

VJ Carey

May 21, 2008

## 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")
```

Slots:

Name:	smlEnv	snpLocPathMaker	chromInds
Class:	environment	function	numeric
Name:	organism	snpLocPackage	snpLocRef
Class:	character	character	character
Name:	activeSnpInds	assayData	phenoData
Class:	numeric	AssayData	AnnotatedDataFrame
Name:	featureData	experimentData	annotation
Class:	AnnotatedDataFrame	MIAME	character
Name:	.__classVersion__		
Class:	Versions		

```

Extends:
Class "eSet", directly
Class "VersionedBiobase", by class "eSet", distance 2
Class "Versioned", by class "eSet", distance 3

> library(GGtools)
> showMethods(class = "smlSet")

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

Function: exprs (package Biobase)
object="smlSet"

Function: getAlleles (package GGtools)
x="smlSet", rs="rsNum"

Function: getSnpChroms (package GGtools)
x="smlSet", filterActive="logical"
x="smlSet", filterActive="missing"

Function: getSnpLocs (package GGtools)
x="smlSet", filterActive="logical"
x="smlSet", filterActive="missing"

Function: gwSnpScreen (package GGtools)
sym="GeneSet", sms="smlSet", cnum="cnumOrMissing"
sym="genesym", sms="smlSet", cnum="cnumOrMissing"

Function: nsamp (package GGtools)
x="smlSet"

Function: plot_EvG (package GGtools)
gsym="genesym", rsn="rsNum", sms="smlSet"

Function: rawSNP (package GGtools)
x="smlSet", chrind="numeric"

Function: sampleNames (package Biobase)
object="smlSet"

Function: show (package methods)

```

```
object="smlSet"
```

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

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

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

### 3 Example data structure

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

```
snp.matrix-based genotype set:  
number of samples: 90  
number of snp.matrix: 2  
annotation:  
  exprs: illuminaHumanv1.db  
  snps: snp locs package: GGBase ; SQLite ref: hmceuAmbB36_23a_dbconn  
Expression data: 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  
    ...: ...  
    isAdad: logical TRUE if person is a father  
    (9 total)
```

### 4 SNP location infrastructure

We have decided to use a dynamically exported SQLite connection to provide access to SNP location information.

```
> conn = grep("dbconn", objects("package:GGBase"), value = TRUE)  
> dbListTables(get(conn))
```

```
[1] "hmceuAmbB36_23a"
```

## 5 Conclusion

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