

Infrastructure classes for high-throughput SNP data

Robert Scharpf and Benilton Carvalho

April 23, 2010

This document describes some of the infrastructure classes used for high-throughput genomic data. For the classes used to organize SNP data, we provide examples for initialization and illustrate some of the accessors. We should add a diagram showing the relationships of these classes here.

[Insert diagram of classes here]

1 Feature-level classes

2 Locus-level classes

The examples below are completely simulated and are not meant to convey any biological plausibility.

2.1 SnpSet

2.1.1 Initialization

```
> theCalls <- matrix(sample(1:3, 20, rep = TRUE), nc = 2)
> p <- matrix(runif(20), nc = 2)
> theConfs <- round(-1000 * log2(1 - p))
> obj <- new("SnpSet", call = theCalls, callProbability = theConfs)
```

2.1.2 Accessors

```
> calls(obj)
```

```
  1 2
1  1 3
2  3 3
3  2 2
4  1 3
5  1 1
6  3 2
7  1 3
8  3 3
9  2 3
10 3 1
```

```
> confs(obj)
```

```
      1      2
1 0.28251268 0.57894845
2 0.84649160 0.79505221
3 0.99977446 0.05823547
4 0.79749688 0.57044264
5 0.81985623 0.86826988
```

```

6 0.88695847 0.31064576
7 0.76636612 0.44953946
8 0.02175976 0.88328264
9 0.58811043 0.57343904
10 0.96431411 0.75414172

```

2.1.3 Annotating

```

> if (require("genomewidesnp6Crlmm")) {
+   ids <- c("SNP_A-2131660", "SNP_A-1967418", "SNP_A-1969580",
+           "SNP_A-4263484", "SNP_A-1978185", "SNP_A-4264431",
+           "SNP_A-1980898", "SNP_A-1983139", "SNP_A-4265735",
+           "SNP_A-1995832")
+   rownames(theCalls) <- rownames(p) <- rownames(theConfs) <- ids
+   obj <- new("SnpSet", call = theCalls, callProbability = theConfs,
+             annotation = "genomewidesnp6")
+   featureData(obj) <- addFeatureAnnotation(obj)
+   fvarLabels(obj)
+   isSnp(obj)
+   position(obj)
+   chromosome(obj)
+ }

[1] 1 1 1 1 1 1 1 1 1 1

```

2.2 CopyNumberSet

2.2.1 Initialization

2.2.2 Accessors

2.2.3 Annotating

2.3 CNSet

2.3.1 Initialization

```

> theCalls <- matrix(2, nc = 2, nrow = 10)
> A <- matrix(sample(1:1000, 20), 10, 2)
> B <- matrix(sample(1:1000, 20), 10, 2)
> CA <- matrix(rnorm(20, 1), nrow = 10)
> CB <- matrix(rnorm(20, 1), nrow = 10)
> p <- matrix(runif(20), nc = 2)
> theConfs <- round(-1000 * log2(1 - p))
> obj <- new("CNSet", alleleA = A, alleleB = B, call = theCalls,
+   callProbability = theConfs, CA = CA, CB = CB)

```

2.3.2 Accessors

```

> calls(obj)

  1 2
1  2 2
2  2 2
3  2 2
4  2 2
5  2 2

```

```

6 2 2
7 2 2
8 2 2
9 2 2
10 2 2

```

```
> confs(obj)
```

```

      1      2
1 0.1588624 0.9976633
2 0.6877653 0.4270747
3 0.3175049 0.5383739
4 0.8609609 0.4166687
5 0.6535442 0.6576768
6 0.6724125 0.4949006
7 0.9203410 0.3343558
8 0.9908225 0.4631304
9 0.6085901 0.9938049
10 0.7806310 0.2457263

```

```
> A(obj)
```

```

      1      2
1 347 899
2 673 587
3 676 915
4 635 494
5 815 484
6 651 121
7 141 316
8 217 723
9 689 283
10 877 820

```

```
> B(obj)
```

```

      1      2
1 45 862
2 380 29
3 498 891
4 924 342
5 590 746
6 824 869
7 956 543
8 565 568
9 167 345
10 609 657

```

```
> CA(obj)
```

```

      1      2
1 -0.07136137 2.1496627
2 -0.64879098 1.7766283
3 0.31819615 0.9688061
4 0.84603506 0.9993472
5 0.07840211 1.4285334

```

```

6  0.61193170 1.4393261
7  1.12563561 0.2411302
8  0.51695168 1.5089723
9  0.53264438 0.8345067
10 0.35703155 1.1735133

```

```
> CB(obj)
```

```

      1      2
1  2.14309748 -0.9813741
2  1.23043199  2.7021148
3  0.79560197  0.6277000
4  0.68796105 -0.1914998
5 -1.15229639  1.8255127
6  0.63692726  0.3648543
7 -0.08950161  0.5031745
8  1.61579794  0.8832304
9 -1.75972923  2.6174919
10 -0.81660838  2.0699319

```

2.3.3 Annotating

Annotating with chromosome and physical position:

```

> if (require("genomewidesnp6Crlmm")) {
+   ids <- c("SNP_A-2131660", "SNP_A-1967418", "SNP_A-1969580",
+           "SNP_A-4263484", "SNP_A-1978185", "SNP_A-4264431",
+           "SNP_A-1980898", "SNP_A-1983139", "SNP_A-4265735",
+           "SNP_A-1995832")
+   rownames(theCalls) <- rownames(p) <- rownames(theConfs) <- ids
+   rownames(A) <- rownames(B) <- rownames(CA) <- rownames(CB) <- ids
+   obj2 <- new("CNSet", alleleA = A, alleleB = B, call = theCalls,
+               callProbability = theConfs, CA = CA, CB = CB, annotation = "genomewidesnp6")
+   fvarLabels(obj2)
+   isSnp(obj2)
+   chromosome(obj2)
+   position(obj2)
+ }

```

3 Session Information

The version number of R and packages loaded for generating the vignette were:

- R version 2.11.0 (2010-04-22), x86_64-pc-mingw32
- Locale: LC_COLLATE=English_United States.1252, LC_CTYPE=English_United States.1252, LC_MONETARY=English_United States.1252, LC_NUMERIC=C, LC_TIME=English_United States.1252
- Base packages: base, datasets, graphics, grDevices, methods, stats, tools, utils
- Other packages: Biobase 2.8.0, genomewidesnp6Crlmm 1.0.2, oligoClasses 1.10.0
- Loaded via a namespace (and not attached): affyio 1.16.0, Biostrings 2.16.0, IRanges 1.6.0