

The SNPMap.cdm Package

July 5, 2008

Type Package

Title Annotation for SNP Microarrays and Pooling in R

Version 1.0.0

Date 2008-06-24

Author Oliver SP Davis, Leo C Schalkwyk

Maintainer Oliver SP Davis <snpmap@iop.kcl.ac.uk>

Description Pooling DNA on SNP microarrays is a cost-effective way to carry out genome-wide association studies for heritable disorders or traits. The SNPMap package provides formal SNPMap objects and methods in R as a base for these analyses using Affymetrix genotyping arrays. The SNPMap.cdm package provides cdm objects for the SNPMap package.

License GPL (>= 3)

LazyLoad yes

LazyData yes

Depends R (>= 2.4.0)

Enhances SNPMap

R topics documented:

GenomeWideSNP_5	2
GenomeWideSNP_6	2
Mapping250K_Nsp	3
Mapping250K_Sty	4
SNPMap.cdm-package	5

Index	8
--------------	----------

GenomeWideSNP_5 *cdm for the GenomeWideSNP_5 microarray*

Description

A [SNPMaP](#) `cdm` for the Affymetrix GenomeWideSNP_5 genotyping microarray.

Usage

```
GenomeWideSNP_5
```

Format

```
The format is:
List of 1
$ four.probes:List of 1
..$ perfect: int [1:440794, 1:8] 3594244 3674393 3884091 4063474 3431205 4588391 2636192
4345647 3459144 2686249 ...
.. .. attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:440794] "SNP_A-1780520" "SNP_A-1780618" "SNP_A-1780632" "SNP_A-
1780654" ...
.. .. ..$ : chr [1:8] "pma1" "pma2" "pma3" "pma4" ...
```

Details

The `cdm` is used to identify the probesets in the 'raw' `SNPMaP` object derived from the microarray CELs, achieving a similar function to an Affymetrix CDF file, but faster because it is held in memory once loaded. The `GenomeWideSNP_5` `cdm` has one `set` comprising 440794 probesets with four quartets.

GenomeWideSNP_6 *cdm for the GenomeWideSNP_6 microarray*

Description

A [SNPMaP](#) `cdm` for the Affymetrix GenomeWideSNP_6 genotyping microarray.

Usage

```
GenomeWideSNP_6
```

Format

```

The format is:
List of 2
$ three.probes:List of 1
..$ perfect: int [1:796045, 1:6] 5273360 4463256 2883438 810524 3933238 3961968 3158564
394894 4680958 4119394 ...
.. .. attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:796045] "SNP_A-2131660" "SNP_A-1967418" "SNP_A-1969580" "SNP_A-
4263484" ...
.. .. ..$ : chr [1:6] "pma1" "pma2" "pma3" "pmb1" ...
$ four.probes :List of 1
..$ perfect: int [1:110555, 1:8] 2839656 632482 4786064 1551942 5246520 672608 501298 2752888
6755706 2368568 ...
.. .. attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:110555] "SNP_A-4268887" "SNP_A-2006248" "SNP_A-2010253" "SNP_A-
2017872" ...
.. .. ..$ : chr [1:8] "pma1" "pma2" "pma3" "pma4" ...

```

Details

The `cdm` is used to identify the probesets in the 'raw' `SNPMaP` object derived from the microarray CELs, achieving a similar function to an Affymetrix CDF file, but faster because it is held in memory once loaded. The `GenomeWideSNP_6` `cdm` has two `sets`: The first comprises those 796045 probesets with three quartets, while the second comprises the 110555 probesets with four quartets. These are reached by specifying `set 1` or `2` respectively in a call to `snpmap()`. Alternatively, both sets can be read at once using `msnpmap()`.

Mapping250K_Nsp *cdm for the Mapping250K_Nsp microarray*

Description

A `SNPMaP` `cdm` for the Affymetrix Mapping250K_Nsp genotyping microarray that forms a 500K set with the Mapping250K_Sty.

Usage

```
Mapping250K_Nsp
```

Format

```

The format is:
List of 2
$ six.probes:List of 2
..$ perfect : int [1:252960, 1:12] 78152 2167382 2284272 2931842 548588 5440644 2799720
3330902 5113224 3410564 ...
.. .. attr(*, "dimnames")=List of 2

```

```

.. .. ..$ : chr [1:252960] "SNP_A-1780520" "SNP_A-1780618" "SNP_A-1780632" "SNP_A-1780654" ...
.. .. ..$ : chr [1:12] "pma1" "pma2" "pma3" "pma4" ...
..$ mismatch: int [1:252960, 1:12] 78153 2167383 2284273 2931843 548589 5440645 2799721 3330903 5113225 3410565 ...
.. ..- attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:252960] "SNP_A-1780520" "SNP_A-1780618" "SNP_A-1780632" "SNP_A-1780654" ...
.. .. ..$ : chr [1:12] "mma1" "mma2" "mma3" "mma4" ...
$ ten.probes:List of 2
..$ perfect : int [1:9304, 1:20] 5461834 5808782 1479170 3088726 2972016 3253158 6375046 1861456 6019908 938352 ...
.. ..- attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:9304] "SNP_A-4192498" "SNP_A-4192909" "SNP_A-4192918" "SNP_A-4192955"
...
.. .. ..$ : chr [1:20] "pma1" "pma2" "pma3" "pma4" ...
..$ mismatch: int [1:9304, 1:20] 5461835 5808783 1479171 3088727 2972017 3253159 6375047 1861457 6019909 938353 ...
.. ..- attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:9304] "SNP_A-4192498" "SNP_A-4192909" "SNP_A-4192918" "SNP_A-4192955"
...
.. .. ..$ : chr [1:20] "mma1" "mma2" "mma3" "mma4" ...

```

Details

The `cdm` is used to identify the probesets in the 'raw' `SNPMaP` object derived from the microarray CELs, achieving a similar function to an Affymetrix CDF file, but faster because it is held in memory once loaded. The `Mapping250K_Nsp cdm` includes mismatch probes as well as perfect match probes and has two `sets`: The first comprises those 252960 probesets with six quartets, while the second comprises the 9304 probesets with ten quartets. These are reached by specifying `set 1` or `set 2` respectively in a call to `snpmap()`. Alternatively, both sets can be read at once using `msnpmap()`. The mismatch probes are extracted by specifying `useMM=TRUE`.

Mapping250K_Sty *cdm for the Mapping250K_Sty microarray*

Description

A `SNPMaP cdm` for the Affymetrix `Mapping250K_Sty` genotyping microarray that forms a 500K set with the `Mapping250K_Nsp`.

Usage

```
Mapping250K_Sty
```

Format

```

The format is:
List of 2
$ six.probes:List of 2
..$ perfect : int [1:195567, 1:12] 3110922 763382 679536 1173994 2062218 1863554 6463496
531018 3433740 508600 ...
.. .. attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:195567] "SNP_A-1855402" "SNP_A-4249904" "SNP_A-2174835" "SNP_A-
1880271" ...
.. .. ..$ : chr [1:12] "pma1" "pma2" "pma3" "pma4" ...
..$ mismatch: int [1:195567, 1:12] 3110923 763383 679537 1173995 2062219 1863555 6463497
531019 3433741 508601 ...
.. .. attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:195567] "SNP_A-1855402" "SNP_A-4249904" "SNP_A-2174835" "SNP_A-
1880271" ...
.. .. ..$ : chr [1:12] "mma1" "mma2" "mma3" "mma4" ...
$ ten.probes:List of 2
..$ perfect : int [1:42737, 1:20] 1979704 3466744 6515338 457896 1439108 4471518 1145816
739218 4053004 5022546 ...
.. .. attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:42737] "SNP_A-4246731" "SNP_A-4247531" "SNP_A-4289595" "SNP_A-4258429"
...
.. .. ..$ : chr [1:20] "pma1" "pma2" "pma3" "pma4" ...
..$ mismatch: int [1:42737, 1:20] 1979705 3466745 6515339 457897 1439109 4471519 1145817
739219 4053005 5022547 ...
.. .. attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:42737] "SNP_A-4246731" "SNP_A-4247531" "SNP_A-4289595" "SNP_A-4258429"
...
.. .. ..$ : chr [1:20] "mma1" "mma2" "mma3" "mma4" ...

```

Details

The `cdm` is used to identify the probesets in the 'raw' SNPMaP object derived from the microarray CELs, achieving a similar function to an Affymetrix CDF file, but faster because it is held in memory once loaded. The `Mapping250K_Sty` `cdm` includes mismatch probes as well as perfect match probes and has two `sets`: The first comprises those 195567 probesets with six quartets, while the second comprises the 42737 probesets with ten quartets. These are reached by specifying `set 1` or `2` respectively in a call to `snpmap()`. Alternatively, both sets can be read at once using `msnpmap()`. The mismatch probes are extracted by specifying `useMM=TRUE`.

SNPMaP.cdm-package *Annotation for SNP Microarrays and Pooling in R*

Description

Pooling DNA on SNP microarrays is a cost-effective way to carry out genome-wide association studies for heritable disorders or traits. The SNPMaP package provides formal SNPMaP objects and

methods in R as a base for these analyses using Affymetrix genotyping arrays. The SNPMaP.cdm package provides cdm objects for the SNPMaP package.

Details

Package: SNPMaP.cdm
Type: Package
Version: 1.0.0
Date: 2008-06-24
License: GPL (>= 3)

Once the CEL files of the microarrays used in a SNPMaP study have been read into a 'raw' format SNPMaP object, a cdm matrix is used to identify the probesets, achieving a similar function to an Affymetrix CDF file, but faster because it is held in memory once loaded. Each cdm object may comprise several matrices, each referring to a different 'set' of probes. For example, the GenomeWideSNP_6 cdm contains two sets: one with three probes per SNP, the other with four probes per SNP. The cdm objects for older arrays may contain corresponding sets of mismatch probes. The SNPMaP package is an evolution of the scripts referred to in Meaburn et al (2006) and is described in Davis, Plomin and Schalkwyk (submitted for publication); please cite this paper if you find the package useful. Additional supporting material is available at <http://sgdp.iop.kcl.ac.uk/snpmap/>.

Array types

This distribution of the SNPMaP.cdm package includes cdm matrices for the following arrays:

[Mapping250K_Sty](#)
[Mapping250K_Nsp](#)
[GenomeWideSNP_5](#)
[GenomeWideSNP_6](#)

Future work

Future releases of the package will build on the range of chips covered, and custom cdm's will be made available on the SNPMaP Web site.

Author(s)

Oliver SP Davis and Leo C Schalkwyk
Maintainer: Oliver SP Davis <snpmap@iop.kcl.ac.uk>

References

Davis, OSP, Plomin, R, and Schalkwyk, LC. (submitted for publication) The SNPMaP package for R: A framework for genome-wide association using DNA pooling on microarrays.

Meaburn E, Butcher LM, Schalkwyk LC, and Plomin R. (2006) Genotyping pooled DNA using 100K SNP microarrays: a step towards genomewide association scans. *Nucleic Acids Research*, 34(4):e28. <http://dx.doi.org/10.1093/nar/gnj027>

See Also

The [SNPMaP](#) package.

Index

*Topic **datasets**

GenomeWideSNP_5, [2](#)

GenomeWideSNP_6, [2](#)

Mapping250K_Nsp, [3](#)

Mapping250K_Sty, [4](#)

*Topic **package**

SNPMaP.cdm-package, [5](#)

GenomeWideSNP_5, [2](#), [6](#)

GenomeWideSNP_6, [2](#), [6](#)

Mapping250K_Nsp, [3](#), [6](#)

Mapping250K_Sty, [4](#), [6](#)

msnpmap, [3-5](#)

set, [2](#)

sets, [3-5](#)

SNPMaP, [2-4](#), [7](#)

snpmap, [3-5](#)

SNPMaP.cdm-package, [5](#)