

Groups, tags, and snp.picker

Chris Wallace

2014-04-25 Fri

Contents

1	Grouping SNPs	1
2	The groups and tags classes	1
3	Automatic group building based on marginal posterior probabilities	2
4	Manual corrections	3

1 Grouping SNPs

We often (usually) cannot resolve an association to a single causal variant, and one of the aims of GUESSFM is to generate groups of SNPs, one of which is likely to be causal.

2 The groups and tags classes

The `groups` and `tags` classes are two ways to represent the same information, but use different structures which might be more convenient for one function or another.

```
> library(GUESSFM)
> showClass("groups")
```

```
Class "groups" [package "GUESSFM"]
```

```
Slots:
```

```
Name:      .Data      tags
Class:     list character
```

```
Extends:
```

```
Class "list", from data part
Class "vector", by class "list", distance 2
Class "AssayData", by class "list", distance 2
Class "vectorORfactor", by class "list", distance 3
```

```

> showClass("tags")

Class "tags" [package "GUESSFM"]

Slots:

Name:      .Data      tags
Class: character character

Extends:
Class "character", from data part
Class "vector", by class "character", distance 2
Class "data.frameRowLabels", by class "character", distance 2
Class "SuperClassMethod", by class "character", distance 2
Class "atomicVector", by class "character", distance 2
Class "index", by class "character", distance 2
Class "characterORconnection", by class "character", distance 2
Class "atomic", by class "character", distance 2
Class "characterORNULL", by class "character", distance 2
Class "EnumerationValue", by class "character", distance 2
Class "characterORMIAME", by class "character", distance 2
Class "characterOrRle", by class "character", distance 2
Class "vectorORfactor", by class "character", distance 3

```

3 Automatic group building based on marginal posterior probabilities

Once models have been fitted, we can average over them to find marginal posterior probabilities for each SNP. We make the assumption that SNPs in a group should display the following properties:

- the sum of MPI should be no more than 1
- SNPs in a group should not occur in the same model together very often
- SNPs in a group should be correlated ($r^2 > 0.5$), and the MPI should tail off from the primary SNP with r^2

`snp.picker()` aims to create such groups in an automated manner, given the summary of a GUESS run and a genotype dataset from which r^2 may be calculated. It returns an object of class `snppicker` which is an extension of the `groups` class and contains information to allow plotting of the relationship between r^2 and MPI that was used to define a group.

```

> showClass("snppicker")

Class "snppicker" [package "GUESSFM"]

Slots:

```

```
Name:  plotsdata    groups
Class:    list      list
```

4 Manual corrections

Plotting an object of class `snppicker` should be informative. The algorithm underlying `snppicker()` has been designed so it is more likely to err by splitting a single group in two than by joining what might be two separate groups. If you suspect two groups should be merged, you can check the assumptions above using `check.merge()`, and perform a merge using `group.merge()`.