vignette

```
library(hlacoloc)
```

Introduction

HLA-coloc performs colocalization in two steps, as described in the manuscript. These are:

1. HLA gene causal signature: this uses SuSiE to assign a posterior inclusion probability (PIP) to each HLA allele.
2. HLA PIP Bayesian regression: this uses stanR to perform Bayesian regression of the PIPs in each genes separately.

Required inputs

The following dataframes are necessary to run it (for each of the two phenotypes):

1. The HLA association summary statistics. This is a dataframe with four columns:

- `Name`: the name of the allele. These must be given in IMGT-HLA format, but can be of any field resolution though (as long as it's the same for both phenotypes).
- `beta`: the beta of the associations.
- `se`: the standard errors.
- `N`: the sample sizes.

2. The LD matrix (can be either in dataframe or matrix format). This is a square matrix of R coefficient for the HLA alleles above. Importantly, the order of the alleles needs to be the same in the summary statistics dataframe and in the LD matrix here.

Example

First, load the data:

```
data("ebna","ms","r_ebna","r_ms")
```

Here's a snippet of the EBNA data:

```
head(ebna)
#>         Name          beta       se    N
#> 1 G*01:01:01  0.014137500 0.015906 7247
#> 2 G*01:04:04  0.076847400 0.070867 7247
#> 3 G*01:01:02 -0.017568100 0.018972 7247
```

```
#> 4 G*01:01:03  0.000931893 0.033494 7247
#> 5   G*01:05N -0.293138000 0.078049 7247
#> 6 G*01:06:01  0.002110380 0.035066 7247
r_ebna[1:5,1:5]
#>            G*01:01:01 G*01:04:04 G*01:01:02 G*01:01:03   G*01:05N
#> G*01:01:01   1.000000 -0.1573650 -0.5404320 -0.2478420 -0.0925080
#> G*01:04:04  -0.157365  1.0000000 -0.0359779 -0.0317737 -0.0171784
#> G*01:01:02  -0.540432 -0.0359779  1.0000000 -0.1419840 -0.0672568
#> G*01:01:03  -0.247842 -0.0317737 -0.1419840  1.0000000 -0.0257557
#> G*01:05N    -0.092508 -0.0171784 -0.0672568 -0.0257557  1.0000000
```

And here's a snippet of the MS data:

```
head(ms)
#>         Name      beta       se      N
#> 1 G*01:01:01 -0.082764 0.032331 429822
#> 2 G*01:04:04  0.059263 0.144366 429822
#> 3 G*01:01:02 -0.050757 0.043057 429822
#> 4 G*01:01:03  0.120691 0.061987 429822
#> 5   G*01:05N -0.112573 0.158904 429822
#> 6 G*01:06:01  0.155703 0.059848 429822
r_ms[1:5,1:5]
#>            G*01:01:01  G*01:04:04 G*01:01:02 G*01:01:03    G*01:05N
#> G*01:01:01  1.0000000 -0.15260200 -0.5331010 -0.2476440 -0.09820330
#> G*01:04:04 -0.1526020  1.00000000 -0.0378230 -0.0167532 -0.00766048
#> G*01:01:02 -0.5331010 -0.03782300  1.0000000 -0.1369080 -0.05690370
#> G*01:01:03 -0.2476440 -0.01675320 -0.1369080  1.0000000 -0.02644930
#> G*01:05N   -0.0982033 -0.00766048 -0.0569037 -0.0264493  1.00000000
```

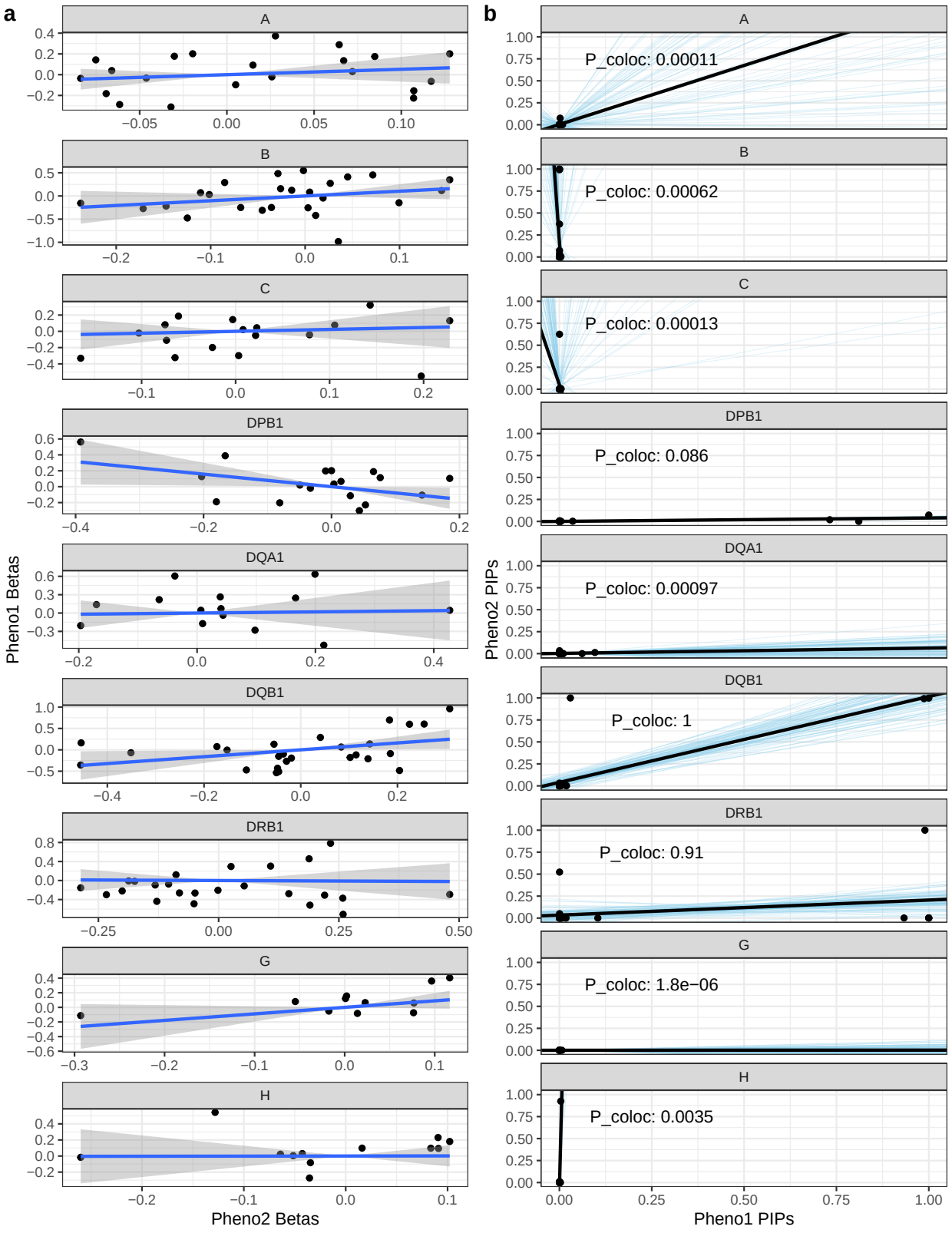
Now we run coloc:

```
coloc_res<-hla_coloc(pheno1=ebna,
                     pheno1R=r_ebna,
                     pheno2=ms,
                     pheno2R=r_ms)
#> Joining, by = "Name"
#> Joining, by = "gene"
```

Results:

```
coloc_res[["hla_colocalization"]]
#> # A tibble: 9 x 5
#>   gene  susie_coloc_prob bayes_pd direction_of_correlat~1 hla_colocalization_p~2
#>   <chr>            <dbl>    <dbl> <chr>                                   <dbl>
#> 1 DQB1          1.00          1     Correct                               1.00
#> 2 DRB1          0.990         0.918 Correct                               0.908
#> 3 DPB1          0.0856        1     Correct                               0.0856
#> 4 H             0.00349       0.994 Correct                               0.00347
#> 5 DQA1          0.00129       0.753 Correct                               0.000970
#> 6 B             0.000803      0.776 Incorrect                             0.000623
#> 7 C             0.000192      0.673 Incorrect                             0.000130
#> 8 A             0.000152      0.747 Correct                               0.000113
```

```
#> 9 G           0.00000352    0.514 Correct                                0.00000181
#> # i abbreviated names: 1: direction_of_correlation,
#> #   2: hla_colocalization_probability
coloc_res[["plot"]]
```


a
b
A
B
C
DPB1
DQA1
DQB1
DRB1
G
H
Pheno1 Betas
Pheno2 Betas
Pheno2 PIPs
Pheno1 PIPs
P_coloc: 0.00011
P_coloc: 0.00062
P_coloc: 0.00013
P_coloc: 0.086
P_coloc: 0.00097
P_coloc: 1
P_coloc: 0.91
P_coloc: 1.8e−06
P_coloc: 0.0035