HyPrColoc

Ref: HyPrColoc: Foley CN, Staley JR, et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. BioRxiv 2019. doi: https://doi.org/10.1101/592238

Three key assumptions

1. GWAS results are from the same underlying population (same LD pattern)
2. at most one causal variant in the genomic region for each trait
3. Causal variant are either directly typed or well imputed

R package: hyprcoloc 安裝

1. install.packages("devtools")
2. library(devtools)
3. install_github("cnfoley/hyprcoloc", build_opts = c("--resave-data", "--no-manual"), build_vignettes = TRUE)
4. library(hyprcoloc)
5. browseVignettes("hyprcoloc")

If issue with installation (owing to c++ compiler)

Try replacing 3 above with previous package version:

3. install_github("jrs95/hyprcoloc", build_opts = c("--no-resave-data", "--no-manual"), build_vignettes = TRUE)

Note there is no "prior.c" parameter in this version, instead use "prior.2 = 1 - prior.c". Default settings are matched.

Otherwise, on a Windows machine try updating Rtools: remove the previous version of Rtools (probably located C:\Rtools) and download Rtools40 from CRAN [https://cran.r-project.org/bin/windows/Rtools/]

我的 R 版本: 4.3.2

Example

執行以下指令

browseVignettes("hyprcoloc")

HyPrColoc

Christopher N Foley & James R Staley 2024-09-01

https://github.com/lihsinchien/MWAS-HyPrColoc/blob/main/hyprcoloc.md#6-analysing-correlated-traits

Continuous traits, ind. SNPs & traits

i. a cluster of putatively colocalized traits
ii. the posterior probability that these traits are colocalized
iii. the 'regional association' probability* (which is always > the posterior probability)
iv. a candidate causal variant explaining the shared association
v. the proportion of the posterior probability explained by this variant (which represents the HyPrColoc multi-trait fine-mapping probability).

```
> res;
$results
```

	iteration	traits	posterior_prob	regional_prob	candidate_snp	posterior_explained_by_snp	dropped_trait
1	1	T1, T2, T3, T4, T5	1.0000	1	rs11591147	1.0000	NA
2	2	T6, T7, T8	0.9164	1	rs12117612	0.4197	NA
3	3	T9, T10	0.9018	1	rs7524677	0.0763	NA

i. ii. iii. iv. v.

```
attr(,"class")
[1] "hyprcoloc"
```

ii. 越大代表，i. 中的 traits 受同一個 SNP 影響的可能性越高 (colocalization)
v. 越大，代表 iv. 中的 SNPs 為 causal 的可能性越大
Ex: T1~5 等五個 traits 為 colocalized traits可能性高(1.00), 其 causal variant 為 rs11591147可能性高(1.00)
Ex: T9, T10 為 colocalized traits 可能性高(0.90)，其 causal variant 為 rs7524677 的可能性低(0.08)

Binary traits, ind. SNPs & traits

```
binary.traits = c(1,1,1,rep(0,dim(betas)[2]-3));
res <- hyprcoloc(betas, ses, trait.names=traits, snp.id=rsid, binary.outcomes = binary.traits);
res
#> $results
#>   iteration             traits posterior_prob regional_prob candidate_snp
#> 1         1 T1, T2, T3, T4, T5         1.0000             1    rs11591147
#> 2         2         T6, T7, T8         0.9164             1    rs12117612
#> 3         3            T9, T10         0.9018             1     rs7524677
#>   posterior_explained_by_snp dropped_trait
#> 1                     1.0000           NA
#> 2                     0.4197           NA
#> 3                     0.0763           NA
#>
#> attr(,"class")
#> [1] "hyprcoloc"
```

- Binary.outcomes=c(1,1,1,0,0,0,0,0,0,0)
- T1~T3 continuous, T4~T10 binary

測試某些traits (T1, T2) 是否有 colocalization

- Trait.subset = c(“T1”,”T2”)

```
res <- hyprcoloc(betas, ses, trait.names=traits, snp.id=rsid, trait.subset = c("T1","T2"));
res
#> $results
#>   iteration traits posterior_prob regional_prob candidate_snp
#> 1         1 T1, T2              1             1    rs11591147
#>   posterior_explained_by_snp dropped_trait
#> 1                          1            NA
#>
#> attr(,"class")
#> [1] "hyprcoloc"
```

Computing a credible set of snps

```
res <- hyprcoloc(betas, ses, trait.names=traits, snp.id=rsid, snpscores = TRUE);
```

```
res[[1]];
#>   iteration             traits posterior_prob regional_prob candidate_snp
#> 1         1 T1, T2, T3, T4, T5         1.0000             1    rs11591147
#> 2         2         T6, T7, T8         0.9164             1    rs12117612
#> 3         3            T9, T10         0.9018             1     rs7524677
#>   posterior_explained_by_snp dropped_trait
#> 1                     1.0000           NA
#> 2                     0.4197           NA
#> 3                     0.0763           NA
```

```
head(res[[2]][[1]]);
#>    rs6694014   rs11206477     rs978479    rs6684892  rs149881092    rs2081705
#> 4.109458e-67 3.004278e-68 5.637121e-67 6.486572e-68 6.063410e-66 5.765875e-67
```

fine-mapping score(snpscore): the probability that a snp is likely to be causal

identifies a credible set of snps which explains 95% of the posterior probability of colocalization for each cluster of colocalized traits

```
cred.sets(res, value = 0.95);
#> [[1]]
#> rs11591147
#>         1
#>
#> [[2]]
#>  rs12117612   rs7532349  rs11206481  rs12145624  rs12724445  rs12126037
#> 0.419677651 0.419677651 0.025324539 0.022969148 0.022969148 0.022332481
#>  rs13375783   rs1544909
#> 0.014513746 0.009991906
#>
#> [[3]]
#>  rs7524677  rs7547776  rs7524783  rs6701789  rs1035817  rs7530321  rs7520033
#> 0.07626197 0.07626197 0.07626197 0.07626197 0.07626197 0.05349125 0.05349125
#>  rs7553410 rs11206486  rs7524899  rs6681189  rs7537797 rs11206485 rs10888890
#> 0.05349125 0.02231383 0.02231383 0.02231383 0.02231383 0.02098626 0.02098626
#>  rs2114574  rs6664660  rs6588539 rs11206489  rs6668051 rs10218512 rs11206492
#> 0.02098626 0.02098626 0.02098626 0.02098626 0.02098626 0.02098626 0.02098626
#>  rs4634950  rs7539163  rs7529244  rs6687414  rs7538808 rs11206487 rs12745865
#> 0.02098626 0.02098626 0.02098626 0.02098626 0.02098626 0.01203687 0.01203687
#> rs12725873
#> 0.01203687
```

```
ld.matrix["rs12117612","rs7532349"];
#> [1] 1
```

3 Assessing stability of clusters (sensitivity analysis)

3.1 Change prior

3.2 The Bayesian divisive clustering algorithm

6 Analysing correlated traits

(i) a matrix containing the putative correlation between the traits;

(ii) a matrix containing the LD between the snps in the region

(iii) a matrix containing the proportion of shared participants between each pair of studies.

- Note the matrix of overlapping sample proportions has diagonal elements set to 1 and off diagonal elements (i,j) denote the proportion of overlap between the i'th and j'th studies
  - ex:if N denotes the total number of overlapping participants and Ni,Nj denote the numbers of participants in the i'th and j'th studies respectively, then N/(NiNj) is the (i,j)=(j,i) element.

```
# default assumption, presented for clarity:
sample.overlap = matrix(1, dim(betas)[2], dim(betas)[2]);
```

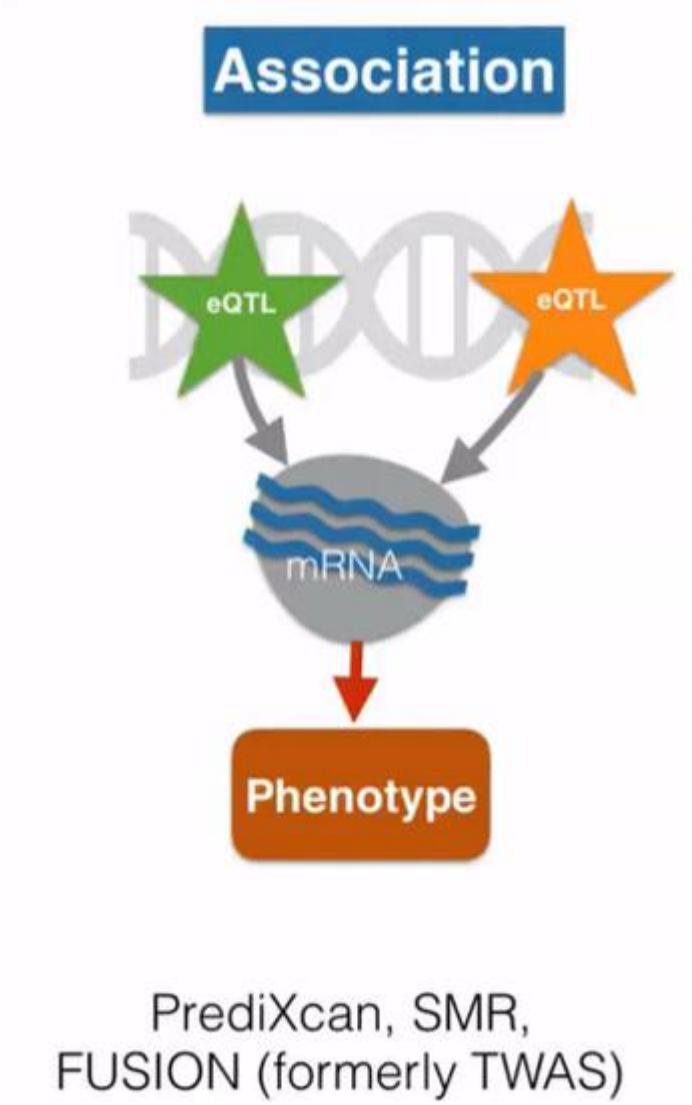
```
traits <- paste0("T", 1:dim(betas)[2]);
ptm = proc.time();
res <- hyprcoloc(betas, ses, trait.names=traits, snp.id=rsid, trait.cor = trait.cor,
                 ld.matrix = ld.matrix, sample.overlap = sample.overlap, uniform.priors = FALSE);
time.corr = proc.time() - ptm;
res
#> $results
#>   iteration             traits posterior_prob regional_prob candidate_snp
#> 1         1 T1, T2, T3, T4, T5         1.0000             1    rs11591147
#> 2         2         T6, T7, T8         0.8257             1    rs12117612
#> 3         3            T9, T10         0.8944             1     rs7524677
#>   posterior_explained_by_snp dropped_trait
#> 1                     1.0000           NA
#> 2                     0.3535           NA
#> 3                     0.0696           NA
#>
#> attr(,"class")
#> [1] "hyprcoloc"
```

```
res <- hyprcoloc(betas, ses, trait.names=traits, snp.id=rsid, trait.cor = trait.cor,
                 ld.matrix = ld.matrix, sample.overlap = sample.overlap, uniform.priors = TRUE);
res
#> $results
#>   iteration             traits posterior_prob regional_prob candidate_snp
#> 1         1 T1, T2, T3, T4, T5         1.0000             1    rs11591147
#> 2         2         T6, T7, T8         0.9744             1    rs12117612
#> 3         3            T9, T10         0.9782             1     rs7524677
#>   posterior_explained_by_snp dropped_trait
#> 1                     1.0000           NA
#> 2                     0.3535           NA
#> 3                     0.0696           NA
#>
#> attr(,"class")
#> [1] "hyprcoloc"
```

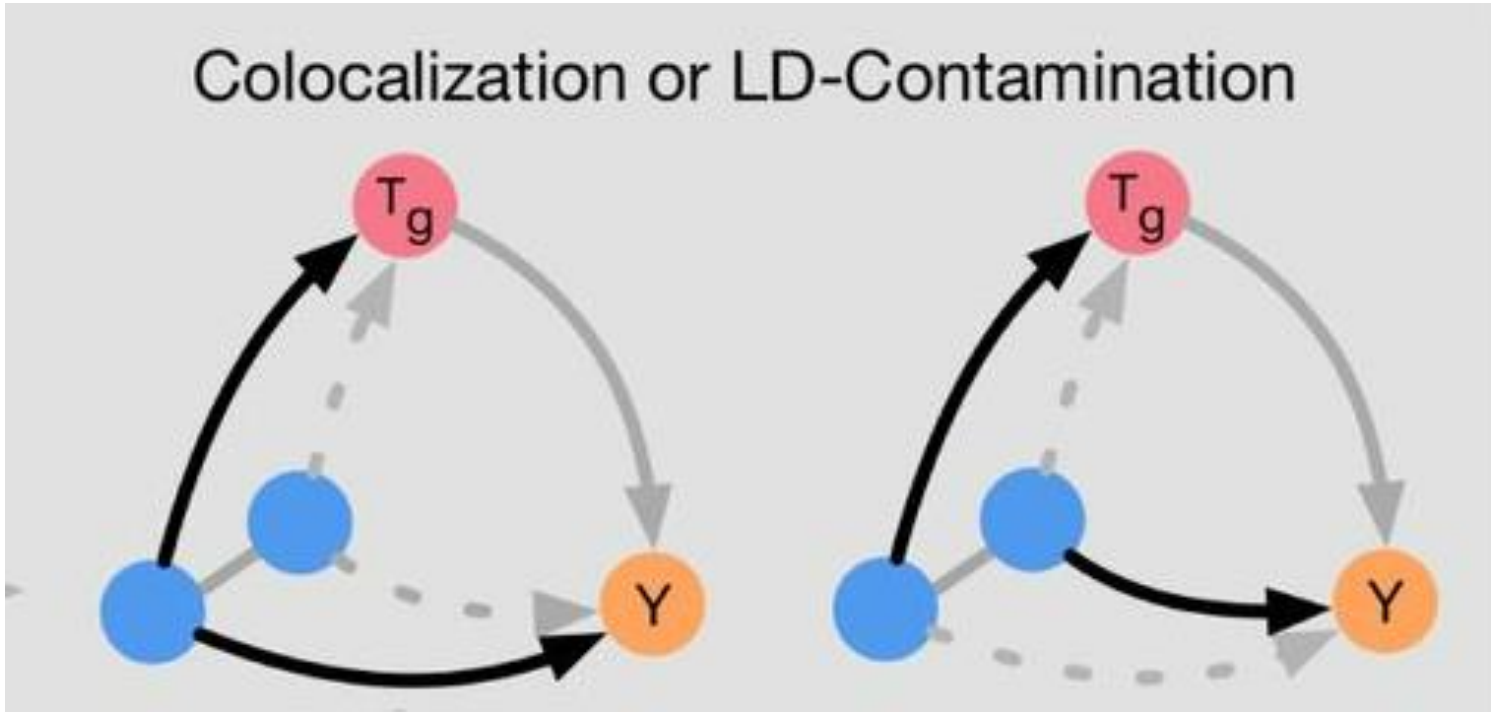
```
res <- hyprcoloc(betas, ses, trait.names=traits, snp.id=rsid);
res;
#> $results
#>   iteration             traits posterior_prob regional_prob candidate_snp
#> 1         1 T1, T2, T3, T4, T5         1.0000             1    rs11591147
#> 2         2         T6, T7, T8         0.9164             1    rs12117612
#> 3         3            T9, T10         0.9018             1     rs7524677
#>   posterior_explained_by_snp dropped_trait
#> 1                     1.0000           NA
#> 2                     0.4197           NA
#> 3                     0.0763           NA
#>
#> attr(,"class")
#> [1] "hyprcoloc"
```

```
ptm = proc.time();
res <- hyprcoloc(betas, ses, trait.names=traits, snp.id=rsid);
time.ind = proc.time() - ptm;
time.ind;
#> 使用者    系統    流逝
#>   0.00    0.01    0.01
time.corr;
#> 使用者    系統    流逝
#>  33.02    0.37   33.40
```

Source: https://www.youtube.com/watch?v=FBflh7jR4X4&list=WL&index=28&t=1904s



PrediXcan, SMR,
FUSION (formerly TWAS)



https://www.biorxiv.org/content/10.1101/045260v7.full

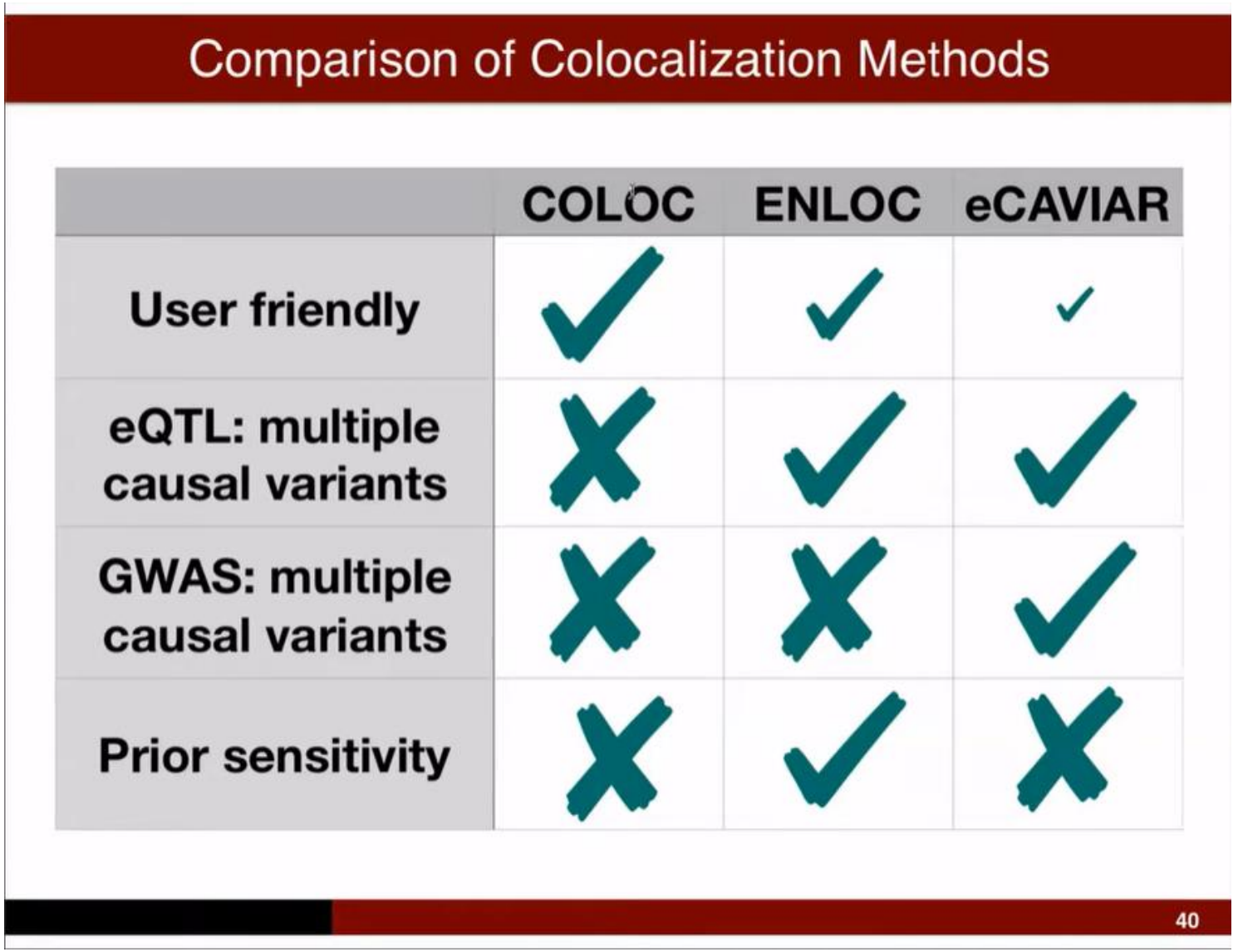
Lung cancer, methylation, gene expression

M 個 SNPs
3 個 trait: T1, T2, T3

資料來源	所需資料 1
Disease (T1) ~ SNPs	1. $\hat{\beta}_{T1}$ 2. $se(\hat{\beta}_{T1})$
Expression (T2)~ SNPs	1. $\hat{\beta}_{T2}$ 2. $se(\hat{\beta}_{T2})$
Methylation (T3)~ SNPs	1. $\hat{\beta}_{T3}$ 2. $se(\hat{\beta}_{T3})$

所需資料 2 (預設為單位矩陣)

1. LD matrix (MxM 矩陣)
2. Trait correlation matrix (3x3 矩陣)
3. Proportion of shared participants (3x3 矩陣)



https://www.youtube.com/watch?v=FBflh7jR4X4&list=WL&index=28&t=1904s

- The posterior odds:

$$\frac{P(H_m|D)}{P(H_0|D)} = \sum_{S \in S_m} \boxed{\frac{P(D|S)}{P(D|S_0)}} \times \boxed{\frac{p(S)}{p(S_0)}}$$

Bayes factor Prior odds

$$S^* = \max_{S \in S_m} P(S|D)$$

$$P(H|D) = \frac{\sum_{S \in S_H} BF(S) \frac{p(S)}{p(S_0)}}{\sum_{H_i \in \Omega} \sum_{S \in S_{H_i}} BF(S) \frac{p(S)}{p(S_0)}}, \quad (5)$$

BF(S): Bayes factor which is the likelihood of the data being generated under $S \in S_H$ relative to the likelihood of the data being generated S_0

prior

1. variant-level priors

A genetic variant is associated with k traits

$$p_0 + \sum_{k=1}^{m}\left(\sum_{j_1=1}^{m}\sum_{j_2>j_1}\cdots\sum_{j_k>j_{k-1}} p_{j_1 j_2 \ldots j_k}\right) = 1.$$

$$p_{12\ldots k} = p\prod_{i=2}^{k}\left(1-\gamma^{i-1}\right), k = 2, \ldots m, \quad (10)$$

p: the probability of the genetic variant being associated with one trait
γ: controls the probability that a genetic variant is associated with an additional trait
$p_c = 1 - \gamma$: the probability of a variant being causal for a second trait given it is causal for one trait (the conditional colocalization prior), $p_c = p_{12}/p_1$

$$\frac{p(S)}{p(S_0)} = \frac{p_{12\ldots k}}{p_0} = \frac{p}{p_0}\prod_{i=2}^{k}\left(1-\gamma^{i-1}\right), k = 2, \ldots, m,$$

for configurations $S \in S_{\mathcal{H}_k}$, where k traits share a causal variant and the remaining $m - k$ traits do not have a casual variant, and

prior

2. Conditionally uniform prior probabilities

(i) not setting variant-level information

(ii) implicitly accounting for large differences in the causal configuration space between hypotheses

$$\frac{P(S|H)}{P(S_0|H_0)} = \frac{1/|\mathcal{S}_H|}{1/|S_0|} = 1/|\mathcal{S}_H|, \tag{13}$$

where $\left|\mathcal{S}_{\mathcal{H}_k}\right| = Q$ and

$$\left|\mathcal{S}_{\mathcal{H}_{(m-1,1)}}\right| = \begin{cases} Q(Q-1) : m = 2, \\ mQ(Q-1) : m > 2. \end{cases} \tag{14}$$