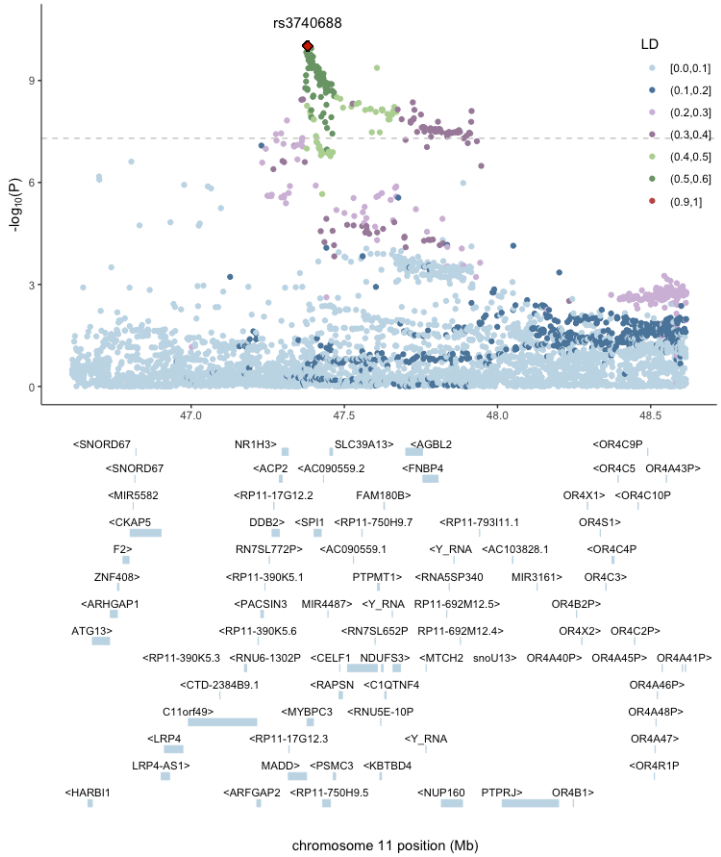
clocus R pipeline

Jiawen Chen

August 2019

1 Introduction

In order to make GWAS plot points more distinguishable, separating the colors by r^2 using 10 levels is a good choice. Here is an example about gene CELF1 region on chromosome 11. clous_upper to make the upper part and clous_lower to make the lower part.



2 Input dataset

For making these plot, GWAS summary stats(including P values and BP), LD, and Gene list files are required. Including only 2Mb downstream and upstream of the gene you want to plot is a good practice.

Calculating LD

First select the most significant SNP from your gene region and use plink to calculate the LD. You can use the script follows.

```
#!/bin/bash
/z/Comp/lu_group/Software/plink/plink_1.9_linux_x86_64/plink \
--bfile <yout bfile here> \
--keep <your sample ID file here> \
--extract <your GENE_SNP.txt here> \
--r2 \
--ld-snp <your top SNP ID here> \
--ld-window-r2 0 \
--ld-window-kb 2000 \
--ld-window 99999  \ #keep LD of all other SNPs
--out <your output file name here>
```

And sample out put will be like (Here rs3740688 is my top SNP)

CHR_A	BP_A	SNP_A	CHR_B	BP_B	SNP_B	R2
11	47380340	rs3740688	11	46611239	rs61882757	0.00747222
11	47380340	rs3740688	11	46611805	rs61882758	0.00749763
11	47380340	rs3740688	11	46612332	rs73449983	0.00747222
11	47380340	rs3740688	11	46612511	rs150757221	0.00201535
11	47380340	rs3740688	11	46612524	rs55737385	0.0359818
11	47380340	rs3740688	11	46614726	rs145698246	0.00136195
11	47380340	rs3740688	11	46615682	rs74553660	0.0155928
11	47380340	rs3740688	11	46616036	rs185540637	0.00271718

FYI, 1KG genotype files are in

/z/Comp/lu_group/Resource/1000G.

List of EUR samples path is

/z/Comp/lu_group/Members/jiawen/clocus/1000GP_Phase3.EUR.sample

Gene list

Thanks to Kunling. The formatted file is in

/z/Comp/lu_group/Members/jiawen/clocus/gencode.v19.withdir.txt

Then extract the genes within corresponding chromosome.

Gene GWAS

Combine the LD and GWAS(2Mb) together and keeps the P, position, snp id, r^2 columns. And change the column name as follows.

```
          P          r2         snp      pos
3.243e-01 7.47222e-03  rs61882757 46611239
4.358e-01 7.49763e-03  rs61882758 46611805
3.512e-01 7.47222e-03  rs73449983 46612332
1.642e-01 2.01535e-03 rs150757221 46612511
```

Also extract the top SNP P value and position.

3 clocus methods

clocus_upper

clocus_upper is for plotting the upper part figure with the input:

```
clocus_upper(gene_GWAS, color, grid.show, top_pos, top_p, top_snp)
```

gene_GWAS: the file mentioned in the Gene GWAS part, the column name has to be the same
color: default is

"c1d8e6","537ea3","d0b8d9","a283a3","739e6f"

"b3d49b","ffc887","e3987f","ffb5b5","bf4d4d"

grid.show: show grid or not, default is not, you can show by using

```
grid.show = element\_line(color = "grey94")
```

top_pos: position of top SNP you want to highlight
top_p: P value of top snp
top_snp: name of top snp

clocus_lower

clocus_lower is for plotting the lower part with the input:

```
clocus_lower(gene_GWAS, gene_chr , gene_everyline , x_title , grid.show ,
             omit_gene)
```

gene_GWAS: the file mentioned in the Gene GWAS part, the column name has to be the same
gene_chr: chromosome specific gene list file which can be extracted by the file described in Gene List. And the column name has to be same as follows. (Note: you do not have to add "<" or ">" by your self, the figure here just to claim the column names)

chr	gene	gene_id	start	end	direction	gene_name
11	gene	ENSG00000180423.4	46624411	46638777	-	<HARBI1
11	gene	ENSG00000175224.12	46638826	46696368	+	ATG13>
11	gene	ENSG00000175220.7	46698630	46722149	-	<ARHGAP1
11	gene	ENSG00000175213.2	46722368	46727462	+	ZNF408>
11	gene	ENSG00000180210.10	46740730	46761056	+	F2>
11	gene	ENSG00000175216.10	46764598	46867843	-	<CKAP5

gene_everyline: The maximum number of genes in every line, the default number is 6.
x_title: The name of xlab.
grid.show: As described in the upper part.
omit_gene: The genes you do not want to show in the figure. Default is a empty list.

4 R package required

```
library(gggenes)
library(ggplot2)
library(ggfittext)
library(gplots)
library(data.table)
library(ggrepel)
library(latex2exp)
```

Check the website of gggenes to get more information about that, interesting package!

https://CRAN.R-project.org/package=gggenes

And you can use the files in my folder to have a try. The R code is also there.

/z/Comp/lu_group/Members/jiawen/clocus