

Readme

The name of the package will imply the value of taskid. For example, `taskid_200_res.zip` implies that `taskid = 200`. Name of `taskid_xxx_res.zip` will give `taskid = xxx`.



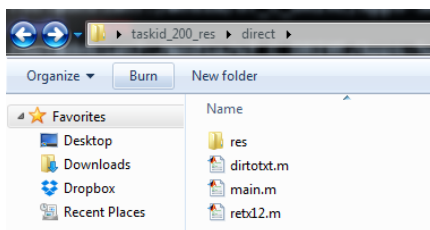
Inside the `taskid_xxx_res` there would be two folders: `direct` and `indirect`



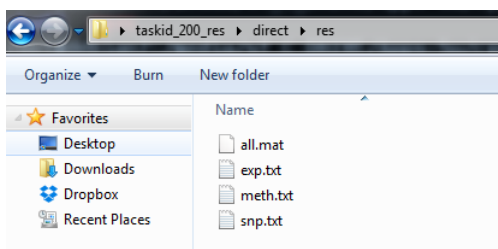
Here folder `direct` is to store all result about direct relationship, so for `indirect`. Maybe there will be only direct with indirect missing, which means I can't retrieve indirect relationship from current taskid.

direct relationship part

In folder `taskid_xxx_res/direct`, there will be some `.m` files and a folder named `res`. Ignore all `.m` files since they are only used to generate folder `res`.

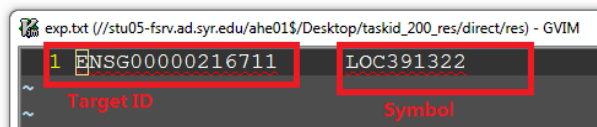


In `taskid_xxx_res/direct/res`, There would be three `.txt` files and one `.mat` file. just focus on these `.txt` files since it's readable.



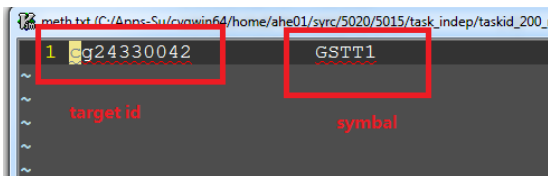
1. exp.txt

exp.txt is to record the **Target ID** and **Symbol** of **Expression**. There is only one line for exp.txt since for each task we only focus on one Expression variable.



2. meth.txt

This file is to record all methylations that are selected when doing regression to expression. There may be more than one lines inside (depending on how many methylation are selected), but if there are no methylation selected, this file won't be generated.



For this there is only one line in meth.txt since there is only one selected methylation.

3. snp.txt

This file is to store all SNP's selected when doing regression to expression. There are three columns in the file.

```

14 snp.txt (C:/Apps-Su/cygwin64/home/ahel01/sync/5020/5015/task_indep/taskid_200_res/d
1 1 rs5992911
2 1 rs73877834
3 1 rs5992912
4 1 rs34653617
5 1 rs11089214
6 1 rs11089215
7 1 rs11089216
8 1 rs11089218
9 1 rs10775697
10 1 rs5746554
11 1 rs62228662
12 1 rs5746564
13 1 rs5747680
14 1 rs5746567
15 1 rs5747698
16 1 rs113710173
17 1 rs9618684
18 1 rs12628685
19 1 rs12170243
20 1 rs5748929
21 1 rs2108582
22 1 rs7285493
23 1 rs9617962
24 1 rs9617963
25 1 rs12157365
26 1 rs5992649
27 1 rs5994238
28 1 rs5992650
29 2 rs178273
30 2 rs680548
31 2 rs650538
32 2 rs12483894
33 2 rs111719983
34 2 rs115091689
35 2 rs113994637
36 3 rs8139870
37 4 rs16981003
38 5 rs4414
39 5 rs114959216
40 5 rs113402025
41 5 rs112343331
42 5 rs75151479
43 5 rs80005710
44 5 rs3788394
45 5 rs5752309
46 5 rs6005025
47 5 rs5997077
48 5 rs111633165
49 5 rs112794880

```

The first line is just to index all involved SNP's (including tagSNP and SNP tagged by tagSNP).

The second column is used to express the tagSNP and "SNP tagged by tagSNP". As we can see in the example, there are some repeating for the second line (1, 2, 5 are not unique, there are repeating, but 3, 4 are unique without repeating.)

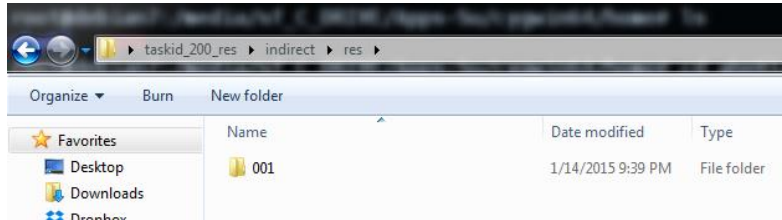
for all SNP's with the same value of the second line, the first one appeared in the file is the tagSNP and the rest is the "SNP's tagged by the tagSNP".

i.e. in the picture there are 28 SNP's are of value 1 for the second line. Then the first one is tagSNP(rs5992911), and the rest of the 27 SNP's are just SNP's tagged by the tagSNP.

but if they are unique of the second line, which just means the selected tagSNP doesn't tag any other SNP in the dataset(i.e. rs8139870 of line 36, means rs8139870 is a tagSNP and it doesn't tag any other SNP. And the same for rs1698003 of line 37)

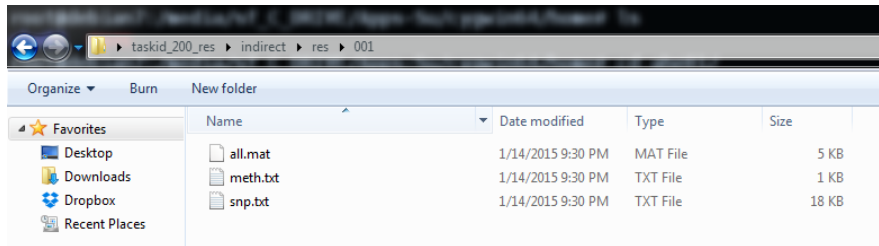
Indirect relationship part

in folder `taskid_xxx_res/indirect/res` , Again let ignore all .m files. there would be more than one folders inside. it's named as 001 002 003 004



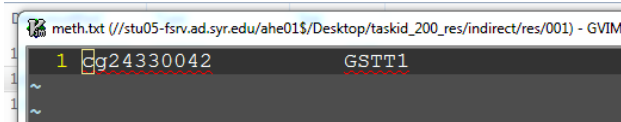
for taskid = 200 since there is only one methylation selected, then there is only 001.

If I can't retrieve indirect relationship(means no methylation selected), then the `taskid_xxx_res/indirect/` folder won't be generated. In folder `taskid_xxx_res/indirect/res/xxx` there will be three files all.mat, meth.txt, snp.txt, let's focus on .txt file only.



1. meth.txt

to record the methylation for this indirect relationship task. there is only one line inside. The first column is for target id and the second is for symbol.



2. snp.txt

to record all SNP's when doing regression from SNP to methylation. I skip it here since it's exactly the same as I explained in the direct relationship part.