How to analyze interSNP results of DB_TEST_BestMarkerCombi2.txt

The main purpose here is the same that at SingleMarker but a little more tricky.

The interSNP file contents are,

```
No
Chr_1
rs_No_1
Pos_No_1
p-Single-marker_1
Chr_2
rs_No_2
Pos_No_2
p-Single-marker_2
p-value
p-value_corr
SNP1_Allele_A
SNP1_Allele_B
SNP2_Allele_A
SNP2_Allele_B
beta_x1
se_x1
beta_x1D
se_x1D
beta_x2
se_x2
beta_x2D
se_x2D
beta_x1x2
se_x1x2
beta_x1x2D
se_x1x2D
beta_x1Dx2
se_x1Dx2
beta_x1Dx2D
se_x1Dx2D
```

Lot of helpless info here. But don't despair, we are going to make it quickly now.

First, parse it again with *awk* for sanity. Now we will try just CHR SNP BP and P, let see if plink can do the job

```
echo "CHR SNP BP P" > tmpfile; tail -n +2  myresults.file | awk -F'\t'
'{print $2,$3,$4,$5"\n"$6,$7,$8,$9}' >> tmpfile
```

I tried this,

```
echo "CHR SNP BP P" > test_clean_combi.txt; tail -n +2
/home/data/Bonn_GWAS/TEST5/intersnp_TEST5/outputs/ADMURimpQC2_TEST5_BestMark
```

```
erCombi2.txt | awk -F'\t' '{print $2,$3,$4,$5"\n"$6,$7,$8,$9}' >>
test_clean_combi.txt
```

and then

```
plink --bfile ~/data/Variomics/ADMURimpQC2 --clump test_clean_combi.txt --
clump-p1 0.01
```

with bad luck,

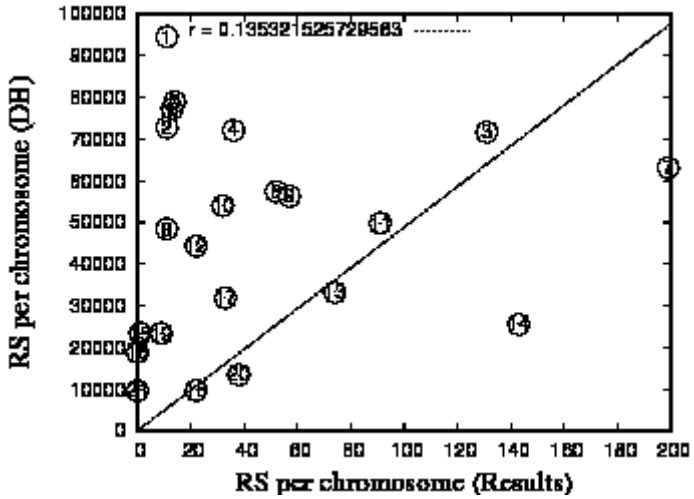
```
$ wc -l ADMURimpQC2_TEST5_BestMarkerCombi2.clumped
236 ADMURimpQC2_TEST5_BestMarkerCombi2.clumped
```

So what now? Let's try with P=0.1. This is not wrong at all. The point here is that each SNP has a low effect over the disease but a combination of them could has some noticeable effect.

```
plink --bfile ~/data/Variomics/ADMURimpQC2 --clump test_clean_combi.txt --
clump-p1 0.1
```

```
$ wc -l ADMURimpQC2_TEST5_BestMarkerCombi2.clumped
1979 ADMURimpQC2_TEST5_BestMarkerCombi2.clumped
```

So now run the test and,



But of course, it is too good to be true. A closer look at the *awk* command shows we are using the *p values* for single marker of each SNP. This is at least curious and I think it must be taken into account in some way.

But, let's see what happens if the combined *p value* is used.

```
$ echo "CHR SNP BP P" > test_clean_combi2.txt; tail -n +2
/home/data/Bonn_GWAS/TEST5/intersnp_TEST5/outputs/ADMURimpQC2_TEST5_BestMark
erCombi2.txt | awk -F'\t' '{print $2,$3,$4,$10"\n"$6,$7,$8,$10}' >>
test_clean_combi2.txt
$ plink --bfile ~/data/Variomics/ADMURimpQC2 --clump test_clean_combi2.txt -
-clump-p1 0.1
```

So far so good,

```
$ plink --bfile ~/data/Variomics/ADMURimpQC2 --clump test_clean_combi2.txt --clump-p1 0.1
```

give us,

```
@----------------------------------------------------------@
|        PLINK!       |     v1.07      |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@

Web-based version check ( --noweb to skip )
Connecting to web...  OK, v1.07 is current

Writing this text to log file [ plink.log ]
Analysis started: Thu Mar  7 09:22:41 2013

Options in effect:
        --bfile /home/osotolongo/data/Variomics/ADMURimpQC2
        --clump test_clean_combi2.txt
        --clump-p1 0.1

Reading map (extended format) from [
/home/osotolongo/data/Variomics/ADMURimpQC2.bim ]
1034238 markers to be included from [
/home/osotolongo/data/Variomics/ADMURimpQC2.bim ]
Reading pedigree information from [
/home/osotolongo/data/Variomics/ADMURimpQC2.fam ]
1088 individuals read from [ /home/osotolongo/data/Variomics/ADMURimpQC2.fam
]
1088 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
319 cases, 769 controls and 0 missing
511 males, 577 females, and 0 of unspecified sex
Reading genotype bitfile from [
/home/osotolongo/data/Variomics/ADMURimpQC2.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 1034238 SNPs
1088 founders and 0 non-founders found
1130 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ plink.hh ]
Total genotyping rate in remaining individuals is 0.986847
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 1034238 SNPs
```

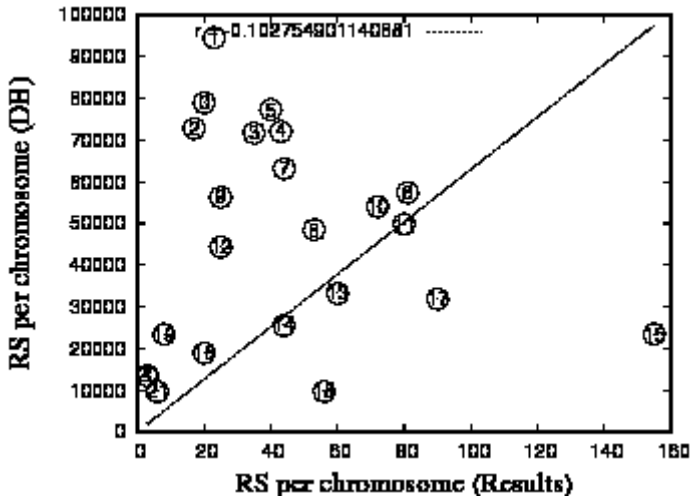
```
After filtering, 319 cases, 769 controls and 0 missing
After filtering, 511 males, 577 females, and 0 of unspecified sex

Parameters for --clump:
          p-value threshold for index SNPs = 0.1
      Physical (kb) threshold for clumping = 250
     LD (r-squared) threshold for clumping = 0.5
        p-value threshold for clumped SNPs = 0.01

Reading results for clumping from [ test_clean_combi2.txt ]
Extracting fields SNP and P
Indexing on all files
Writing clumped results file to [ plink.clumped ]

Analysis finished: Thu Mar  7 09:27:40 2013
```

and surprising, the results are also very good,



But quite different. **Why?**