Procedure 1

- First we took the every provided file (listTEST5.txt, listTEST6.txt, listTEST15.txt, listTEST16.txt) and look for the proxies for every marker at http://www.broadinstitute.org/mpg/snap/ldsearch.php with,

```
SNP data set: 1000 Genomes Pilot 1
r2 threshold: 0.7
Population panel: CEU
Distance limit: 100
```

- We substitute every marker for all its proxies and construct every possible pair for each file
- Then we choose the common pairs to all files and construct a single one (**p4.txt**)
- This file must be converted in *plink* format (just put a column after the other)
- We run *plink* over each database looking for epistasis,

```
plink --bfile bed_DB --epistasis --epi1 1 --epi2 1 --set-test --set
sets.txt --allow-no-sex --out bed_DB_sets
```

- Now we extract from the results just the pairs corresponding to **p4.txt** for each DB.
- Finally, following Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach, M. C. WHITLOCK, doi: 10.1111/j.1420-9101.2005.00917.x, we calculate the combined *p-value*,

$$ p = \phi \left(\frac{1}{\sqrt{k}} \sum_{i=1}^{k} \phi^{-1}(p_i) \right) $$ where $ \phi $ is the Normal distribution cumulative function.