

## BonnGWAS

Organizando SNPs de T. Becker

1.- Poniendolos en una linea

```
for x in list*.txt; do cat $x | awk {'print $1'} >> ${x%.txt}.cont; cat $x |
awk {'print $2'} >> ${x%.txt}.cont; sort ${x%.txt}.cont > ${x%.txt}.sc; rm
${x%.txt}.cont; done
```

😬 Que estupidez!!!!!! Asi mejor:

```
$ for x in list*.txt; do cat $x | awk {'print $1"\n"$2'} | sort | uniq >
${x%.txt}.sc; done
```

2.- Buscando proxys

<http://www.broadinstitute.org/mpg/snap/ldsearch.php>

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

→ .proxies

3.- Proxies reales

```
$ find_pairs.pl x.txt ->construye las parejas de proxies tomando las
parejas originales de x.txt y los proxies de cada marcador de x.proxies.
Escribe el resultado en x.pairs
```

4.- Despues hemos hecho esto:

```
$ cat *.pairs | sort | uniq -cd > pairs.txt
$ sort -r pairs.txt > p2.txt
$ cat p3.txt | awk '{print $2" "$3}' > p4.txt
```

Resumiendo:

```
$ cat *.pairs | sort | uniq -cd | sort -r | awk '{print $2" "$3}| > p4.txt
```

y para los sets:

```
$ awk '{print $1}' p4.txt > set1.txt
$ awk '{print $2}' p4.txt > set2.txt
$ sort set1.txt | uniq > set1a.txt
$ sort set2.txt | uniq > set2a.txt
$ cat set1a.txt set2a.txt > sets.txt
```

Resumiendo:

```
$ awk '{print $1}' p4.txt | sort | uniq > set1.txt
$ awk '{print $2}' p4.txt | sort | uniq > set2.txt
$ cat set1.txt set2.txt > sets.txt
```

O incluso en una linea, (no estoy seguro de que funcione)

```
$ cat `awk '{print $1}' p4.txt | sort | uniq` `awk '{print $2}' p4.txt |
sort | uniq` > sets.txt
```

y entonces para cada DB:

```
$ nohup plink --bfile /home/data/Variomics/TGEN_impQC2 --epistasis --epil 1
--epi2 1 --set-test --set sets.txt --allow-no-sex --out
/home/aruiz/Downloads/p4/TGEN_impQC2_sets &
```

Nota: Antes hay que editar sets.txt,

```
$ head ../sets.txt
SET_1 <----- Añadir esto!!!!
rs10000964
rs10002814
rs10002954
rs10003224
rs10005124
rs10005324
rs10007504
rs10008064
rs10008274
```

5.- Luego que ya tenemos los resultados de plink, usar [/opt/gntics/grep4dbs.sh](#) o [/opt/gngtics/gpcc.pl](#)

```
$ grep4dbs.sh ../p4.txt ADMURimpQC2_sets.epi.cc GenADA_impQC2_sets.epi.cc
TGEN_impQC2_sets.epi.cc
ADNIimpQC2_sets.epi.cc NIA_AD_impQC2_sets.epi.cc

$ awk {'print $1" "$3" "$2" "$4" "$5" "$6" "$7'} db.epi.cc >
db.epi.cc_tmp.reorder
$ gpcc.pl ../p4.txt db.epi.cc_tmp.reorder
$ mv db.epi.cc_tmp.reorder.cool db.epi.cc.ok
```

6.- Meta-analisis!!!!!!!

Segun, [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, el meta se hace con [este codigo](#) (corriendo [/opt/gntics/zscore.pl](#)) .

La forma mas simple sigue la expresión,

$$Z = \frac{1}{\sqrt{k}} \sum_{i=1}^k Z_i$$

siendo  $Z(p) = \Phi^{-1}(p)$  y  $\Phi$  la CDF de la normal. Luego,

$$p = \Phi\left(\frac{1}{\sqrt{k}} \sum_{i=1}^k \Phi^{-1}(p_i)\right)$$

Para la CDF e ICDF se usa <http://search.cpan.org/~callahan/Math-CDF-0.1/CDF.pm>

From:

<http://imagen.fundacioace.com/wiki/> - **Detritus Wiki**

Permanent link:

<http://imagen.fundacioace.com/wiki/doku.php?id=genetica:bonngwas>

Last update: **2020/08/04 10:58**

