



Imputando con 1000 genome e impute2

La base de datos a usar ya ha pasado el QC así que me voy a saltar esa parte.

Reorientado SNPs

Primeramente me bajo de aquí:

http://www.ncbi.nlm.nih.gov/projects/SNP/snp_viewBatch.cgi?sbid=33767 el Assay correspondiente al chip utilizado.

El archivo es algo así (mas un encabezado que le he quitado):

ss#	loc_snp_id	allele	samplesize	rs#	ss2rs_orien	chr
chr_pos	contig_acc		contig_pos	rs2genome_orien		
assembly	weight					
ss66315736	SNP_A-1883517	C/T	270	rs41498448	0	1
779743	NT_004350.19	258375	1	GRCh37.p10	3	
ss66317030	SNP_A-1886933	A/G	270	rs2980300	1	1
785989	NT_004350.19	264621	0	GRCh37.p10	1	
ss66323101	SNP_A-1902458	A/C	270	rs10907175	0	1
1130727	NT_004350.19	609359	0	GRCh37.p10	1	
ss66416875	SNP_A-2131660	C/T	270	rs2887286	1	1
1156131	NT_004350.19	634763	1	GRCh37.p10	1	
ss66516979	SNP_A-4221087	G/T	270	rs307378	0	1
1268847	NT_004350.19	747479	0	GRCh37.p10	1	
ss66316151	SNP_A-1884606	A/G	270	rs7540231	0	1
1506035	NT_004350.19	984667	0	GRCh37.p10	1	
ss66458131	SNP_A-2235839	A/G	270	rs2272908	1	1
1721479	NT_004350.19	1200111	0	GRCh37.p10	1	
ss66451308	SNP_A-2218153	A/G	270	rs12141314	0	1
1753686	NT_004350.19	1232318	0	GRCh37.p10	1	

Primero selecciono los SNPs que están en reverse (1 en la sexta columna)

```
$ awk '{if($6==1) print $5}' snpBatch_AFFY_33767 > list_snp_rw_oq1.txt
```

y los invierto con *plink*

```
$ plink --noweb --bfile admurcia --flip list_snp_rw_oq1.txt --make-bed --out admurciav36f
```

```
@-----@
|          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/          |
|-----|
@-----@
```

```
Skipping web check... [ --noweb ]
Writing this text to log file [ admurciav36f.log ]
Analysis started: Thu Jan  2 13:41:21 2014
```

```
Options in effect:
  --noweb
  --bfile admurcia
  --flip list_snp_rw_oq1.txt
  --make-bed
  --out admurciav36f
```

```
Reading map (extended format) from [ admurcia.bim ]
198429 markers to be included from [ admurcia.bim ]
Reading pedigree information from [ admurcia.fam ]
1088 individuals read from [ admurcia.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 [ admurcia.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Reading SNPs to flip strand from [ list_snp_rw_oq1.txt ]
Flipped strand of 299983 SNPs
Before frequency and genotyping pruning, there are 198429 SNPs
1088 founders and 0 non-founders found
1130 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ admurciav36f.hh ]
Total genotyping rate in remaining individuals is 0.993951
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 198429 SNPs
After filtering, 319 cases, 769 controls and 0 missing
After filtering, 511 males, 577 females, and 0 of unspecified sex
Writing pedigree information to [ admurciav36f.fam ]
Writing map (extended format) information to [ admurciav36f.bim ]
Writing genotype bitfile to [ admurciav36f.bed ]
Using (default) SNP-major mode
```

Analysis finished: Thu Jan 2 13:41:31 2014

Ahora voy a convertirlo en *.ped* y *.map*

```
$ plink --noweb --bfile admurciav36f --recode --tab --out admurciav36f
```

```
@-----@
|          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/          |
|-----|
@-----@
```

```
Skipping web check... [ --noweb ]
Writing this text to log file [ admurciav36f.log ]
Analysis started: Thu Jan 2 13:50:09 2014
```

Options in effect:

```
--noweb
--bfile admurciav36f
--recode
--tab
--out admurciav36f
```

```
Reading map (extended format) from [ admurciav36f.bim ]
198429 markers to be included from [ admurciav36f.bim ]
Reading pedigree information from [ admurciav36f.fam ]
1088 individuals read from [ admurciav36f.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 [ admurciav36f.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 198429 SNPs
1088 founders and 0 non-founders found
1130 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ admurciav36f.hh ]
Total genotyping rate in remaining individuals is 0.993951
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 198429 SNPs
After filtering, 319 cases, 769 controls and 0 missing
After filtering, 511 males, 577 females, and 0 of unspecified sex
Writing recoded ped file to [ admurciav36f.ped ]
Writing new map file to [ admurciav36f.map ]
```

Analysis finished: Thu Jan 2 13:51:24 2014

Construcción 37/hg19

Así que sigo. Ahora hay que hacer el cambio a construcción 37/hg19

Ahora, a partir del *.map*, creo el siguiente archivo:

```
#Columna      Información
#1  # Cromosoma
#2  Posición de inicio (Columna3-1)
#3  Posición final
#4  Nombre del snp (rs...)

# Reemplazar      chr23 por chrX
#                chr24 por chrY
#                chr25 por chrX
```

```
$ awk '{print "chr"$1"\t"($4-1)"\t"$4"\t"$2}' admurciav36f.map > v36.txt
```

y subo *v36.txt* a <http://genome.ucsc.edu/cgi-bin/hgLiftOver>

Obtengo algo así:

```
Successfully converted 194109 records: View Conversions
```

```
Conversion failed on 4320 records.      Display failure file      Explain
failure messages
```

```
Failed input regions:
```

```
#Deleted in new
chr3    195567371    195567372    rs4974514
#Deleted in new
chr4    32575207     32575208     rs11096892
#Deleted in new
chr4    40025321     40025322     rs4512036
#Deleted in new
chr4    103677906    103677907    rs230525
#Deleted in new
chr4    103951974    103951975    rs223413
#Deleted in new
chr4    176359371    176359372    rs1711984
#Deleted in new
chr5    17739883     17739884     rs1849834
#Deleted in new
chr5    97752918     97752919     rs2124263
.....
```

De aqui obtengo dos archivos, los convertidos: *hglft_genome_1156_bb5460.bed* y los errores: *hglft_genome_1156_bb5460.err*. Primeramente tengo que eliminar los convertidos.

```
$ grep -v "#Deleted in new" hglft_genome_1156_bb5460.err | awk {'print $4'}
> delete_snps.txt
$ plink --noweb --bfile admurciav36f --exclude delete_snps.txt --make-bed -
-out admurciav36f_nofailures
$ plink --noweb --bfile admurciav36f_nofailures --recode --tab --out
admurciav37
```

Luego convierto el archivo de snps convertidos en un *.map*. Lo que tengo es,

```
$ head hglft_genome_1156_bb5460.bed
chr1    785988    785989    rs2980300
chr1    1130726   1130727   rs10907175
chr1    1268846   1268847   rs307378
chr1    1506034   1506035   rs7540231
chr1    1721478   1721479   rs2272908
chr1    1759025   1759026   rs9786963
chr1    1759053   1759054   rs10907187
chr1    1823921   1823922   rs6688000
chr1    2104980   2104981   rs262641
chr1    2105877   2105878   rs3128309
```

y haciendo

```
$ sed 's/^chr//' hglft_genome_1156_bb5460.bed | awk {'print
$1"\t"$4"\t0"\t"$3'} > v37.map
```

Obtengo,

```
$ head v37.map
1    rs2980300    0    785989
1    rs10907175   0    1130727
1    rs307378     0    1268847
1    rs7540231    0    1506035
1    rs2272908    0    1721479
1    rs9786963    0    1759026
1    rs10907187   0    1759054
1    rs6688000    0    1823922
1    rs262641     0    2104981
1    rs3128309    0    2105878
```

Por ultimo tomo este *.map* y obtengo mi archivo final.

```
$ cp admurciav37.ped v37.ped
$ plink --noweb --file v37 --make-bed --out v37
```

Prephasing

Primero voy a separar por cromosomas la base de datos que he obtenido. Esto es muy rapido.

```
$ parallel plink --noweb --bfile v37 --chr {} --recode --make-bed --out v37chr{ }\.unphased ::: {1..22..1}
```

Ahora empieza lo divertido. Con [shapeit](#) revisamos las inconsistencias entre los SNPs de nuestra base de datos y 1000Genome,

```
$ parallel shapeit -check -B v37chr{ }.unphased --input-ref /nas/data/GNlib/1000Genome/ALL_1000G_phase1integrated_v3_impute/ALL_1000G_phase1integrated_v3_chr{ }_impute.hap.gz /nas/data/GNlib/1000Genome/ALL_1000G_phase1integrated_v3_impute/ALL_1000G_phase1integrated_v3_chr{ }_impute.legend.gz /nas/data/GNlib/1000Genome/ALL_1000G_phase1integrated_v3_impute/ALL_1000G_phase1integrated_v3.sample --output-log v37chr{ }.alignments ::: {1..22..1}
```

Con esto obtengo para cada cromosoma (XX):

- v37chrXX.alignments.log
- v37chrXX.alignments.snp.strand (incluye los errores encontrados por incompatibilidad de alelos strand y SNPs que se encuentran en nuestro gwas que no se encuentran en el panel de referencia missing)
- v37chrXX.alignments.snp.strand.exclude (posiciones cromosómicas de todos los SNPs que dan error)

Primeramente voy a intentar hacer un *flip* con los archivos ue dan errores,

```
$ for x in {1..22..1}; do awk '{if($1=="strand") print $3}' v37chr$x.alignments.snp.strand > strand$x; done  
$ parallel plink --noweb --bfile v37chr{ }.unphased --flip strand{ } --make-bed --out v37chr{ }f.unphased ::: {1..22..1}
```

Ahora miro los errores que no se han solucionado con el *flip* y los excluyo de la base de datos.

```
$ parallel shapeit -check -B v37chr{ }f.unphased --input-ref /nas/data/GNlib/1000Genome/ALL_1000G_phase1integrated_v3_impute/ALL_1000G_phase1integrated_v3_chr{ }_impute.hap.gz /nas/data/GNlib/1000Genome/ALL_1000G_phase1integrated_v3_impute/ALL_1000G_phase1integrated_v3_chr{ }_impute.legend.gz /nas/data/GNlib/1000Genome/ALL_1000G_phase1integrated_v3_impute/ALL_1000G_phase1integrated_v3.sample --output-log v37chr{ }f.alignments ::: {1..22..1}  
$ for x in {1..22..1}; do awk '{print $3}' v37chr{x}f.alignments.snp.strand > exclude{x}; done  
$ parallel plink --bfile v37chr{ }f.unphased --exclude exclude{ } --make-bed --out admurchr{ }.unphased ::: {1..22..1}
```

Y felizmente ahora vamos a hacer el pre-phasing,

```
$ screen bash
$ for x in {1..22..1}; do shapeit --input-bed admurchr${x}.unphased --input-map
/nas/data/GNlib/1000Genome/ALL_1000G_phase1integrated_v3_impute/genetic_map_chr${x}_combined_b37.txt --thread 30 --effective-size 11418 -0
admurchr${x}.haps admurchr${x}.sample --output-log admurchr${x}.log; done
```

Imputando

Tras esto ya estamos preparados para imputar. la orden general para imputar la ventana de 15000000 a 16999999 en el cromosoma 22 es algo así,

```
impute2 -m
/nas/data/GNlib/1000Genome/ALL_1000G_phase1integrated_v3_impute/genetic_map_chr22_combined_b37.txt -known_haps_g admurchr22.haps -h
/nas/data/GNlib/1000Genome/ALL_1000G_phase1integrated_v3_impute/ALL_1000G_phase1integrated_v3_chr22_impute.hap.gz -l
/nas/data/GNlib/1000Genome/ALL_1000G_phase1integrated_v3_impute/ALL_1000G_phase1integrated_v3_chr22_impute.legend.gz -Ne 20000 -int 15000000 16999999 -o
admur_chr22.15000000-16999999.impute2 --allow_large_regions --seed 367946
```

Lo que he hecho es un sencillo script en perl que me ejecute secuencialmente la imputacion en todas las ventanas de un cromosoma. Como se fabrican estas ventanas es algo oscuro para mi asi que las he copiado y pegado.

[imputechr.pl](#)

```
#!/usr/bin/perl

# Copyright 2013 O. Sotolongo <osotolongo@fundacioace.com>

use strict; use warnings;
my @regs = (
    [qw (100000 2100000 4100000 6100000 8100000 10100000 12100000
14100000 16100000 18100000 20100000 22100000 24100000 26100000 28100000
30100000 32100000 34100000 36100000 38100000 40100000 42100000 44100000
46100000 48100000 50100000 52100000 54100000 56100000 58100000 60100000
62100000 64100000 66100000 68100000 70100000 72100000 74100000 76100000
78100000 80100000 82100000 84100000 86100000 88100000 90100000 92100000
94100000 96100000 98100000 100100000 102100000 104100000 106100000
108100000 110100000 112100000 114100000 116100000 118100000 120100000
122100000 124100000 126100000 128100000 130100000 132100000 134100000
136100000 138100000 140100000 142100000 144100000 146100000 148100000
150100000 152100000 154100000 156100000 158100000 160100000 162100000
164100000 166100000 168100000 170100000 172100000 174100000 176100000
178100000 180100000 182100000 184100000 186100000 188100000 190100000
192100000 194100000 196100000 198100000 200100000 202100000 204100000
206100000 208100000 210100000 212100000 214100000 216100000 218100000
220100000 222100000 224100000 226100000 228100000 230100000 232100000
```

```
234100000 236100000 238100000 240100000 242100000 244100000 246100000
248100000)],
[qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000 64010000 66010000 68010000 70010000 72010000 74010000 76010000
78010000 80010000 82010000 84010000 86010000 88010000 90010000 92010000
94010000 96010000 98010000 100010000 102010000 104010000 106010000
108010000 110010000 112010000 114010000 116010000 118010000 120010000
122010000 124010000 126010000 128010000 130010000 132010000 134010000
136010000 138010000 140010000 142010000 144010000 146010000 148010000
150010000 152010000 154010000 156010000 158010000 160010000 162010000
164010000 166010000 168010000 170010000 172010000 174010000 176010000
178010000 180010000 182010000 184010000 186010000 188010000 190010000
192010000 194010000 196010000 198010000 200010000 202010000 204010000
206010000 208010000 210010000 212010000 214010000 216010000 218010000
220010000 222010000 224010000 226010000 228010000 230010000 232010000
234010000 236010000 238010000 240010000 242010000)],
[qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000 64010000 66010000 68010000 70010000 72010000 74010000 76010000
78010000 80010000 82010000 84010000 86010000 88010000 90010000 92010000
94010000 96010000 98010000 100010000 102010000 104010000 106010000
108010000 110010000 112010000 114010000 116010000 118010000 120010000
122010000 124010000 126010000 128010000 130010000 132010000 134010000
136010000 138010000 140010000 142010000 144010000 146010000 148010000
150010000 152010000 154010000 156010000 158010000 160010000 162010000
164010000 166010000 168010000 170010000 172010000 174010000 176010000
178010000 180010000 182010000 184010000 186010000 188010000 190010000
192010000 194010000 196010000)],
[qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000 64010000 66010000 68010000 70010000 72010000 74010000 76010000
78010000 80010000 82010000 84010000 86010000 88010000 90010000 92010000
94010000 96010000 98010000 100010000 102010000 104010000 106010000
108010000 110010000 112010000 114010000 116010000 118010000 120010000
122010000 124010000 126010000 128010000 130010000 132010000 134010000
136010000 138010000 140010000 142010000 144010000 146010000 148010000
150010000 152010000 154010000 156010000 158010000 160010000 162010000
164010000 166010000 168010000 170010000 172010000 174010000 176010000
178010000 180010000 182010000 184010000 186010000 188010000
190010000)],
[qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
```



```
62010000 64010000 66010000 68010000 70010000 72010000 74010000 76010000
78010000 80010000 82010000 84010000 86010000 88010000 90010000 92010000
94010000 96010000 98010000 100010000 102010000 104010000 106010000
108010000 110010000 112010000 114010000 116010000 118010000 120010000
122010000 124010000 126010000 128010000 130010000 132010000 134010000
136010000 138010000 140010000 142010000 144010000 146010000 148010000
150010000 152010000 154010000 156010000 158010000 160010000 162010000
164010000 166010000 168010000 170010000 172010000 174010000 176010000
178010000 180010000)],
[qw (100000 2100000 4100000 6100000 8100000 10100000 12100000
14100000 16100000 18100000 20100000 22100000 24100000 26100000 28100000
30100000 32100000 34100000 36100000 38100000 40100000 42100000 44100000
46100000 48100000 50100000 52100000 54100000 56100000 58100000 60100000
62100000 64100000 66100000 68100000 70100000 72100000 74100000 76100000
78100000 80100000 82100000 84100000 86100000 88100000 90100000 92100000
94100000 96100000 98100000 100100000 102100000 104100000 106100000
108100000 110100000 112100000 114100000 116100000 118100000 120100000
122100000 124100000 126100000 128100000 130100000 132100000 134100000
136100000 138100000 140100000 142100000 144100000 146100000 148100000
150100000 152100000 154100000 156100000 158100000 160100000 162100000
164100000 166100000 168100000 170100000)],
[qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000 64010000 66010000 68010000 70010000 72010000 74010000 76010000
78010000 80010000 82010000 84010000 86010000 88010000 90010000 92010000
94010000 96010000 98010000 100010000 102010000 104010000 106010000
108010000 110010000 112010000 114010000 116010000 118010000 120010000
122010000 124010000 126010000 128010000 130010000 132010000 134010000
136010000 138010000 140010000 142010000 144010000 146010000 148010000
150010000 152010000 154010000 156010000 158010000)],
[qw (100000 2100000 4100000 6100000 8100000 10100000 12100000
14100000 16100000 18100000 20100000 22100000 24100000 26100000 28100000
30100000 32100000 34100000 36100000 38100000 40100000 42100000 44100000
46100000 48100000 50100000 52100000 54100000 56100000 58100000 60100000
62100000 64100000 66100000 68100000 70100000 72100000 74100000 76100000
78100000 80100000 82100000 84100000 86100000 88100000 90100000 92100000
94100000 96100000 98100000 100100000 102100000 104100000 106100000
108100000 110100000 112100000 114100000 116100000 118100000 120100000
122100000 124100000 126100000 128100000 130100000 132100000 134100000
136100000 138100000 140100000 142100000 144100000 146100000)],
[qw (100000 2100000 4100000 6100000 8100000 10100000 12100000
14100000 16100000 18100000 20100000 22100000 24100000 26100000 28100000
30100000 32100000 34100000 36100000 38100000 40100000 42100000 44100000
46100000 48100000 50100000 52100000 54100000 56100000 58100000 60100000
62100000 64100000 66100000 68100000 70100000 72100000 74100000 76100000
78100000 80100000 82100000 84100000 86100000 88100000 90100000 92100000
94100000 96100000 98100000 100100000 102100000 104100000 106100000
108100000 110100000 112100000 114100000 116100000 118100000 120100000
122100000 124100000 126100000 128100000 130100000 132100000 134100000
```

```
136100000 138100000 140100000)],
[qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000 64010000 66010000 68010000 70010000 72010000 74010000 76010000
78010000 80010000 82010000 84010000 86010000 88010000 90010000 92010000
94010000 96010000 98010000 100010000 102010000 104010000 106010000
108010000 110010000 112010000 114010000 116010000 118010000 120010000
122010000 124010000 126010000 128010000 130010000 132010000
134010000)],
[qw (100000 2100000 4100000 6100000 8100000 10100000 12100000
14100000 16100000 18100000 20100000 22100000 24100000 26100000 28100000
30100000 32100000 34100000 36100000 38100000 40100000 42100000 44100000
46100000 48100000 50100000 52100000 54100000 56100000 58100000 60100000
62100000 64100000 66100000 68100000 70100000 72100000 74100000 76100000
78100000 80100000 82100000 84100000 86100000 88100000 90100000 92100000
94100000 96100000 98100000 100100000 102100000 104100000 106100000
108100000 110100000 112100000 114100000 116100000 118100000 120100000
122100000 124100000 126100000 128100000 130100000 132100000
134100000)],
[qw (100000 2100000 4100000 6100000 8100000 10100000 12100000
14100000 16100000 18100000 20100000 22100000 24100000 26100000 28100000
30100000 32100000 34100000 36100000 38100000 40100000 42100000 44100000
46100000 48100000 50100000 52100000 54100000 56100000 58100000 60100000
62100000 64100000 66100000 68100000 70100000 72100000 74100000 76100000
78100000 80100000 82100000 84100000 86100000 88100000 90100000 92100000
94100000 96100000 98100000 100100000 102100000 104100000 106100000
108100000 110100000 112100000 114100000 116100000 118100000 120100000
122100000 124100000 126100000 128100000 130100000 132100000)],
[qw (18000000 20000000 22000000 24000000 26000000 28000000 30000000
32000000 34000000 36000000 38000000 40000000 42000000 44000000 46000000
48000000 50000000 52000000 54000000 56000000 58000000 60000000 62000000
64000000 66000000 68000000 70000000 72000000 74000000 76000000 78000000
80000000 82000000 84000000 86000000 88000000 90000000 92000000 94000000
96000000 98000000 100000000 102000000 104000000 106000000 108000000
110000000 112000000 114000000)],
[qw (18000000 20000000 22000000 24000000 26000000 28000000 30000000
32000000 34000000 36000000 38000000 40000000 42000000 44000000 46000000
48000000 50000000 52000000 54000000 56000000 58000000 60000000 62000000
64000000 66000000 68000000 70000000 72000000 74000000 76000000 78000000
80000000 82000000 84000000 86000000 88000000 90000000 92000000 94000000
96000000 98000000 100000000 102000000 104000000 106000000)],
[qw (19000000 21000000 23000000 25000000 27000000 29000000 31000000
33000000 35000000 37000000 39000000 41000000 43000000 45000000 47000000
49000000 51000000 53000000 55000000 57000000 59000000 61000000 63000000
65000000 67000000 69000000 71000000 73000000 75000000 77000000 79000000
81000000 83000000 85000000 87000000 89000000 91000000 93000000 95000000
97000000 99000000 101000000)],
[qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
```

```

30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000 64010000 66010000 68010000 70010000 72010000 74010000 76010000
78010000 80010000 82010000 84010000 86010000 88010000 90010000)],
    [qw (1000 2001000 4001000 6001000 8001000 10001000 12001000
14001000 16001000 18001000 20001000 22001000 24001000 26001000 28001000
30001000 32001000 34001000 36001000 38001000 40001000 42001000 44001000
46001000 48001000 50001000 52001000 54001000 56001000 58001000 60001000
62001000 64001000 66001000 68001000 70001000 72001000 74001000 76001000
78001000 80001000)],
    [qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000 64010000 66010000 68010000 70010000 72010000 74010000 76010000
78010000)],
    [qw (100000 2100000 4100000 6100000 8100000 10100000 12100000
14100000 16100000 18100000 20100000 22100000 24100000 26100000 28100000
30100000 32100000 34100000 36100000 38100000 40100000 42100000 44100000
46100000 48100000 50100000 52100000 54100000 56100000 58100000)],
    [qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000)],
    [qw (10000000 12000000 14000000 16000000 18000000 20000000 22000000
24000000 26000000 28000000 30000000 32000000 34000000 36000000 38000000
40000000 42000000 44000000 46000000 48000000)],
    [qw (15000000 17000000 19000000 21000000 23000000 25000000 27000000
29000000 31000000 33000000 35000000 37000000 39000000 41000000 43000000
45000000 47000000 49000000 51000000)],
);
my $window = 1999999;
my $libdir =
'/nas/data/GNlib/1000Genome/ALL_1000G_phaselintegrated_v3_impute/';
my $chr = shift;
die "Must supply chromosome!\n" unless $chr;
my $base = $regs[$chr-1];
for my $i (0 .. ${$base}) {
    my $w1 = $base->[$i];
    my $w2 = $w1 + $window;
    my $order = "impute2 -m
".$libdir."genetic_map_chr".$chr."_combined_b37.txt -known_haps_g
admurchr".$chr.".haps -h
".$libdir."ALL_1000G_phaselintegrated_v3_chr".$chr."_impute.hap.gz -l
".$libdir."ALL_1000G_phaselintegrated_v3_chr".$chr."_impute.legend.gz -
Ne 20000 -int ".$w1." ".$w2." -o admur_chr".$chr.".".$w1."-
".$w2.".impute2 --allow_large_regions --seed 367946";
    print "$order\n";
    system($order);
}

```

```
}
```

Ahora ejecutar esto ya es muy sencillo

```
parallel imputechr.pl {} ::: {1..22..1}
```

No esta muy optimizado ya que utilizo solo 22 procesadores, los cromosomas no estan balanceados, etc pero tampoco creo que valga la pena romperse la cabeza con esto.

Tras esto hay que pegarlos archivos *.impute2* pero voy a probar con uno a ver que pasa,

```
$ wc -l admur_chr22.*.impute2
  2163 admur_chr22.15000000-16999999.impute2
 26326 admur_chr22.17000000-18999999.impute2
 23261 admur_chr22.19000000-20999999.impute2
 23821 admur_chr22.21000000-22999999.impute2
 29967 admur_chr22.23000000-24999999.impute2
 30049 admur_chr22.25000000-26999999.impute2
 27506 admur_chr22.27000000-28999999.impute2
 24557 admur_chr22.29000000-30999999.impute2
 26377 admur_chr22.31000000-32999999.impute2
 30715 admur_chr22.33000000-34999999.impute2
 29381 admur_chr22.35000000-36999999.impute2
 27769 admur_chr22.37000000-38999999.impute2
 25536 admur_chr22.39000000-40999999.impute2
 24608 admur_chr22.41000000-42999999.impute2
 31071 admur_chr22.43000000-44999999.impute2
 32923 admur_chr22.45000000-46999999.impute2
 36053 admur_chr22.47000000-48999999.impute2
 36468 admur_chr22.49000000-50999999.impute2
  3052 admur_chr22.51000000-52999999.impute2
 491603 total
$ cat admur_chr22.*.impute2 > admur_chr22.gen
$ wc -l admur_chr22.gen
491603 admur_chr22.gen
```

Parece que va bien, pero cuidado, si se hace un **cat** indiscriminado no va a respetar el orden de los archivos,

```
$ wc -l admur_chr1.*.impute2
 20908 admur_chr1.100000-20999999.impute2
 27413 admur_chr1.100100000-102099999.impute2
 28473 admur_chr1.10100000-120999999.impute2
 28073 admur_chr1.102100000-104099999.impute2
 30590 admur_chr1.104100000-106099999.impute2
 29565 admur_chr1.106100000-108099999.impute2
 23361 admur_chr1.108100000-110099999.impute2
 27206 admur_chr1.110100000-112099999.impute2
 26169 admur_chr1.112100000-114099999.impute2
 25630 admur_chr1.114100000-116099999.impute2
```

```
26798 admur_chr1.116100000-118099999.impute2
27182 admur_chr1.118100000-120099999.impute2
 9486 admur_chr1.120100000-122099999.impute2
17754 admur_chr1.121000000-140999999.impute2
31730 admur_chr1.141000000-160999999.impute2
 5863 admur_chr1.144100000-146099999.impute2
15432 admur_chr1.146100000-148099999.impute2
 8211 admur_chr1.148100000-150099999.impute2
22784 admur_chr1.150100000-152099999.impute2
27905 admur_chr1.152100000-154099999.impute2
21489 admur_chr1.154100000-156099999.impute2
27749 admur_chr1.156100000-158099999.impute2
28988 admur_chr1.158100000-160099999.impute2
26522 admur_chr1.160100000-162099999.impute2
24375 admur_chr1.161000000-180999999.impute2
28623 admur_chr1.162100000-164099999.impute2
29061 admur_chr1.164100000-166099999.impute2
29108 admur_chr1.166100000-168099999.impute2
28942 admur_chr1.168100000-170099999.impute2
27559 admur_chr1.170100000-172099999.impute2
24188 admur_chr1.172100000-174099999.impute2
26726 admur_chr1.174100000-176099999.impute2
25819 admur_chr1.176100000-178099999.impute2
26084 admur_chr1.178100000-180099999.impute2
27678 admur_chr1.180100000-182099999.impute2
30910 admur_chr1.181000000-200999999.impute2
27082 admur_chr1.182100000-184099999.impute2
26659 admur_chr1.184100000-186099999.impute2
27929 admur_chr1.186100000-188099999.impute2
31630 admur_chr1.188100000-190099999.impute2
26320 admur_chr1.190100000-192099999.impute2
27083 admur_chr1.192100000-194099999.impute2
30094 admur_chr1.194100000-196099999.impute2
26517 admur_chr1.196100000-198099999.impute2
26585 admur_chr1.198100000-200099999.impute2
27000 admur_chr1.200100000-202099999.impute2
26731 admur_chr1.201000000-220999999.impute2
26190 admur_chr1.202100000-204099999.impute2
24815 admur_chr1.204100000-206099999.impute2
21534 admur_chr1.206100000-208099999.impute2
29269 admur_chr1.208100000-210099999.impute2
28422 admur_chr1.21000000-40999999.impute2
27718 admur_chr1.210100000-212099999.impute2
```

.....

así que tengo que hacerlo *mas o menos* o mano. Por suerte tengo el script perl anterior, asi que voy a reutilizarlo.

[catchr.pl](#)

```
#!/usr/bin/perl
```

```
# Copyright 2013 O. Sotolongo <osotolongo@fundacioace.com>
```

```
use strict; use warnings;
my @regs = (
  [qw (100000 2100000 4100000 6100000 8100000 10100000 12100000
14100000 16100000 18100000 20100000 22100000 24100000 26100000 28100000
30100000 32100000 34100000 36100000 38100000 40100000 42100000 44100000
46100000 48100000 50100000 52100000 54100000 56100000 58100000 60100000
62100000 64100000 66100000 68100000 70100000 72100000 74100000 76100000
78100000 80100000 82100000 84100000 86100000 88100000 90100000 92100000
94100000 96100000 98100000 100100000 102100000 104100000 106100000
108100000 110100000 112100000 114100000 116100000 118100000 120100000
122100000 124100000 126100000 128100000 130100000 132100000 134100000
136100000 138100000 140100000 142100000 144100000 146100000 148100000
150100000 152100000 154100000 156100000 158100000 160100000 162100000
164100000 166100000 168100000 170100000 172100000 174100000 176100000
178100000 180100000 182100000 184100000 186100000 188100000 190100000
192100000 194100000 196100000 198100000 200100000 202100000 204100000
206100000 208100000 210100000 212100000 214100000 216100000 218100000
220100000 222100000 224100000 226100000 228100000 230100000 232100000
234100000 236100000 238100000 240100000 242100000 244100000 246100000
248100000)],
  [qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000 64010000 66010000 68010000 70010000 72010000 74010000 76010000
78010000 80010000 82010000 84010000 86010000 88010000 90010000 92010000
94010000 96010000 98010000 100010000 102010000 104010000 106010000
108010000 110010000 112010000 114010000 116010000 118010000 120010000
122010000 124010000 126010000 128010000 130010000 132010000 134010000
136010000 138010000 140010000 142010000 144010000 146010000 148010000
150010000 152010000 154010000 156010000 158010000 160010000 162010000
164010000 166010000 168010000 170010000 172010000 174010000 176010000
178010000 180010000 182010000 184010000 186010000 188010000 190010000
192010000 194010000 196010000 198010000 200010000 202010000 204010000
206010000 208010000 210010000 212010000 214010000 216010000 218010000
220010000 222010000 224010000 226010000 228010000 230010000 232010000
234010000 236010000 238010000 240010000 242010000)],
  [qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000 64010000 66010000 68010000 70010000 72010000 74010000 76010000
78010000 80010000 82010000 84010000 86010000 88010000 90010000 92010000
94010000 96010000 98010000 100010000 102010000 104010000 106010000
108010000 110010000 112010000 114010000 116010000 118010000 120010000
122010000 124010000 126010000 128010000 130010000 132010000 134010000
136010000 138010000 140010000 142010000 144010000 146010000 148010000
150010000 152010000 154010000 156010000 158010000 160010000 162010000
```

```
164010000 166010000 168010000 170010000 172010000 174010000 176010000
178010000 180010000 182010000 184010000 186010000 188010000 190010000
192010000 194010000 196010000)],
[qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000 64010000 66010000 68010000 70010000 72010000 74010000 76010000
78010000 80010000 82010000 84010000 86010000 88010000 90010000 92010000
94010000 96010000 98010000 100010000 102010000 104010000 106010000
108010000 110010000 112010000 114010000 116010000 118010000 120010000
122010000 124010000 126010000 128010000 130010000 132010000 134010000
136010000 138010000 140010000 142010000 144010000 146010000 148010000
150010000 152010000 154010000 156010000 158010000 160010000 162010000
164010000 166010000 168010000 170010000 172010000 174010000 176010000
178010000 180010000 182010000 184010000 186010000 188010000
190010000)],
[qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000 64010000 66010000 68010000 70010000 72010000 74010000 76010000
78010000 80010000 82010000 84010000 86010000 88010000 90010000 92010000
94010000 96010000 98010000 100010000 102010000 104010000 106010000
108010000 110010000 112010000 114010000 116010000 118010000 120010000
122010000 124010000 126010000 128010000 130010000 132010000 134010000
136010000 138010000 140010000 142010000 144010000 146010000 148010000
150010000 152010000 154010000 156010000 158010000 160010000 162010000
164010000 166010000 168010000 170010000 172010000 174010000 176010000
178010000 180010000)],
[qw (100000 2100000 4100000 6100000 8100000 10100000 12100000
14100000 16100000 18100000 20100000 22100000 24100000 26100000 28100000
30100000 32100000 34100000 36100000 38100000 40100000 42100000 44100000
46100000 48100000 50100000 52100000 54100000 56100000 58100000 60100000
62100000 64100000 66100000 68100000 70100000 72100000 74100000 76100000
78100000 80100000 82100000 84100000 86100000 88100000 90100000 92100000
94100000 96100000 98100000 100100000 102100000 104100000 106100000
108100000 110100000 112100000 114100000 116100000 118100000 120100000
122100000 124100000 126100000 128100000 130100000 132100000 134100000
136100000 138100000 140100000 142100000 144100000 146100000 148100000
150100000 152100000 154100000 156100000 158100000 160100000 162100000
164100000 166100000 168100000 170100000)],
[qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000 64010000 66010000 68010000 70010000 72010000 74010000 76010000
78010000 80010000 82010000 84010000 86010000 88010000 90010000 92010000
94010000 96010000 98010000 100010000 102010000 104010000 106010000
108010000 110010000 112010000 114010000 116010000 118010000 120010000
122010000 124010000 126010000 128010000 130010000 132010000 134010000
```

```
136010000 138010000 140010000 142010000 144010000 146010000 148010000
150010000 152010000 154010000 156010000 158010000)],
[qw (100000 2100000 4100000 6100000 8100000 10100000 12100000
14100000 16100000 18100000 20100000 22100000 24100000 26100000 28100000
30100000 32100000 34100000 36100000 38100000 40100000 42100000 44100000
46100000 48100000 50100000 52100000 54100000 56100000 58100000 60100000
62100000 64100000 66100000 68100000 70100000 72100000 74100000 76100000
78100000 80100000 82100000 84100000 86100000 88100000 90100000 92100000
94100000 96100000 98100000 100100000 102100000 104100000 106100000
108100000 110100000 112100000 114100000 116100000 118100000 120100000
122100000 124100000 126100000 128100000 130100000 132100000 134100000
136100000 138100000 140100000 142100000 144100000 146100000)],
[qw (100000 2100000 4100000 6100000 8100000 10100000 12100000
14100000 16100000 18100000 20100000 22100000 24100000 26100000 28100000
30100000 32100000 34100000 36100000 38100000 40100000 42100000 44100000
46100000 48100000 50100000 52100000 54100000 56100000 58100000 60100000
62100000 64100000 66100000 68100000 70100000 72100000 74100000 76100000
78100000 80100000 82100000 84100000 86100000 88100000 90100000 92100000
94100000 96100000 98100000 100100000 102100000 104100000 106100000
108100000 110100000 112100000 114100000 116100000 118100000 120100000
122100000 124100000 126100000 128100000 130100000 132100000 134100000
136100000 138100000 140100000)],
[qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000 64010000 66010000 68010000 70010000 72010000 74010000 76010000
78010000 80010000 82010000 84010000 86010000 88010000 90010000 92010000
94010000 96010000 98010000 100010000 102010000 104010000 106010000
108010000 110010000 112010000 114010000 116010000 118010000 120010000
122010000 124010000 126010000 128010000 130010000 132010000
134010000)],
[qw (100000 2100000 4100000 6100000 8100000 10100000 12100000
14100000 16100000 18100000 20100000 22100000 24100000 26100000 28100000
30100000 32100000 34100000 36100000 38100000 40100000 42100000 44100000
46100000 48100000 50100000 52100000 54100000 56100000 58100000 60100000
62100000 64100000 66100000 68100000 70100000 72100000 74100000 76100000
78100000 80100000 82100000 84100000 86100000 88100000 90100000 92100000
94100000 96100000 98100000 100100000 102100000 104100000 106100000
108100000 110100000 112100000 114100000 116100000 118100000 120100000
122100000 124100000 126100000 128100000 130100000 132100000
134100000)],
[qw (100000 2100000 4100000 6100000 8100000 10100000 12100000
14100000 16100000 18100000 20100000 22100000 24100000 26100000 28100000
30100000 32100000 34100000 36100000 38100000 40100000 42100000 44100000
46100000 48100000 50100000 52100000 54100000 56100000 58100000 60100000
62100000 64100000 66100000 68100000 70100000 72100000 74100000 76100000
78100000 80100000 82100000 84100000 86100000 88100000 90100000 92100000
94100000 96100000 98100000 100100000 102100000 104100000 106100000
108100000 110100000 112100000 114100000 116100000 118100000 120100000
122100000 124100000 126100000 128100000 130100000 132100000)],
```



```
[qw (18000000 20000000 22000000 24000000 26000000 28000000 30000000
32000000 34000000 36000000 38000000 40000000 42000000 44000000 46000000
48000000 50000000 52000000 54000000 56000000 58000000 60000000 62000000
64000000 66000000 68000000 70000000 72000000 74000000 76000000 78000000
80000000 82000000 84000000 86000000 88000000 90000000 92000000 94000000
96000000 98000000 100000000 102000000 104000000 106000000 108000000
110000000 112000000 114000000)],
[qw (18000000 20000000 22000000 24000000 26000000 28000000 30000000
32000000 34000000 36000000 38000000 40000000 42000000 44000000 46000000
48000000 50000000 52000000 54000000 56000000 58000000 60000000 62000000
64000000 66000000 68000000 70000000 72000000 74000000 76000000 78000000
80000000 82000000 84000000 86000000 88000000 90000000 92000000 94000000
96000000 98000000 100000000 102000000 104000000 106000000)],
[qw (19000000 21000000 23000000 25000000 27000000 29000000 31000000
33000000 35000000 37000000 39000000 41000000 43000000 45000000 47000000
49000000 51000000 53000000 55000000 57000000 59000000 61000000 63000000
65000000 67000000 69000000 71000000 73000000 75000000 77000000 79000000
81000000 83000000 85000000 87000000 89000000 91000000 93000000 95000000
97000000 99000000 101000000)],
[qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000 64010000 66010000 68010000 70010000 72010000 74010000 76010000
78010000 80010000 82010000 84010000 86010000 88010000 90010000)],
[qw (1000 2001000 4001000 6001000 8001000 10001000 12001000
14001000 16001000 18001000 20001000 22001000 24001000 26001000 28001000
30001000 32001000 34001000 36001000 38001000 40001000 42001000 44001000
46001000 48001000 50001000 52001000 54001000 56001000 58001000 60001000
62001000 64001000 66001000 68001000 70001000 72001000 74001000 76001000
78001000 80001000)],
[qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000 64010000 66010000 68010000 70010000 72010000 74010000 76010000
78010000)],
[qw (100000 2100000 4100000 6100000 8100000 10100000 12100000
14100000 16100000 18100000 20100000 22100000 24100000 26100000 28100000
30100000 32100000 34100000 36100000 38100000 40100000 42100000 44100000
46100000 48100000 50100000 52100000 54100000 56100000 58100000)],
[qw (10000 2010000 4010000 6010000 8010000 10010000 12010000
14010000 16010000 18010000 20010000 22010000 24010000 26010000 28010000
30010000 32010000 34010000 36010000 38010000 40010000 42010000 44010000
46010000 48010000 50010000 52010000 54010000 56010000 58010000 60010000
62010000)],
[qw (10000000 12000000 14000000 16000000 18000000 20000000 22000000
24000000 26000000 28000000 30000000 32000000 34000000 36000000 38000000
40000000 42000000 44000000 46000000 48000000)],
[qw (15000000 17000000 19000000 21000000 23000000 25000000 27000000
29000000 31000000 33000000 35000000 37000000 39000000 41000000 43000000
```

```
45000000 47000000 49000000 51000000)],
);
my $window = 1999999;
my $chr = shift;
die "Must supply chromosome!\n" unless $chr;
my $base = $regs[$chr-1];
my $order = "cat ";
for my $i (0 .. ${$base}) {
    my $w1 = $base->[$i];
    my $w2 = $w1 + $window;
    $order .= "admur_chr".$chr.".".$w1."-".$w2.".impute2 ";
}
$order .= "> admur_chr".$chr.".gen";
system($order);
```

y ahora si terminamos de una vez,

```
$ for x in {1..22..1}; do catchr.pl $x; done
```

Esto nos deja con los archivos *.gen* y *.sample* por cada cromosoma.

From:
<https://mail.fundacioace.com/wiki/> - **Detritus Wiki**

Permanent link:
<https://mail.fundacioace.com/wiki/doku.php?id=genetica:impute2>

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

