Julia Smolik

1. Analiza brakujących danych.

```
indmiss<-read.table(file="plink.imiss", header=TRUE)
hist(indmiss[,6])

snpmiss<-read.table(file="plink.lmiss", header=TRUE)
hist(snpmiss[,5])

plink --bfile HapMap --geno 0.2 --make-bed --out HapMap_2
```

Żadne dane nie zostały odrzucone.

```
MacBook-Air-Julia:analiza julia$ plink --bfile HapMap --geno 0.2 --make-bed --out HapMap_2
PLINK v1.90b6.17 64-bit (28 Apr 2020)        www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to HapMap_2.log.
Options in effect:
  --bfile HapMap
  --geno 0.2
  --make-bed
  --out HapMap_2

8192 MB RAM detected; reserving 4096 MB for main workspace.
1457897 variants loaded from .bim file.
165 people (80 males, 85 females) loaded from .fam.
112 phenotype values loaded from .fam.
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 112 founders and 53 nonfounders present.
Calculating allele frequencies... done.
Warning: 225 het. haploid genotypes present (see HapMap_2.hh ); many commands
treat these as missing.
Total genotyping rate is 0.997378.
0 variants removed due to missing genotype data (--geno).
1457897 variants and 165 people pass filters and QC.
Among remaining phenotypes, 56 are cases and 56 are controls.  (53 phenotypes
are missing.)
--make-bed to HapMap_2.bed + HapMap_2.bim + HapMap_2.fam ... done.
MacBook-Air-Julia:analiza julia$
```

```
plink --bfile HapMap --geno 0.02 --make-bed --out HapMap_2
```

```
MacBook-Air-Julia:analiza julia$ plink --bfile HapMap --geno 0.02 --make-bed --out HapMap_2
PLINK v1.90b6.17 64-bit (28 Apr 2020)        www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to HapMap_2.log.
Options in effect:
  --bfile HapMap
  --geno 0.02
  --make-bed
  --out HapMap_2

8192 MB RAM detected; reserving 4096 MB for main workspace.
1457897 variants loaded from .bim file.
165 people (80 males, 85 females) loaded from .fam.
112 phenotype values loaded from .fam.
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 112 founders and 53 nonfounders present.
Calculating allele frequencies... done.
Warning: 225 het. haploid genotypes present (see HapMap_2.hh ); many commands
treat these as missing.
Total genotyping rate is 0.997378.
27454 variants removed due to missing genotype data (--geno).
1430443 variants and 165 people pass filters and QC.
Among remaining phenotypes, 56 are cases and 56 are controls.  (53 phenotypes
are missing.)
--make-bed to HapMap_2.bed + HapMap_2.bim + HapMap_2.fam ... done.
MacBook-Air-Julia:analiza julia$
```

Odrzucone: 27454; zostało: 1430443

Julia Smolik

```
plink --bfile HapMap_2 --mind 0.02 --make-bed --out HapMap_3
```

```
MacBook-Air-Julia:analiza julia$ plink --bfile HapMap_2 --mind 0.02 --make-bed --out HapMap_3
PLINK v1.90b6.17 64-bit (28 Apr 2020)        www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to HapMap_3.log.
Options in effect:
  --bfile HapMap_2
  --make-bed
  --mind 0.02
  --out HapMap_3

8192 MB RAM detected; reserving 4096 MB for main workspace.
1430443 variants loaded from .bim file.
165 people (80 males, 85 females) loaded from .fam.
112 phenotype values loaded from .fam.
0 people removed due to missing genotype data (--mind).
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 112 founders and 53 nonfounders present.
Calculating allele frequencies... done.
Warning: 179 het. haploid genotypes present (see HapMap_3.hh ); many commands
treat these as missing.
Total genotyping rate is 0.997899.
1430443 variants and 165 people pass filters and QC.
Among remaining phenotypes, 56 are cases and 56 are controls.  (53 phenotypes
are missing.)
--make-bed to HapMap_3.bed + HapMap_3.bim + HapMap_3.fam ... done.
MacBook-Air-Julia:analiza julia$
```

Odrzucone: 0; zostało: 1430443

2. Kontrola przypisania płci

```
plink --bfile HapMap_3 —check-sex
```

```
[MacBook-Air-Julia:analiza julia$ plink --bfile HapMap_3 --check-sex
PLINK v1.90b6.17 64-bit (28 Apr 2020)        www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to plink.log.
Options in effect:
  --bfile HapMap_3
  --check-sex

8192 MB RAM detected; reserving 4096 MB for main workspace.
1430443 variants loaded from .bim file.
165 people (80 males, 85 females) loaded from .fam.
112 phenotype values loaded from .fam.
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 112 founders and 53 nonfounders present.
Calculating allele frequencies... done.
Warning: 179 het. haploid genotypes present (see plink.hh ); many commands
treat these as missing.
Total genotyping rate is 0.997899.
1430443 variants and 165 people pass filters and QC.
Among remaining phenotypes, 56 are cases and 56 are controls.  (53 phenotypes
are missing.)
--check-sex: 23424 Xchr and 0 Ychr variant(s) scanned, 1 problem detected.
Report written to plink.sexcheck .
MacBook-Air-Julia:analiza julia$
```

Problem występuje gdy w kolumnie PEDSEX (2) wartość jest inna niż w SNPSEX (1).

3. Usunięcie chromosomu X

```
[MacBook-Air-Julia:analiza julia$ awk '{ if ($1 >= 1 && $1 <= 22) print $2 }' HapMap_4.bim>auto]
somal.txt
[MacBook-Air-Julia:analiza julia$ plink --bfile HapMap_4 --extract autosomal.txt --make-bed --o]
ut HapMap_5
PLINK v1.90b6.17 64-bit (28 Apr 2020)        www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to HapMap_5.log.
Options in effect:
  --bfile HapMap_4
  --extract autosomal.txt
  --make-bed
  --out HapMap_5

8192 MB RAM detected; reserving 4096 MB for main workspace.
1430443 variants loaded from .bim file.
164 people (80 males, 84 females) loaded from .fam.
112 phenotype values loaded from .fam.
--extract: 1398544 variants remaining.
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 112 founders and 52 nonfounders present.
Calculating allele frequencies... done.
Total genotyping rate is 0.998052.
1398544 variants and 164 people pass filters and QC.
Among remaining phenotypes, 56 are cases and 56 are controls.  (52 phenotypes
are missing.)
--make-bed to HapMap_5.bed + HapMap_5.bim + HapMap_5.fam ... done.
MacBook-Air-Julia:analiza julia$
```

Julia Smolik

4. **Kontrola częstości allelu rzadkiego**

```
plink --bfile HapMap_5 --freq --out MAF_check
```

```
maf_freq <- read.table("MAF_check.frq", header =TRUE)
```

```
hist (maf_freq[,5])
```
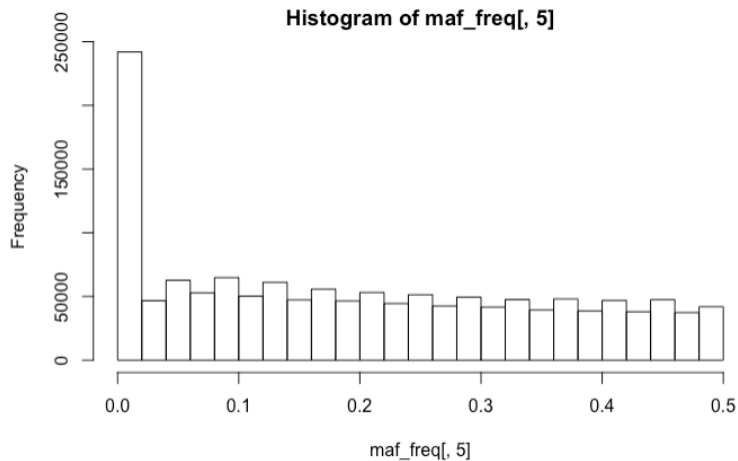


```
plink --bfile HapMap_5 --maf 0.05 --make-bed --out HapMap_6
```

```
MacBook-Air-Julia:analiza julia$ plink --bfile HapMap_5 --maf 0.05 --make-bed --out HapMap_6
PLINK v1.90b6.17 64-bit (28 Apr 2020)      www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to HapMap_6.log.
Options in effect:
  --bfile HapMap_5
  --maf 0.05
  --make-bed
  --out HapMap_6

8192 MB RAM detected; reserving 4096 MB for main workspace.
1398544 variants loaded from .bim file.
164 people (80 males, 84 females) loaded from .fam.
112 phenotype values loaded from .fam.
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 112 founders and 52 nonfounders present.
Calculating allele frequencies... done.
Total genotyping rate is 0.998052.
325318 variants removed due to minor allele threshold(s)
(--maf/--max-maf/--mac/--max-mac).
1073226 variants and 164 people pass filters and QC.
Among remaining phenotypes, 56 are cases and 56 are controls.  (52 phenotypes
are missing.)
--make-bed to HapMap_6.bed + HapMap_6.bim + HapMap_6.fam ... done.
MacBook-Air-Julia:analiza julia$
```

Usunięte: 325318; zostało: 1073226

Powtórz powyższą analizę histogramu częstości MAF dla uzyskanego pliku.

```
plink --bfile HapMap_6 --freq --out MAF_check_1
```

```
maf_freq2 <- read.table("MAF_check_1.frq", header =TRUE)
```

```
hist (maf_freq2[,5])
```

Histogram of maf_freq_2[, 5]

maf_freq_2[, 5]
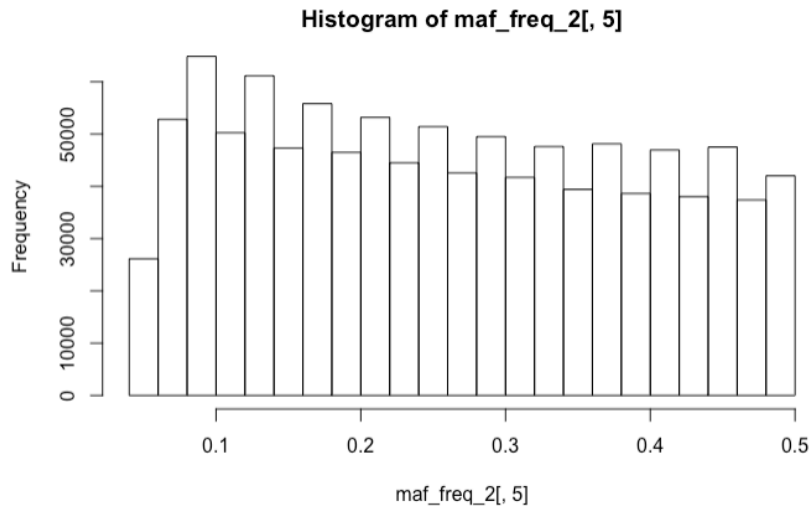
5. Kontrola równowagi Hardy'ego-Weinberga

```
plink --bfile HapMap_6 —hardy
```

```
head plink.hwe
```

```
[MacBook-Air-Julia:analiza julia$ head plink.hwe                                    ]
  CHR       SNP     TEST   A1   A2             GENO    O(HET)    E(HET)            P
    1   rs3131972    ALL    A    G           2/33/77    0.2946    0.2758       0.7324
    1   rs3131972    AFF    A    G           1/19/36    0.3393    0.3047        0.667
    1   rs3131972  UNAFF    A    G           1/14/41      0.25    0.2449            1
    1   rs3131969    ALL    A    G           2/26/84    0.2321     0.232            1
    1   rs3131969    AFF    A    G           1/17/38    0.3036    0.2817            1
    1   rs3131969  UNAFF    A    G            1/9/46    0.1607    0.1771       0.4189
    1   rs1048488    ALL    C    T           2/33/76    0.2973    0.2778       0.7324
    1   rs1048488    AFF    C    T           1/19/35    0.3455    0.3089       0.6661
    1   rs1048488  UNAFF    C    T           1/14/41      0.25    0.2449            1
MacBook-Air-Julia:analiza julia$ ▊
```

```
hwe<-read.table (file="plink.hwe", header=TRUE)
```

```
hist(hwe[,9])
```

Histogram of hwe[, 9]

hwe[, 9]
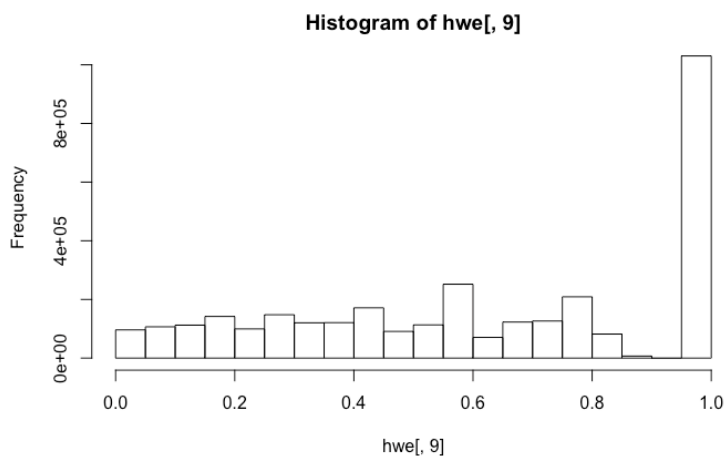
```
awk '{ if ($9 <1e-5) print $0 }' plink.hwe
```

```
plink --bfile HapMap_6 --hwe 1e-6 --make-bed --out HapMap_7
```

```
[MacBook-Air-Julia:analiza julia$ plink --bfile HapMap_6 --hwe 1e-6 --make-bed --out HapMap_7 ]
PLINK v1.90b6.17 64-bit (28 Apr 2020)        www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to HapMap_7.log.
Options in effect:
  --bfile HapMap_6
  --hwe 1e-6
  --make-bed
  --out HapMap_7

8192 MB RAM detected; reserving 4096 MB for main workspace.
1073226 variants loaded from .bim file.
164 people (80 males, 84 females) loaded from .fam.
112 phenotype values loaded from .fam.
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 112 founders and 52 nonfounders present.
Calculating allele frequencies... done.
Total genotyping rate is 0.998039.
--hwe: 0 variants removed due to Hardy-Weinberg exact test.
1073226 variants and 164 people pass filters and QC.
Among remaining phenotypes, 56 are cases and 56 are controls.  (52 phenotypes
are missing.)
--make-bed to HapMap_7.bed + HapMap_7.bim + HapMap_7.fam ... done.
MacBook-Air-Julia:analiza julia$ █
```

```
plink --bfile HapMap_7 --hwe 1e-10 --hwe-all --make-bed --out HapMap_8
```

```
[MacBook-Air-Julia:analiza julia$ plink --bfile HapMap_7 --hwe 1e-10 --hwe-all --make-bed --out HapMap_8 ]
PLINK v1.90b6.17 64-bit (28 Apr 2020)        www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to HapMap_8.log.
Options in effect:
  --bfile HapMap_7
  --hwe 1e-10
  --hwe-all
  --make-bed
  --out HapMap_8

Note: --hwe-all flag deprecated.  Use "--hwe include-nonctrl".
8192 MB RAM detected; reserving 4096 MB for main workspace.
1073226 variants loaded from .bim file.
164 people (80 males, 84 females) loaded from .fam.
112 phenotype values loaded from .fam.
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 112 founders and 52 nonfounders present.
Calculating allele frequencies... done.
Total genotyping rate is 0.998039.
--hwe: 0 variants removed due to Hardy-Weinberg exact test.
1073226 variants and 164 people pass filters and QC.
Among remaining phenotypes, 56 are cases and 56 are controls.  (52 phenotypes
are missing.)
--make-bed to HapMap_8.bed + HapMap_8.bim + HapMap_8.fam ... done.
MacBook-Air-Julia:analiza julia$ █
```

Obie komendy nie usunęły nic z danych. W komendzie awk pojawiło się wiele wyników. Jest wiele SNP, dla których P wynosi mniej niż $1 \cdot 10^{-5}$
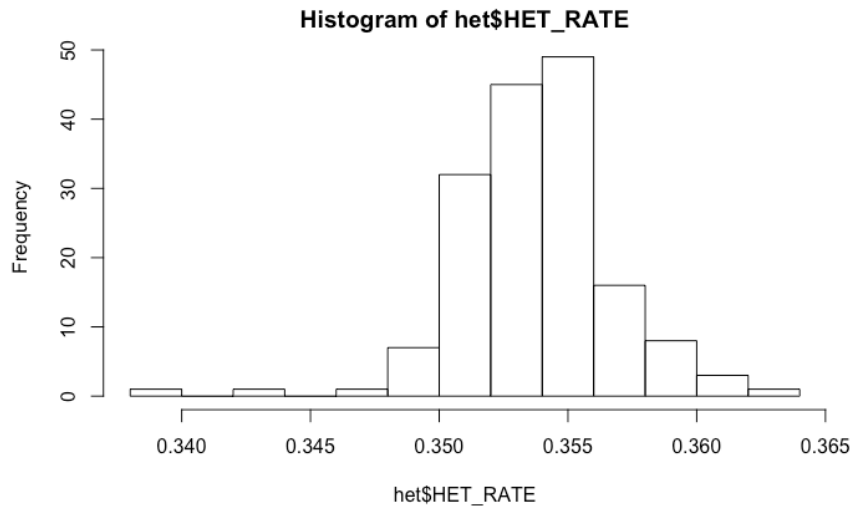
7. Analiza heterozygotyczności

```
plink --bfile HapMap_8 --extract indepSNP.prune.in --het --out R_check
```

```
het <- read.table("R_check.het", head=TRUE)
```

```
het$HET_RATE = (het$"N.NM." - het$"O.HOM.")/het$"N.NM."
```

```
hist(het$HET_RATE)
```

Histogram of het$HET_RATE

W histogramie widać osoby o skrajnych wartościach heterozygotyczności.

```
Rscript --no-save heterozygosity_outliers_list.R
```

```
"FID" "IID" "O.HOM." "E.HOM." "N.NM." "F" "HET_RATE" "HET_DST"
1330 "NA12342" 68049 67240 103571 0.02229 0.342972453679119 -3.66711854374478
1459 "NA12874" 68802 67560 104068 0.0339 0.338874582004074 -5.04839854982741
```

Na liście znalazły się dwie osoby. Odpowiadają one przedziałom: 0.3429 i 0.3388.

8.   Analiza pokrewieństwa

```
plink --bfile HapMap_10 --genome --min 0.2 --out pihat_min0.2
```

```
  FID1     IID1  FID2     IID2 RT    EZ      Z0     Z1       Z2 PI_HAT PHE       DST    PPC   RATIO
 13291 NA07045  1454 NA12813 UN    NA  0.2745 0.5006  0.2250 0.4753   0  0.833661 1.0000  9.8218
```

Została 1 para. Wartość pi-hat = 0.4753. Najprawdopodobniej jest to rodzeństwo.

```
grep identyfikator plink.imiss
```

```
MacBook-Air-Julia:analiza julia$ grep identyfikator plink.imiss
MacBook-Air-Julia:analiza julia$ head plink.imiss
   FID       IID MISS_PHENO   N_MISS   N_GENO    F_MISS
  1328   NA06989          N     4203  1457897  0.002883
  1377   NA11891          N    20787  1457897   0.01426
  1349   NA11843          N     1564  1457897  0.001073
  1330   NA12341          N     6218  1457897  0.004265
  1444   NA12739          Y    29584  1457897   0.02029
  1344   NA10850          Y     2631  1457897  0.001805
  1328   NA06984          N     9638  1457897  0.006611
  1463   NA12877          Y     3788  1457897  0.002598
  1418   NA12275          N     5349  1457897  0.003669
MacBook-Air-Julia:analiza julia$
```

Większą wartość N_MISS ma osoba o FID 1444 (IID: NA12739).

```
plink --bfile HapMap_10 --remove usun.txt --make-bed --out HapMap_11
```

```
[MacBook-Air-Julia:analiza julia$ plink --bfile HapMap_10 --remove usun.txt --make-bed --out HapMap_11      ]
PLINK v1.90b6.17 64-bit (28 Apr 2020)        www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to HapMap_11.log.
Options in effect:
  --bfile HapMap_10
  --make-bed
  --out HapMap_11
  --remove usun.txt

8192 MB RAM detected; reserving 4096 MB for main workspace.
1073226 variants loaded from .bim file.
112 people (57 males, 55 females) loaded from .fam.
112 phenotype values loaded from .fam.
--remove: 112 people remaining.
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 112 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Total genotyping rate is 0.997991.
1073226 variants and 112 people pass filters and QC.
Among remaining phenotypes, 56 are cases and 56 are controls.
--make-bed to HapMap_11.bed + HapMap_11.bim + HapMap_11.fam ... done.
MacBook-Air-Julia:analiza julia$ ▊
```

Zostało 112 osób.

9. Właściwa analiza.

```
plink —bfile HapMap_11 —asoc —out hm11
```

```
[MacBook-Air-Julia:analiza julia$ plink --bfile HapMap_11 --assoc --out hm11
PLINK v1.90b6.17 64-bit (28 Apr 2020)          www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to hm11.log.
Options in effect:
  --assoc
  --bfile HapMap_11
  --out hm11

8192 MB RAM detected; reserving 4096 MB for main workspace.
1073226 variants loaded from .bim file.
112 people (57 males, 55 females) loaded from .fam.
112 phenotype values loaded from .fam.
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 112 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Total genotyping rate is 0.997991.
1073226 variants and 112 people pass filters and QC.
Among remaining phenotypes, 56 are cases and 56 are controls.
Writing C/C --assoc report to hm11.assoc ... done.
MacBook-Air-Julia:analiza julia$ ▊
```

```
wyniki_as <- read.table("hm11.assoc", head=TRUE)

manhattan(wyniki_as)

manhattan(subset(wyniki_as, CHR == 1))

manhattan(subset(wyniki_as, CHR == 1), xlim=c(1.0e07,1.1e07)) #przyblizenia

manhattan(wyniki_as, annotatePval = 1e-5, annotateTop = FALSE) #nazwy SNP
spełniające kryterium

manhattan(subset(wyniki_as, CHR == 1),xlim=c(1.11e08,1.12e08), annotatePval =
1e-5, annotateTop = FALSE)
```
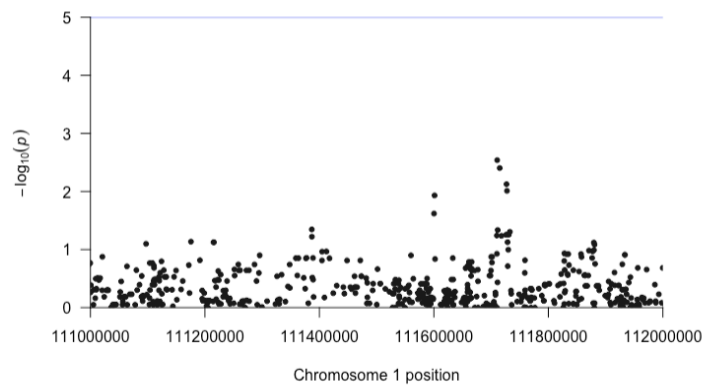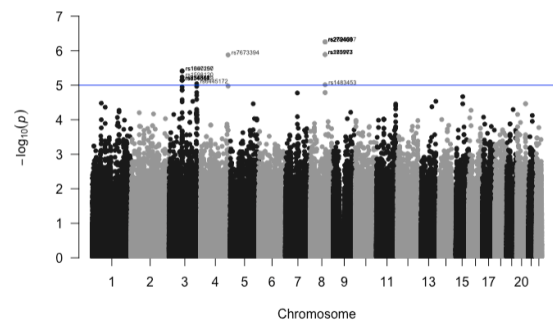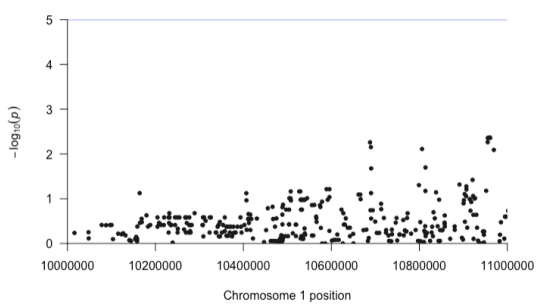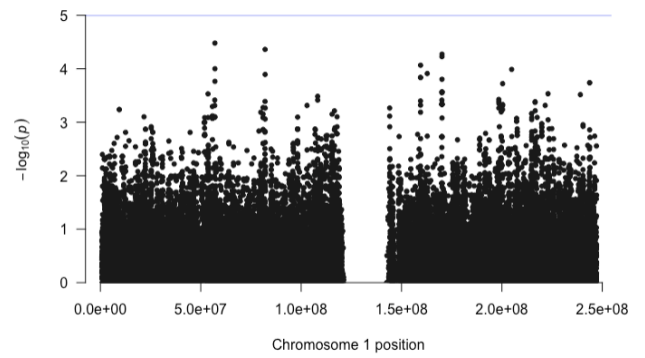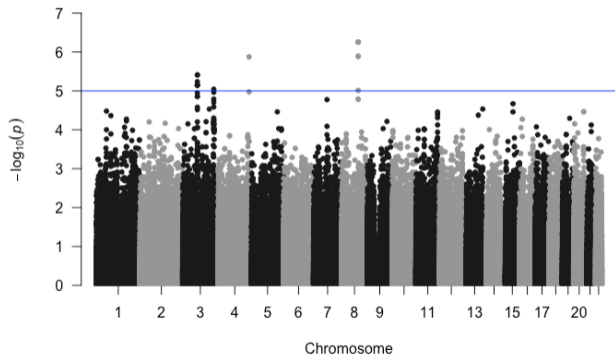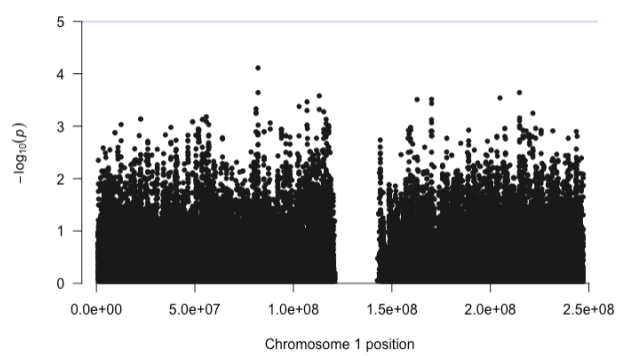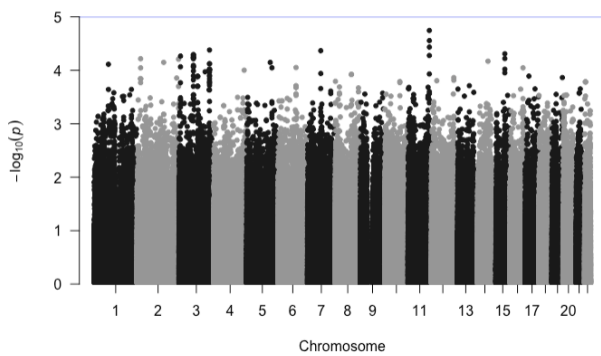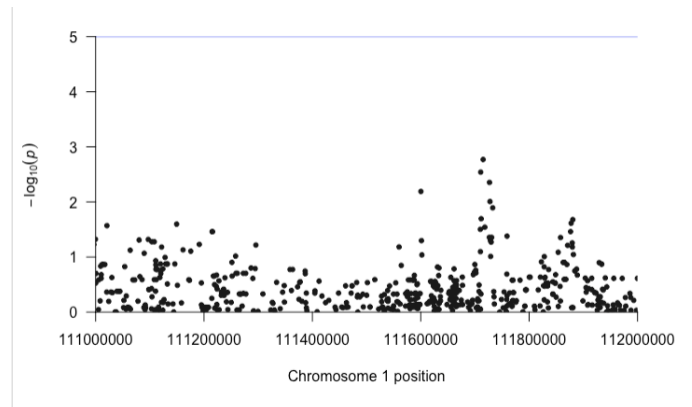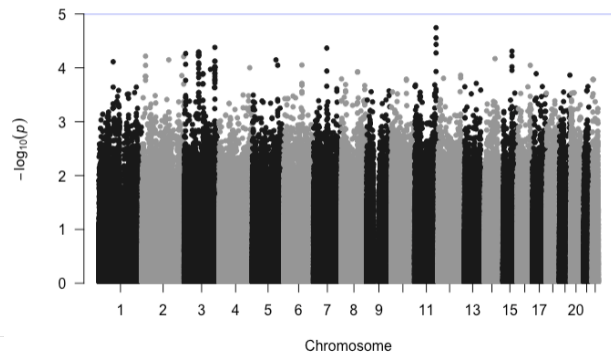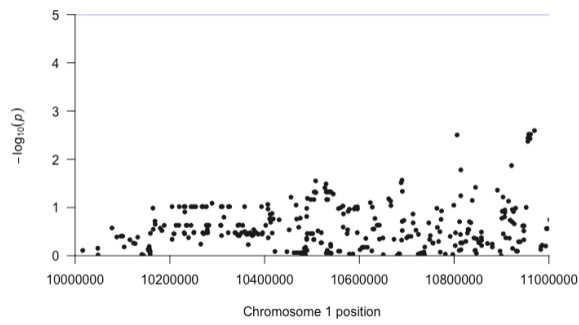
Julia Smolik

10. Analiza stratyfikacji (opcjonalnie)

Wartości P się zmieniły. Teraz są są większe, ponieważ wcześniej np. -log(x)=5, a teraz -log(x)=7.