# PLINK
# gPLINK
# Haploview

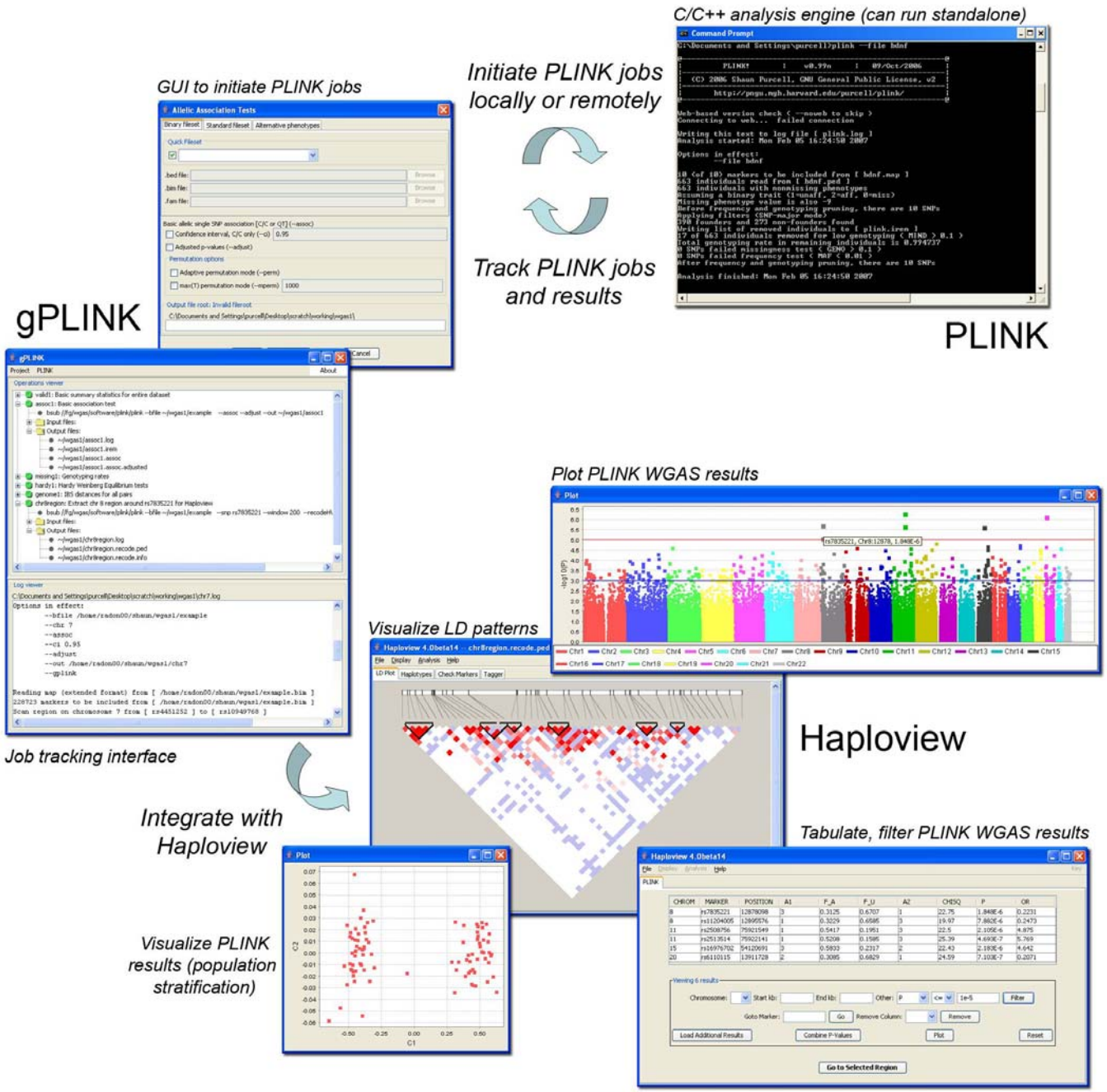
## Whole genome association software tutorial

*Shaun Purcell*

Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA
Broad Institute of Harvard & MIT, Cambridge, MA

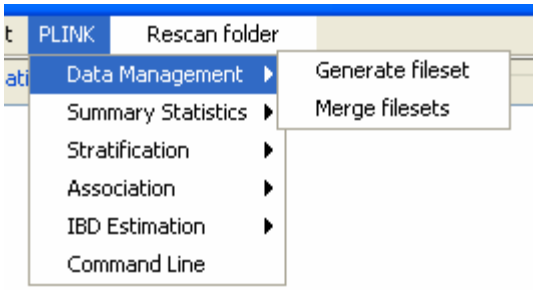
http://pngu.mgh.harvard.edu/purcell/plink/

http://www.broad.mit.edu/mpg/haploview/

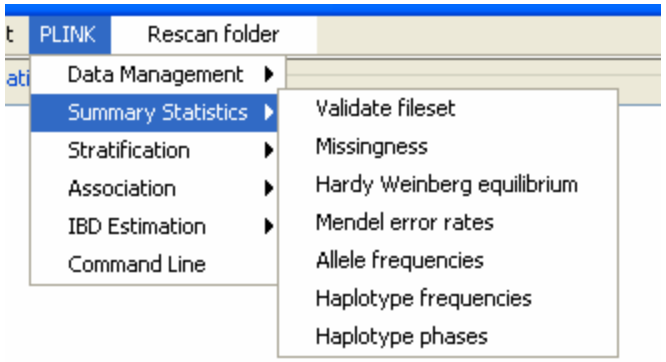

C/C++ analysis engine (can run standalone)
GUI to initiate PLINK jobs
Initiate PLINK jobs locally or remotely
Track PLINK jobs and results
gPLINK
PLINK
Plot PLINK WGAS results
Visualize LD patterns
Haploview
Job tracking interface
Integrate with Haploview
Tabulate, filter PLINK WGAS results
Visualize PLINK results (population stratification)

# GUI for many **PLINK** analyses

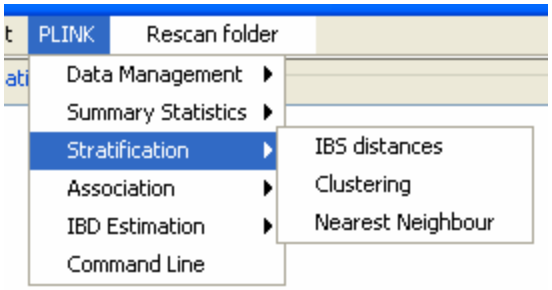
### *Data management*



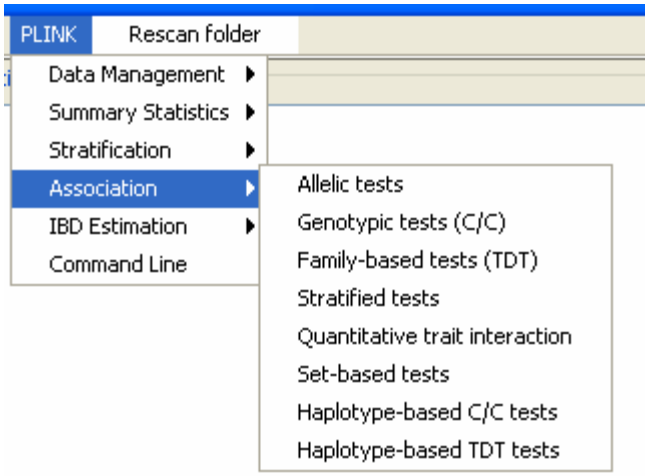

### *Summary statistics*



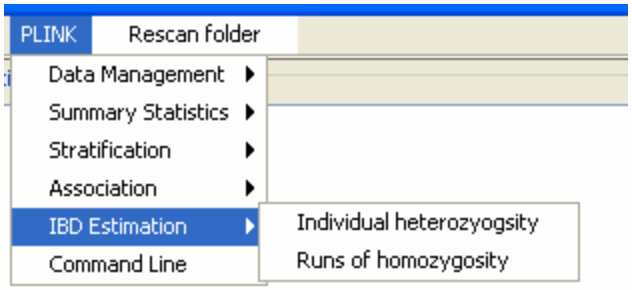

### *Population stratification*




### *Association analysis*




### *IBD-based analysis*

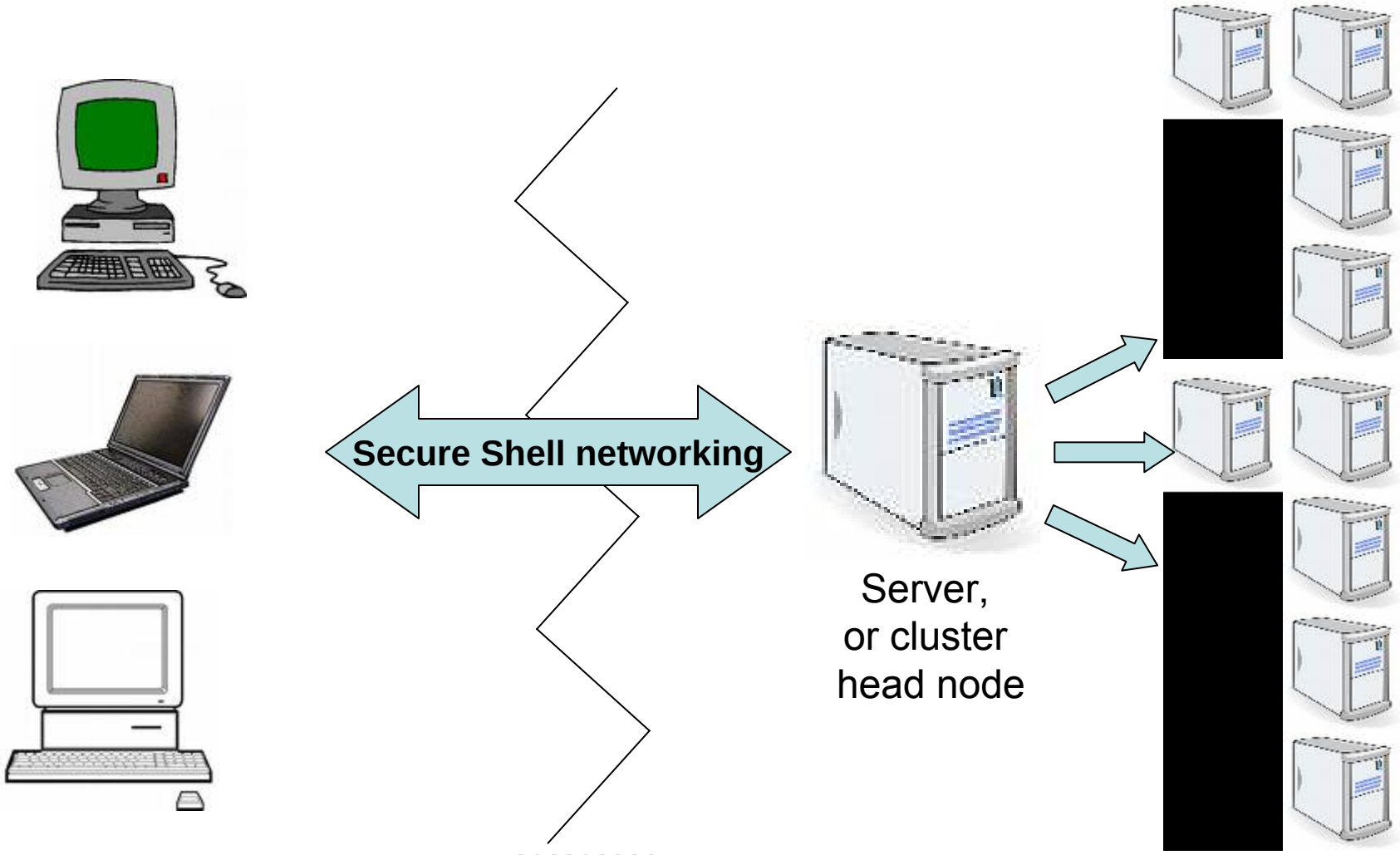


# Computational efficiency

*350 individuals genotyped on 100,000 SNPs*

| | |
|---|---|
| Load, filter and analyze | ~12 seconds |
| 1 permutation (all SNPs) | ~1.6 seconds |

# gPLINK / PLINK in “remote mode”




*gPLINK & Haploview: initiating and viewing jobs*

*PLINK, WGAS data & computation*

A simulated WGAS dataset

↓

Summary statistics and quality control

↓

Whole genome SNP-based association

↓

Whole genome haplotype-based association

↓

Assessment of population stratification

↓

Further exploration of ‘hits’

↓

Visualization and follow-up using Haploview

# In this practical, we will use **gPLINK**, **PLINK** and **Haploview** to…

… examine genotyping rates and look for non-random missing data

… determine SNP frequencies and test Hardy-Weinberg equilibrium

… assess population stratification via clustering, genomic control

… test for allelic, genotypic and haplotypic association

… perform stratified analyses, conditioning on population strata

… assess between-stratum heterogeneity in association signal

… examine linkage disequilibrium patterns around associated SNPs

… select tag SNPs for follow-up and replication studies

# Simulated WGAS dataset

- Real genotypes, but a simulated “disease”
- 90 Asian HapMap individuals
  - 10K autosomal SNPs from Affymetrix 500K product
- Simulated quantitative phenotype; median split to create a disease phenotype
- Illustrative, not realistic!

# Specific questions asked

1) What is the **genotyping rate**?

2) How many **monomorphic SNPs**?

3) Evidence of **non-random genotyping failure**?

4) What is the single **most associated SNP**? Does it reach genome-wide significance? What is the **most associated haplotype**?

5) Is there evidence of **population stratification from genomic control**?

6) Use genotypes to **cluster the sample** into 2 subpopulations. How well does the clustering recover the known Chinese/Japanese split?

7) Is there evidence for **stratification conditional on the two-cluster solution**?

8) What is the **best SNP controlling for stratification**. Is it genome-wide significant?

*For the most highly associated SNP:*

9) Does this SNP pass the **Hardy-Weinberg** equilibrium test?

10) Does this SNP **differ in frequency** between the two populations?

11) Is there evidence that this SNP has a **different association** between the two populations?

12) What are the **allele frequencies** in cases and controls? **Genotype** frequencies? What is the **odds ratio**?

13) Is the rate of **missing data** equal between cases and controls for this SNP?

14) Does an additive model well characterize the association? What about **genotypic, dominant models**, etc?

# Data used in this practical

- Available at http://pngu.mgh.harvard.edu/purcell/affy/purcell.zip

| | |
|---|---|
| `example.bed` | Binary format genotype information (do not attempt to view in a standard text editor) |
| `example.bim` | Map file (6 fields: each row is a SNP: chromosome, RS #, genetic position, physical position, allele 1, allele 2) |
| `example.fam` | Individual information file (first 6 columns of a PED file; disease phenotype is column 6) |
| `pop.phe` | Chinese/Japanese population indicator (FID, IID, population code) |
| `qt.phe` | Alternate quantitative trait phenotype file (Family ID, Individual ID, phenotype) |

# The Truth…

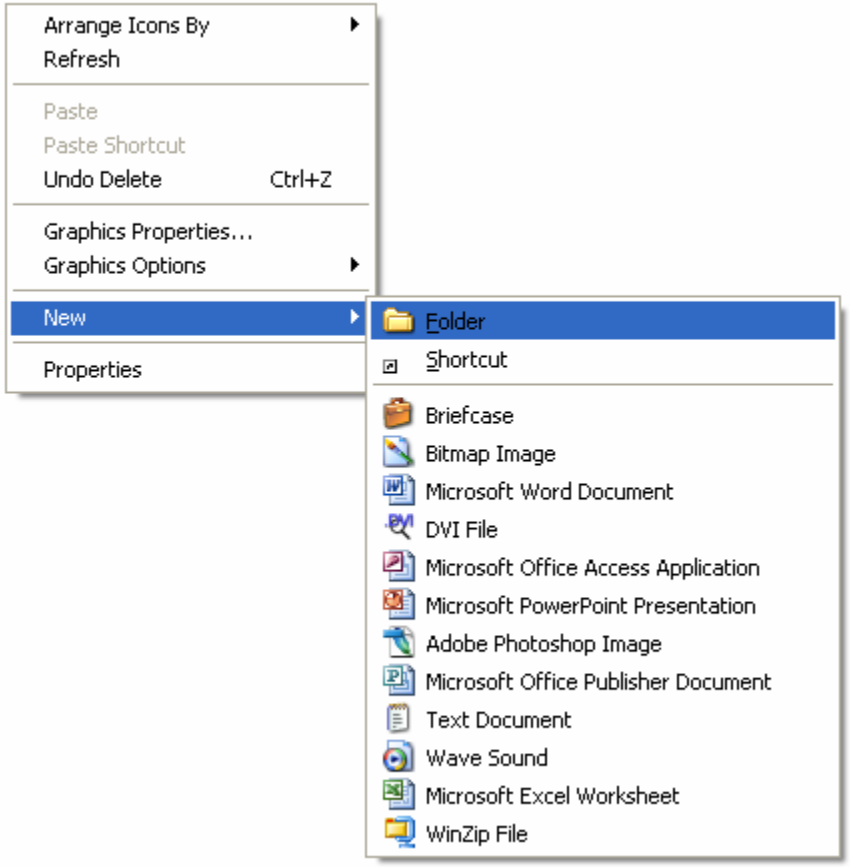
| | Chinese | Japanese |
|---|---|---|
| Case | 34 | 7 |
| Control | 11 | 38 |

*Group difference*

| | “11” | “12” | “22” |
|---|---|---|---|
| Case | 5 | 21 | 23 |
| Control | 16 | 23 | 2 |

*Single common variant*
*rs7835221 chr8*

# A **gPLINK** “project” is a folder

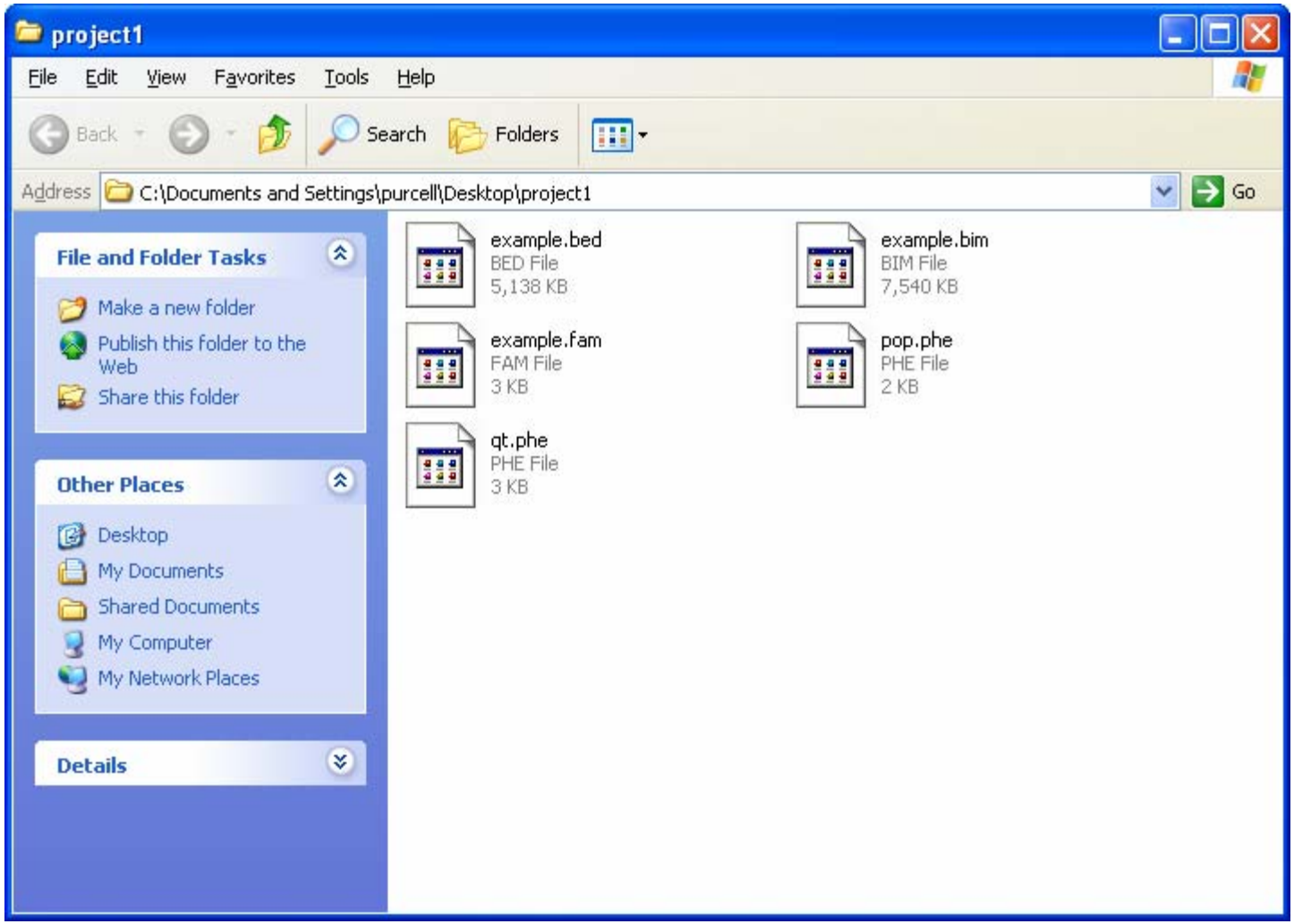




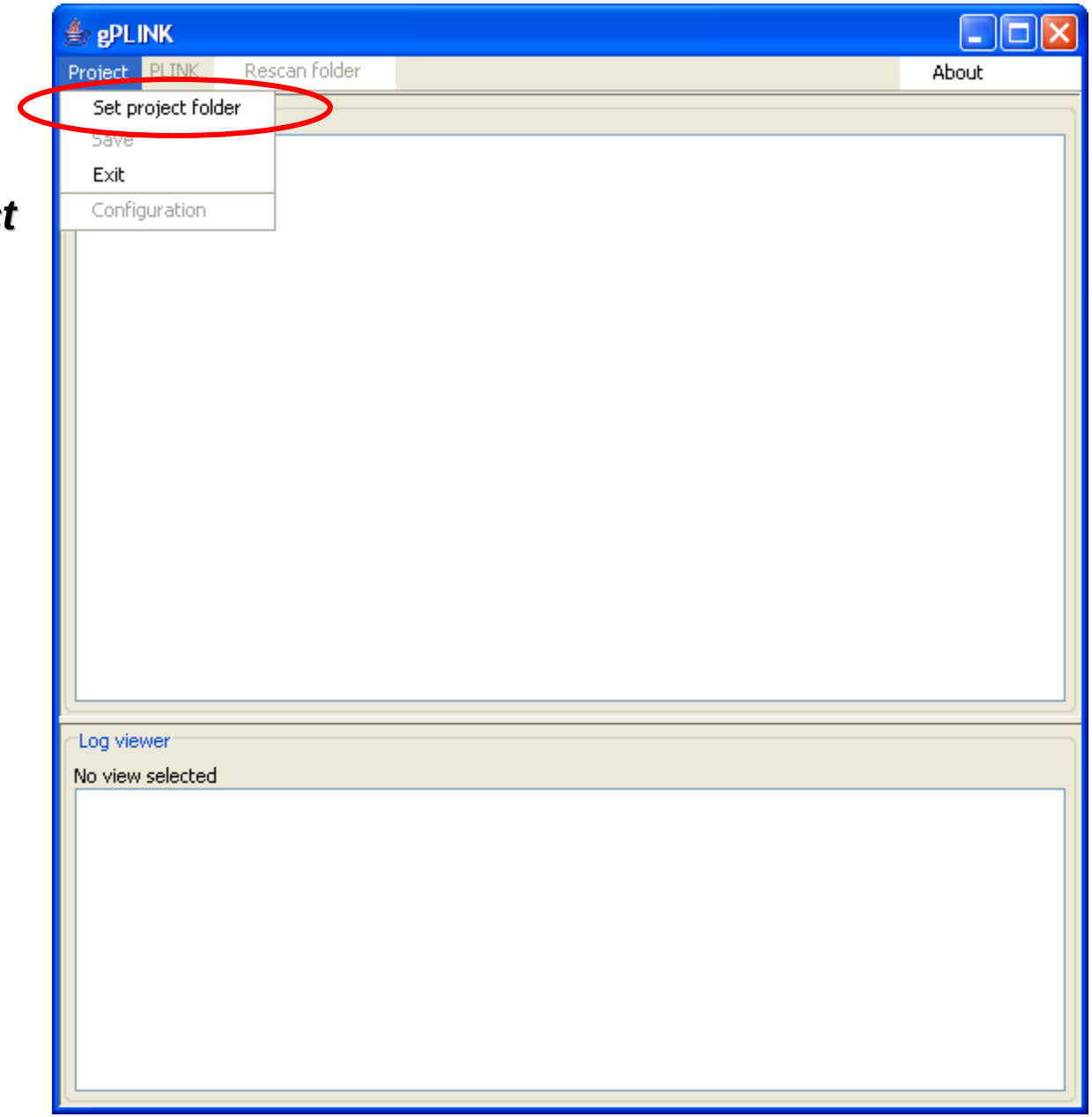
***Right-click on the Desktop to create a project folder…***

***…and rename it* “project1”**

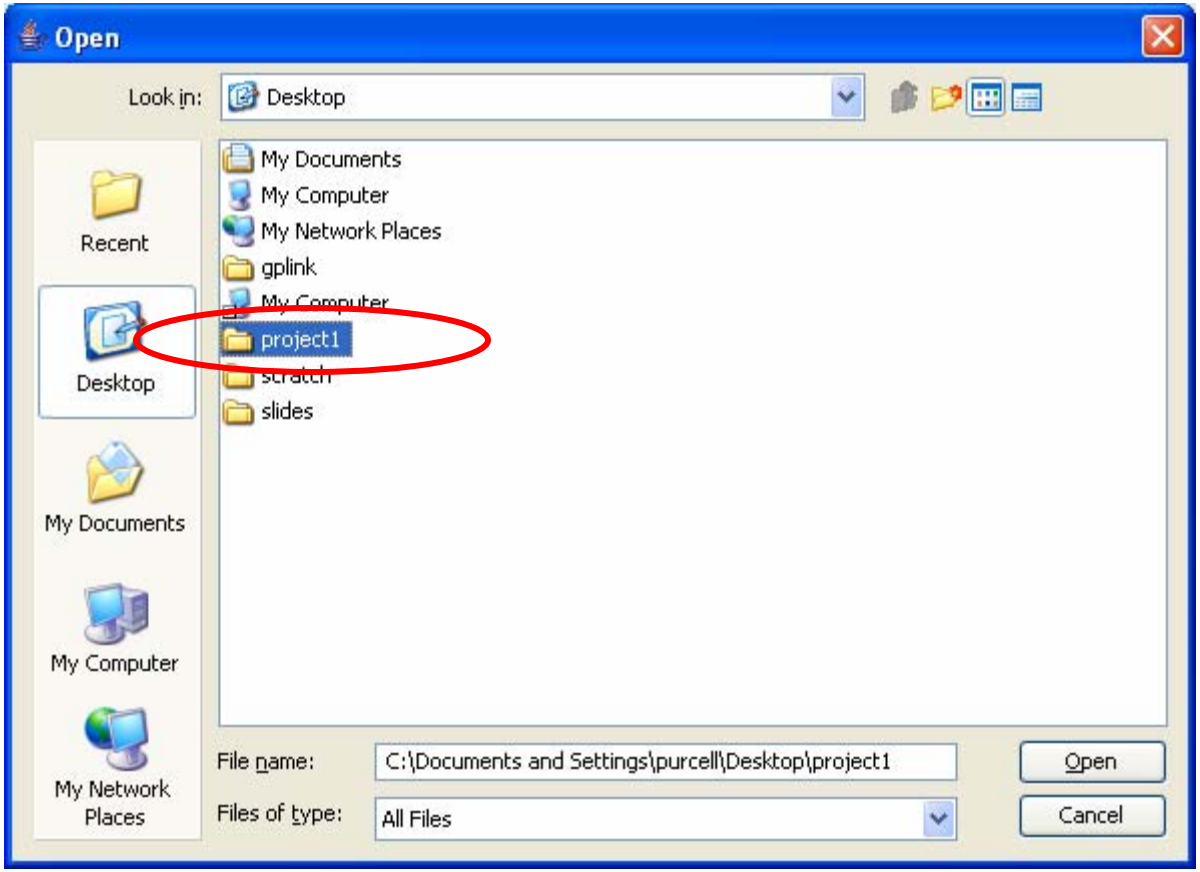
## *Copy the relevant files into this folder*



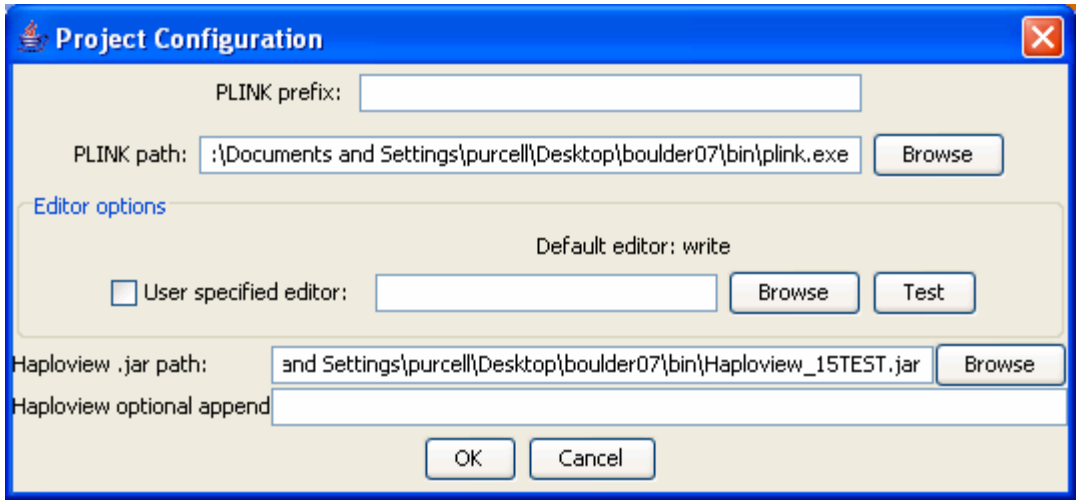
## *Start a new* gPLINK *project*



## *Select the folder you previously created*



# Configuring the new project




***Here, we tell* gPLINK…**

***… where the* PLINK *executable is***

***… specify any* PLINK *prefixes (advanced option for grid computing)***

***… where the* Haploview (version 4.0) *executable is***

***… which text editor to use to view files, e.g.* WordPad (write.exe)**

# Data management

- Recode dataset (A,C,G,T → 1,2)
- Reorder dataset
- Flip DNA strand
- Extract subsets (individuals, SNPs)
- Remove subsets (individuals, SNPs)
- Merge 2 or more filesets
- Compact binary file format

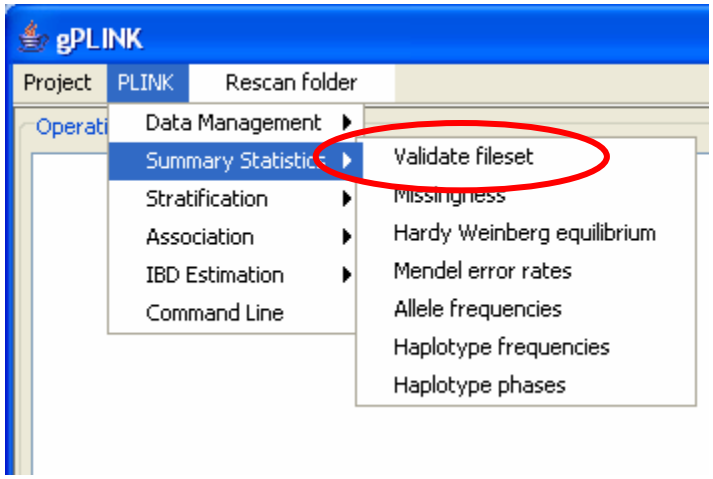
# Summarizing the data

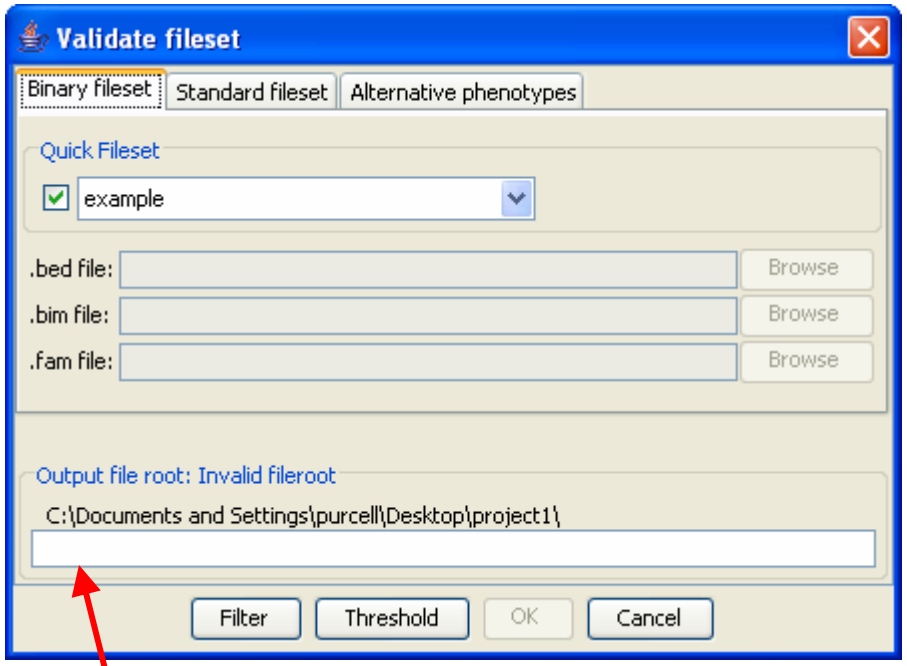
- Hardy-Weinberg
- Mendel errors
- Missing genotypes
- Allele frequencies

- Tests of non-random missingness
  - by phenotype and by (unobserved) genotype
- Individual homozygosity estimates
- Stretches of homozygosity
- Pairwise IBD estimates

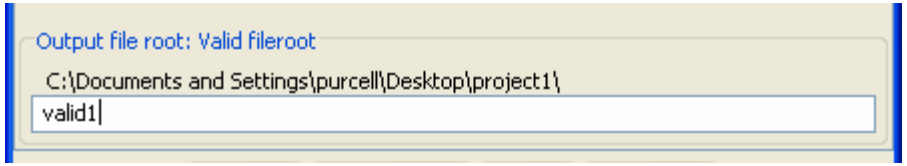
# Validating the fileset

***Doesn't do anything, except (attempt to) load the data and report basic statistics***

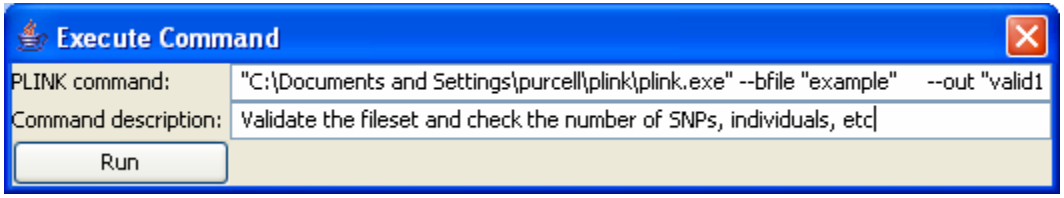



***Need to enter a unique root filename:***




***Then add a description (for logging)***

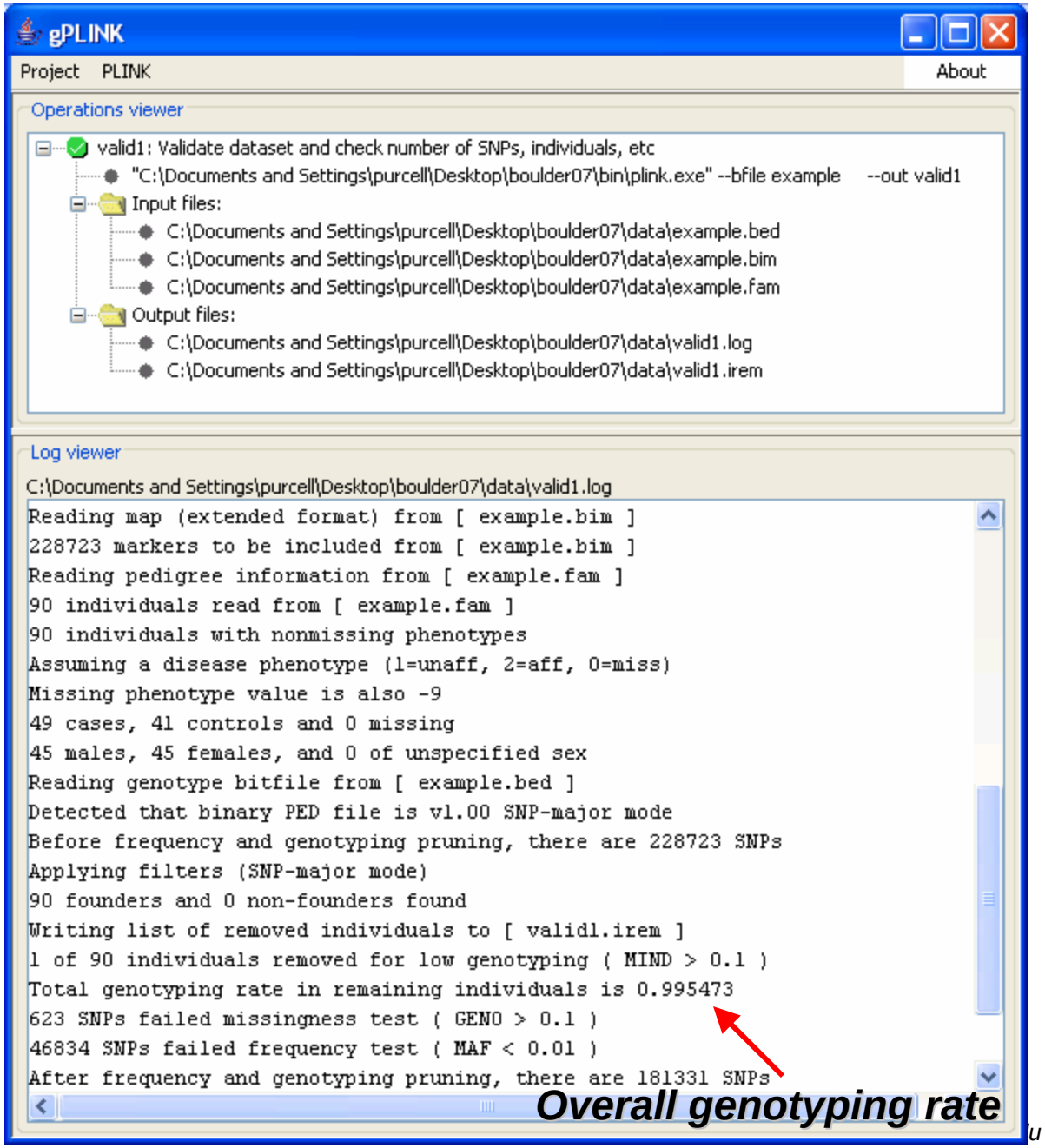


## *Q1) What is the genotyping rate?*

***Clicking on the tree to expand or contract it; individual input or output files can be selected here***

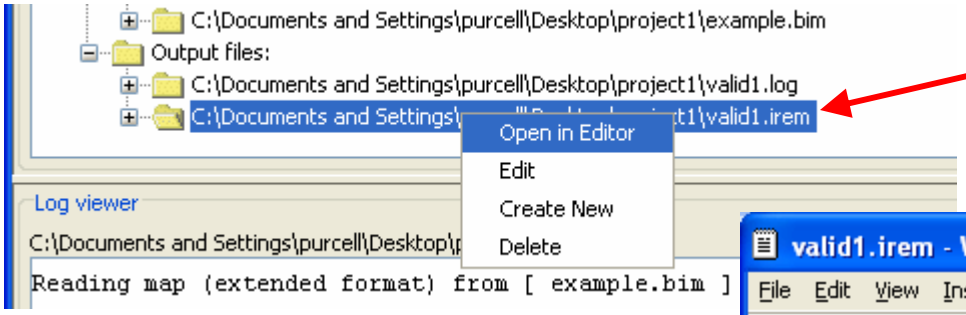
***The log file always gives a lot of useful information: it is good practice always to check it to confirm that an analysis has run okay.***

***Default filters applied here***

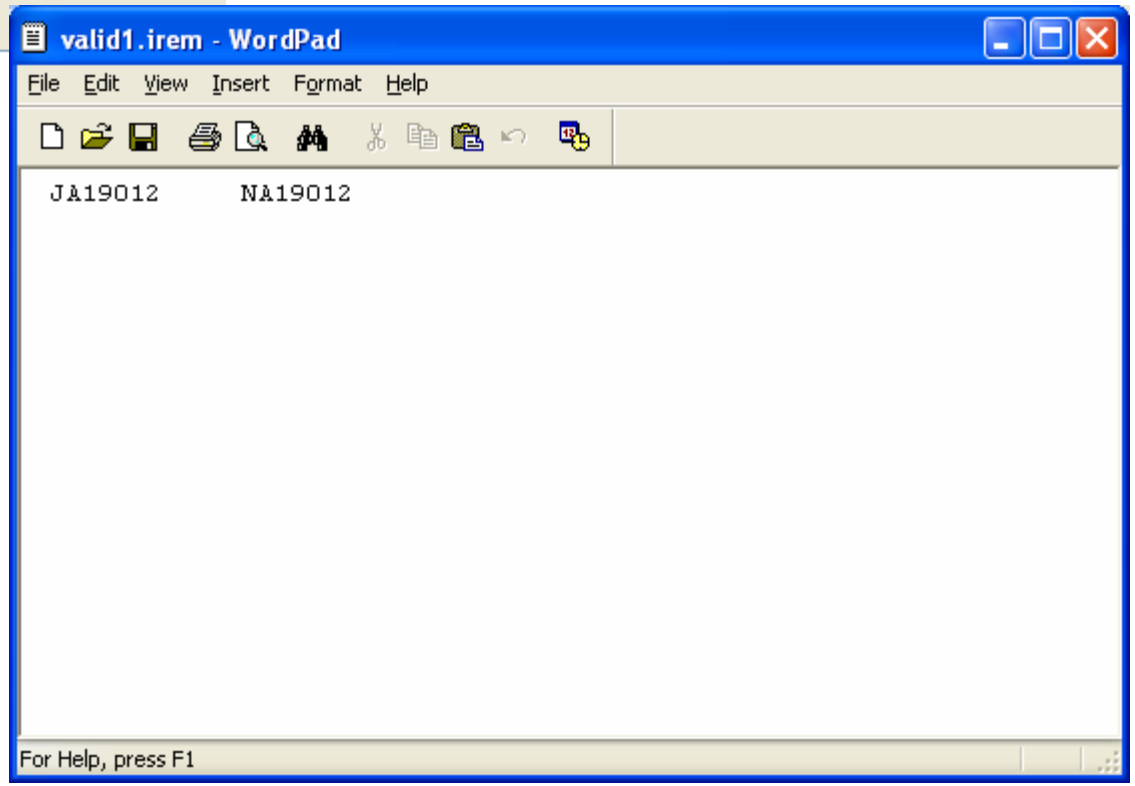



***Overall genotyping rate***

# Viewing an output file

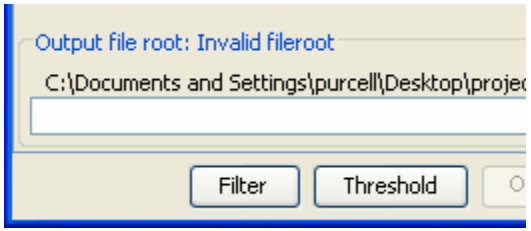



***Right-click on a selected file***



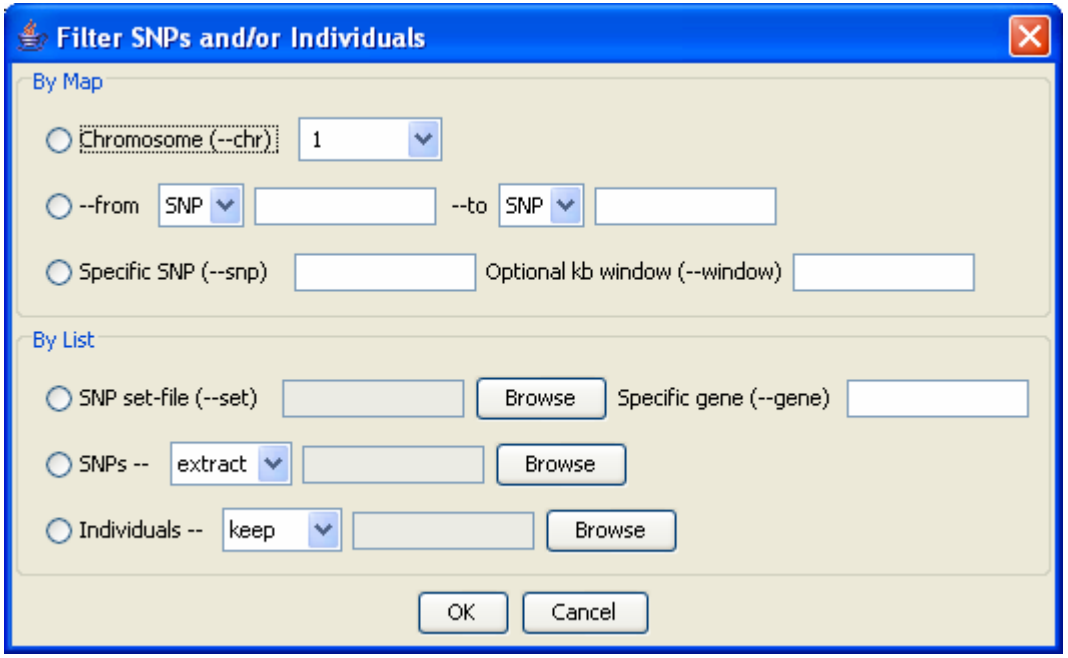

***In this case, a list of individuals excluded due to low genotyping rate (just one person here). (A line contains Family ID and Individual ID)***

# Filters and thresholds

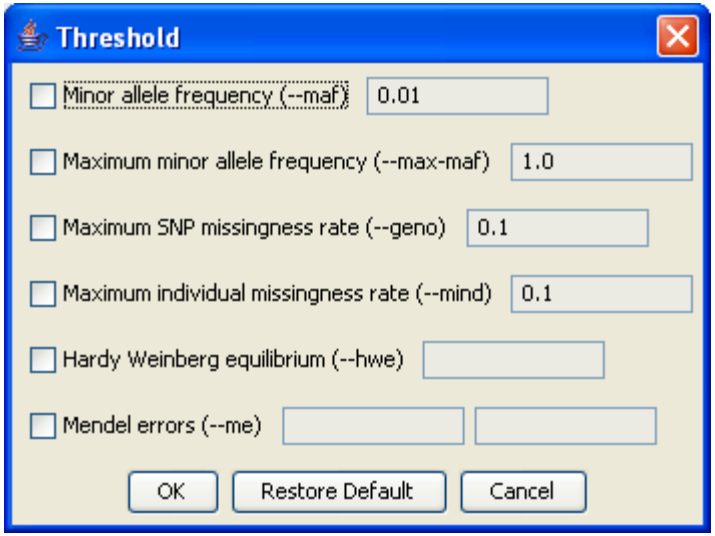



***Most forms have Filter and Thresholds buttons***

***Filters exclude people or SNPs based on prespecified lists, or genomic location***

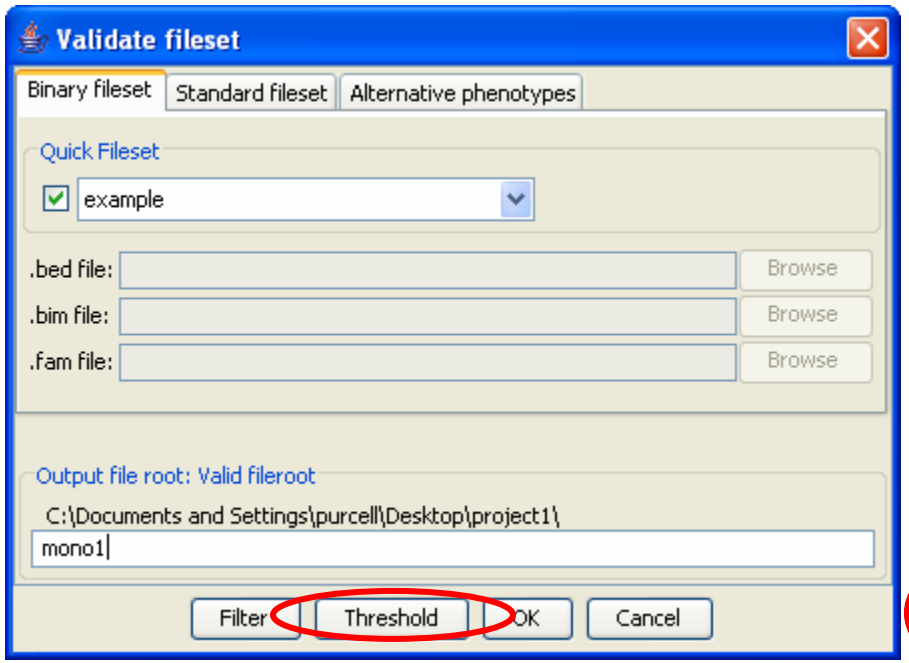


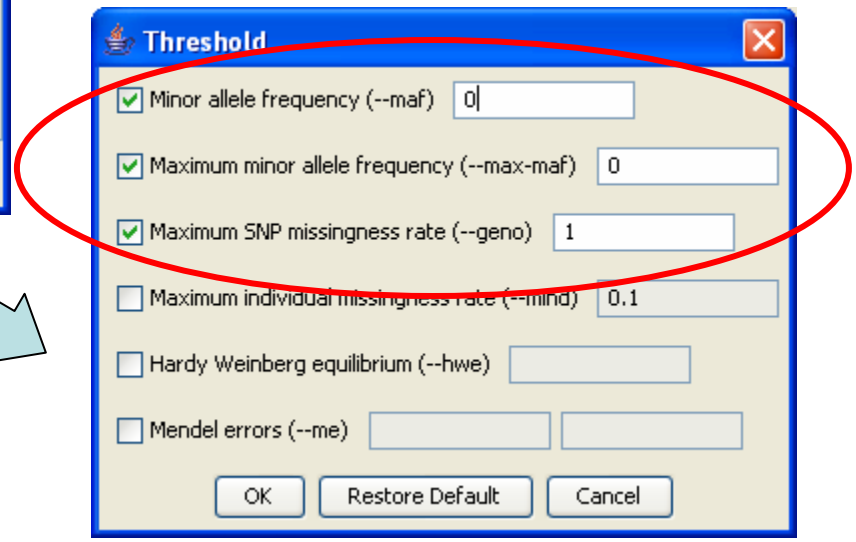

***Thresholds exclude people or SNPs based on genotype data***

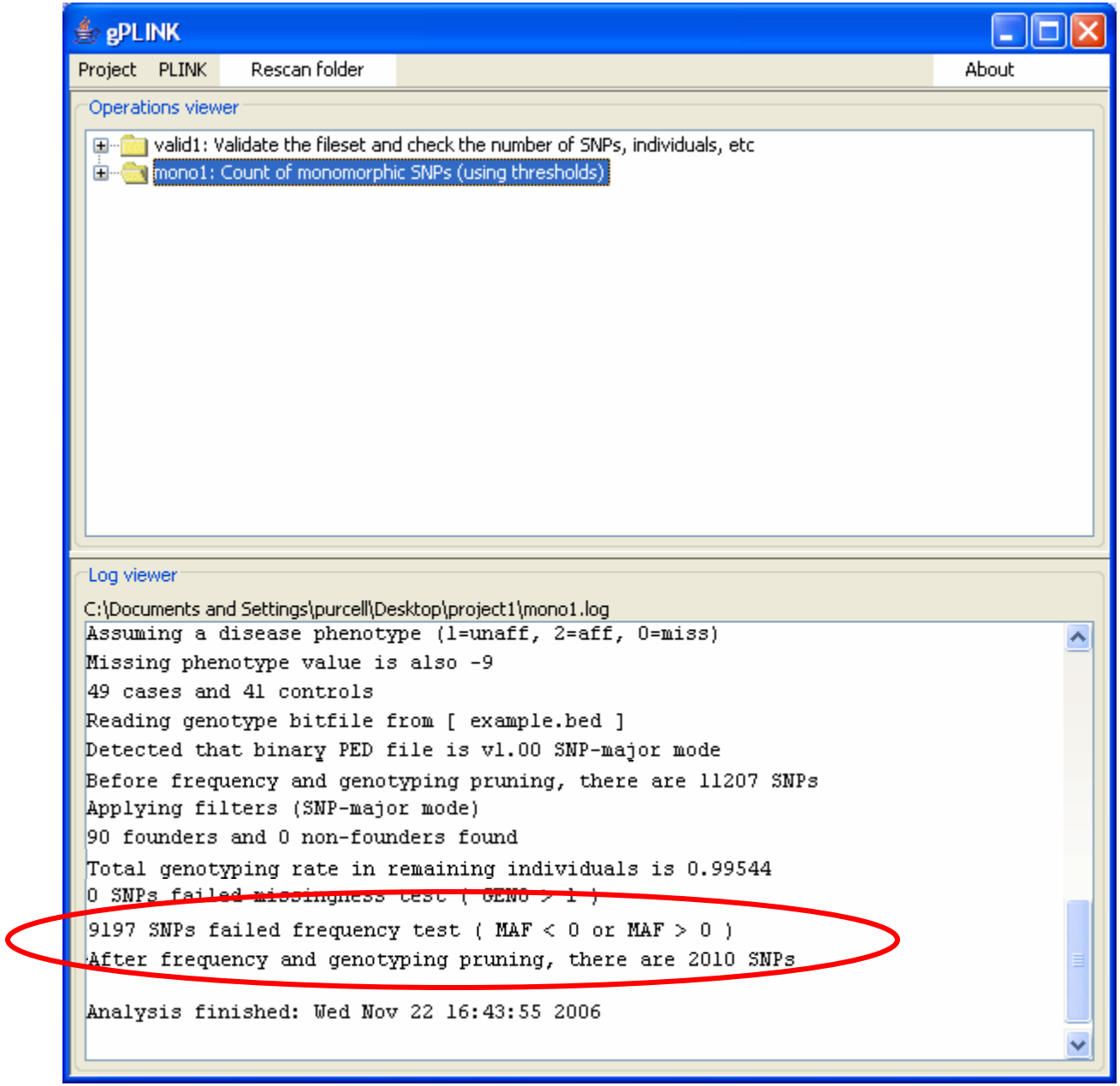


## *Q2) How many monomorphic SNPs?*

***We can use thresholds and the Validate fileset option to answer this:***

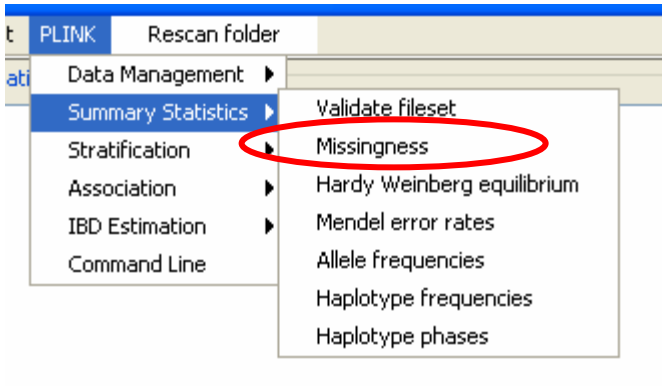


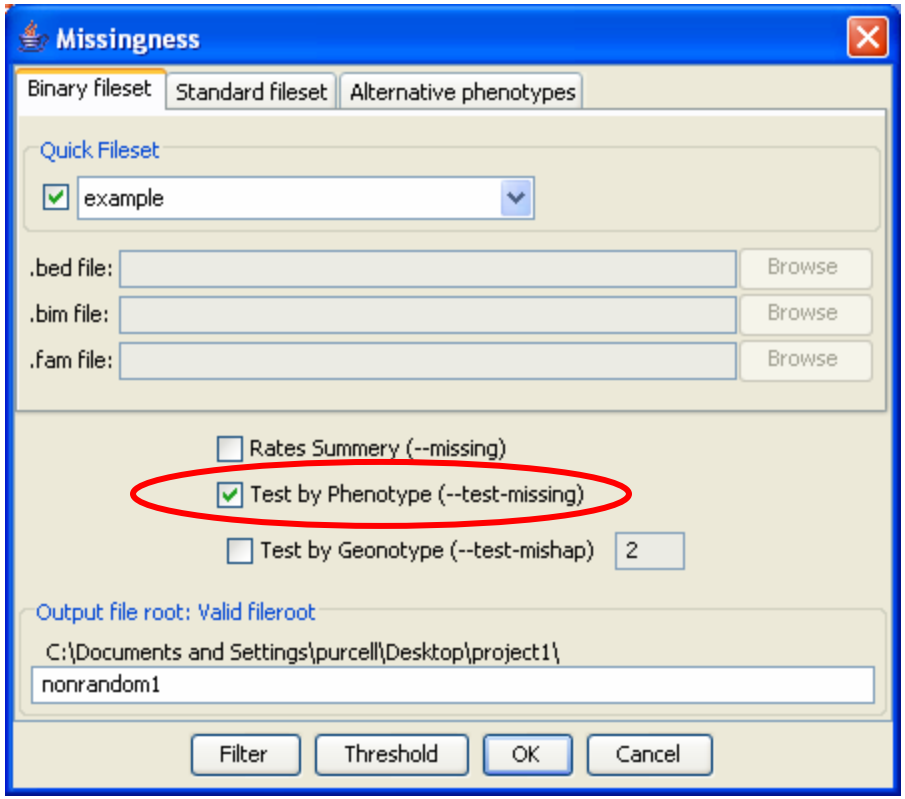



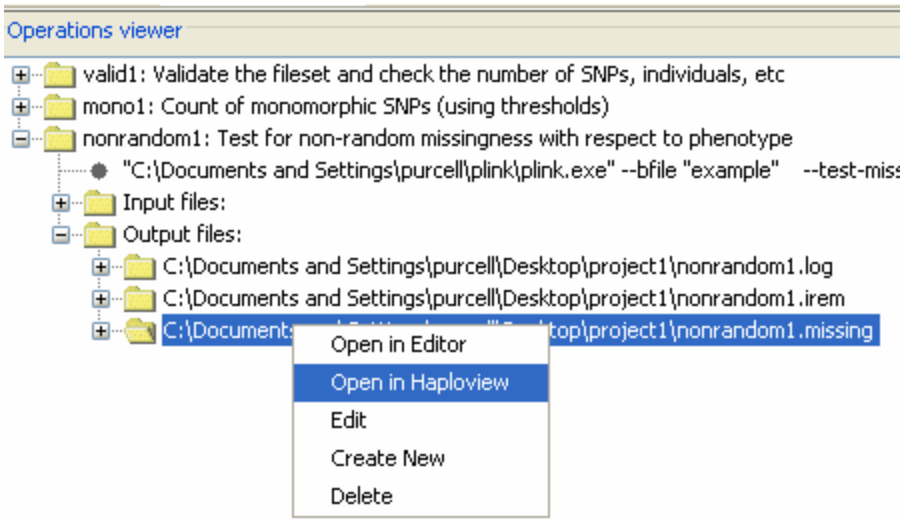

gPLINK
Project
PLINK
Rescan folder
About
Operations viewer
valid1: Validate the fileset and check the number of SNPs, individuals, etc
mono1: Count of monomorphic SNPs (using thresholds)
Log viewer
C:\Documents and Settings\purcell\Desktop\project1\mono1.log
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
49 cases and 41 controls
Reading genotype bitfile from [ example.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 11207 SNPs
Applying filters (SNP-major mode)
90 founders and 0 non-founders found
Total genotyping rate in remaining individuals is 0.99544
0 SNPs failed missingness test ( GENO > 1 )
9197 SNPs failed frequency test ( MAF < 0 or MAF > 0 )
After frequency and genotyping pruning, there are 2010 SNPs
Analysis finished: Wed Nov 22 16:43:55 2006

# Q3) Evidence of non-random genotyping failure?

## The Summary Statistics/Missingness option can answer this:



## *Missing rate in cases (A) and controls (U) and a test for whether rate differs*

Haploview 4.0beta11

File Display Analysis Help Key

PLINK

| Result | Chrom | Marker | Position | F_MISS_A | F_MISS_U | CHISQ_MISS | P_MISS |
|---|---|---|---|---|---|---|---|
| 1 | 1 | rs3094315 | 792429 | 0.02083 | 0.0 | 0.8639 | 0.3527 |
| 2 | 1 | rs4040617 | 819185 | 0.0 | 0.0 | 0.0 | 1.0 |
| 3 | 1 | rs4075116 | 1043552 | 0.0 | 0.0 | 0.0 | 1.0 |
| 4 | 1 | rs9442385 | 1137258 | 0.02083 | 0.0 | 0.8639 | 0.3527 |
| 5 | 1 | rs11260562 | 1205233 | 0.04167 | 0.0 | 1.748 | 0.1862 |
| 6 | 1 | rs6685064 | 1251215 | 0.0 | 0.0 | 0.0 | 1.0 |
| 7 | 1 | rs3766180 | 1563420 | 0.0 | 0.0 | 0.0 | 1.0 |
| 8 | 1 | rs6603791 | 1586208 | 0.0 | 0.0 | 0.0 | 1.0 |
| 9 | 1 | rs7519837 | 1596068 | 0.02083 | 0.0 | 0.8639 | 0.3527 |
| 10 | 1 | rs3737628 | 1755094 | 0.0 | 0.0 | 0.0 | 1.0 |
| 11 | 1 | rs7511905 | 1825948 | 0.0 | 0.0 | 0.0 | 1.0 |
| 12 | 1 | rs3855951 | 1836464 | 0.0 | 0.02439 | 1.184 | 0.2765 |
| 13 | 1 | rs6603803 | 1844850 | 0.02083 | 0.0 | 0.8639 | 0.3527 |
| 14 | 1 | rs2803285 | 1920531 | 0.0 | 0.0 | 0.0 | 1.0 |
| 15 | 1 | rs7513222 | 2060063 | 0.0 | 0.0 | 0.0 | 1.0 |
| 16 | 1 | rs3107146 | 2079746 | 0.0 | 0.0 | 0.0 | 1.0 |
| 17 | 1 | rs3107157 | 2094131 | 0.0 | 0.0 | 0.0 | 1.0 |
| 18 | 1 | rs3753242 | 2101843 | 0.0 | 0.0 | 0.0 | 1.0 |
| 19 | 1 | rs385039 | 2109571 | 0.0 | 0.0 | 0.0 | 1.0 |

Filters

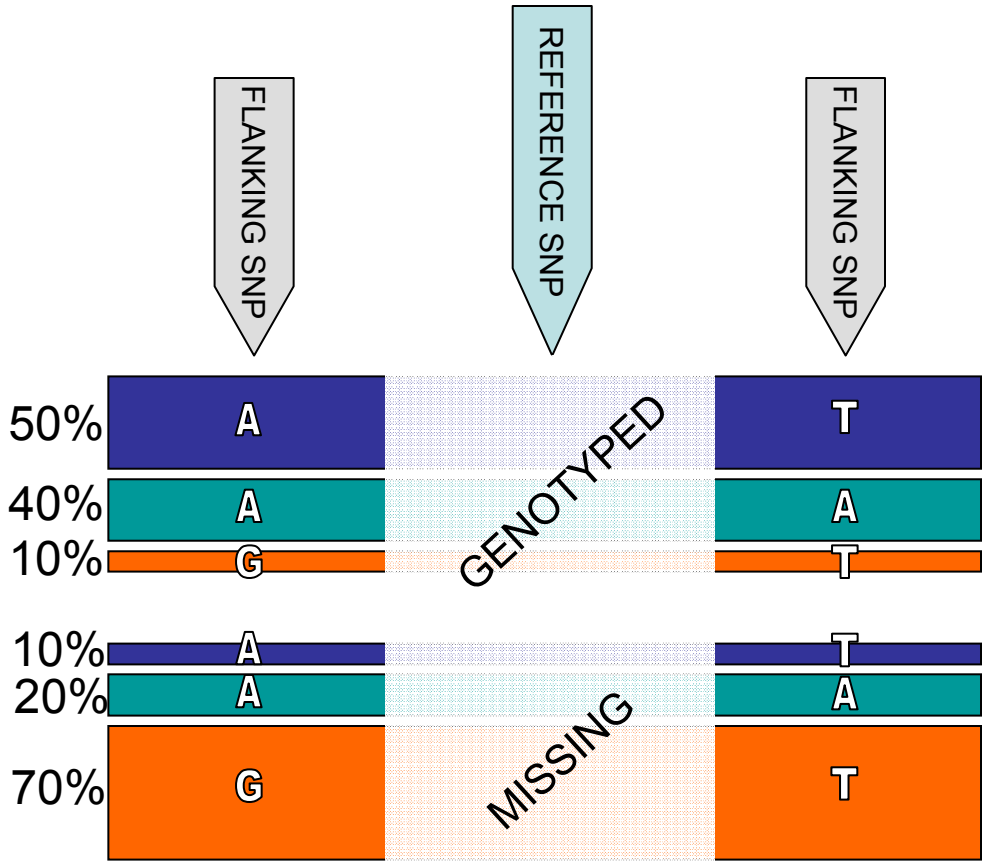
Chromosome: Start kb: End kb: Other: Filter

View top 100 results Go Marker: Go

Load Additional Results Reset Filters

Go to Selected Region

# Non-random genotyping failure




*"Mishap" test*

**~10% (30,824) of SNPs with >5 missing genotypes fail mishap test at *p* < 1e-8**

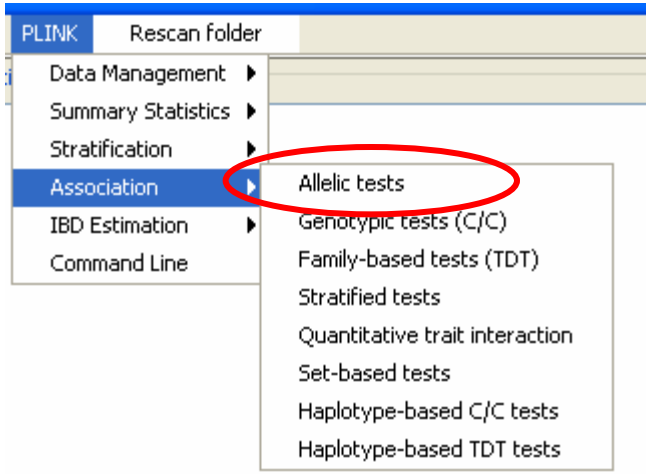
For example: rs7524558 has 68 missing genotypes (~2.6% missing)

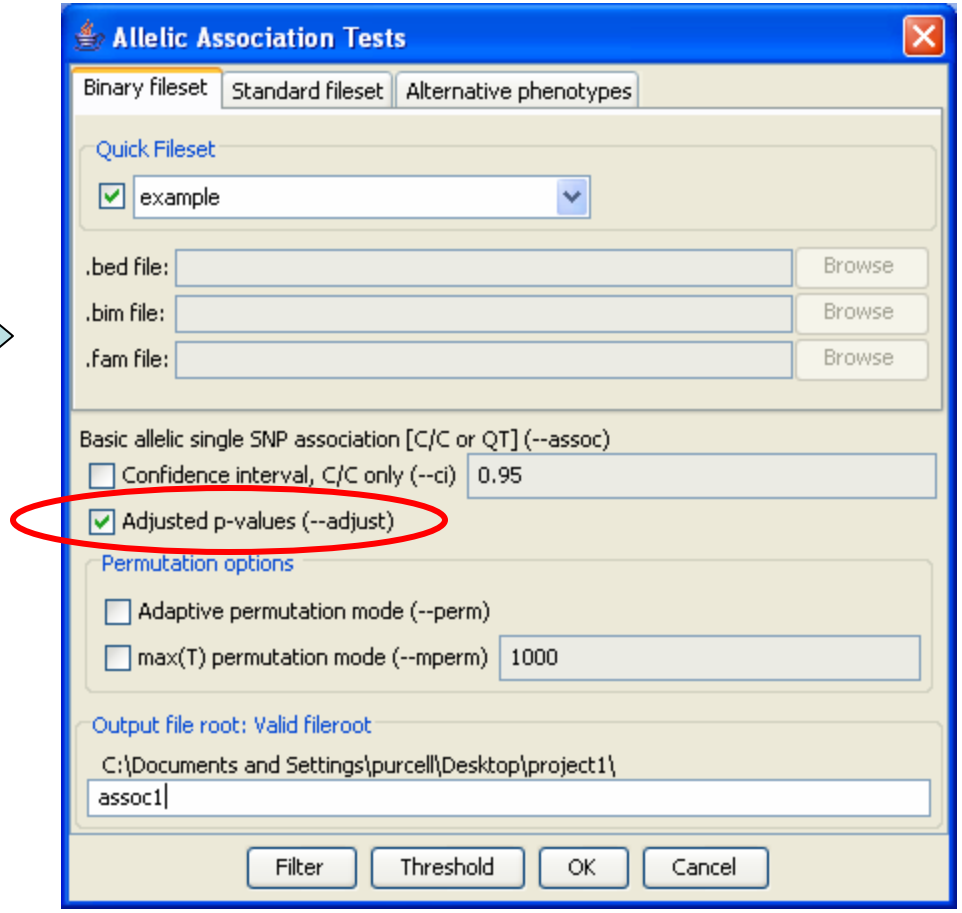
| Flanking haplotypes | GENO | MISSING |
|---|---|---|
| HOM | 2340 | 0 |
| HET | 49 | 68 |

# Association analysis

- Case/control
  - allelic, trend, genotypic
  - general Cochran-Mantel-Haenszel
- Family-based TDT
- Quantitative traits
- Haplotype analysis
  - focus on “multimarker predictors”
- Multilocus tests, covariates, epistasis, etc

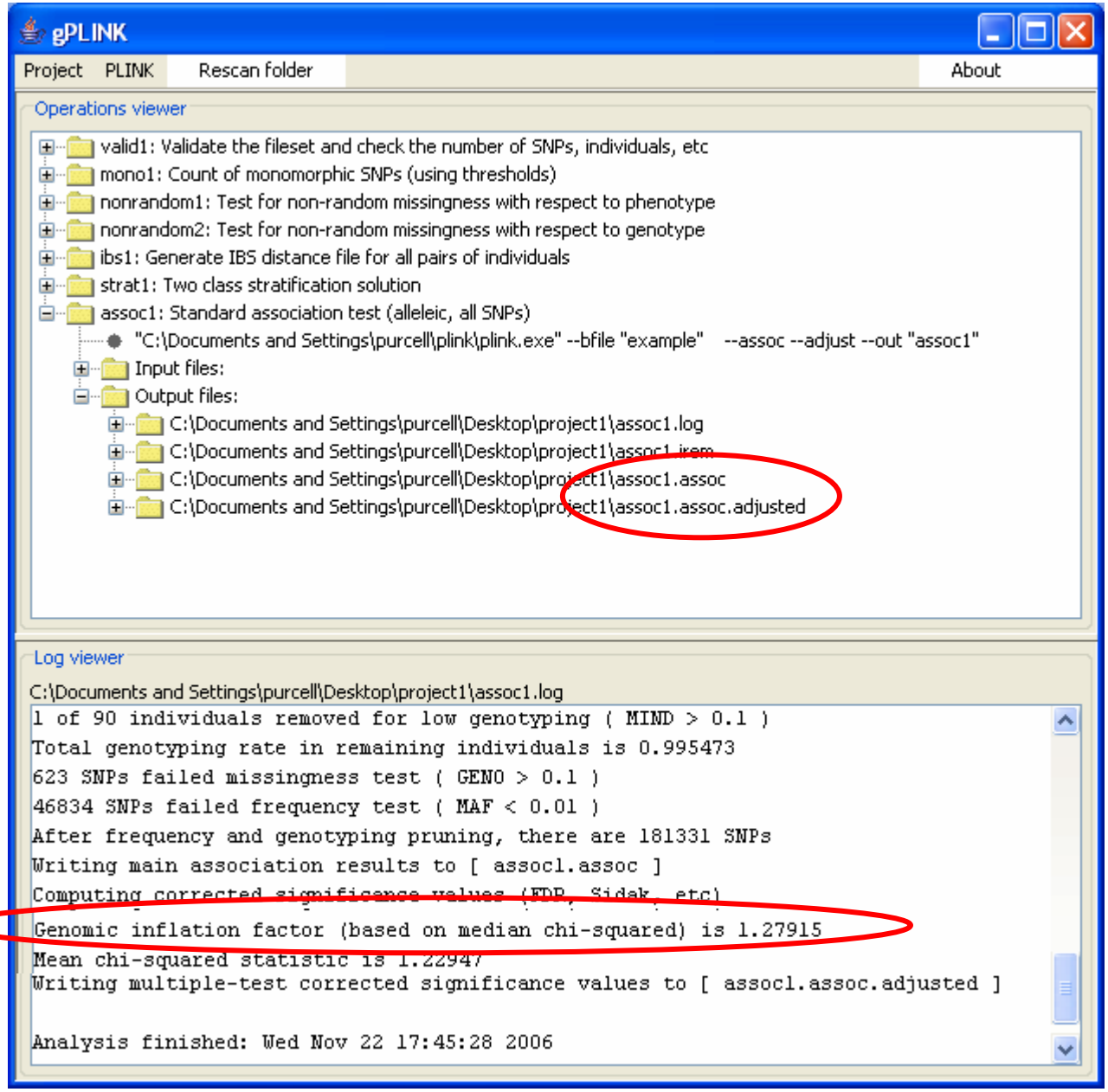
# Standard association tests



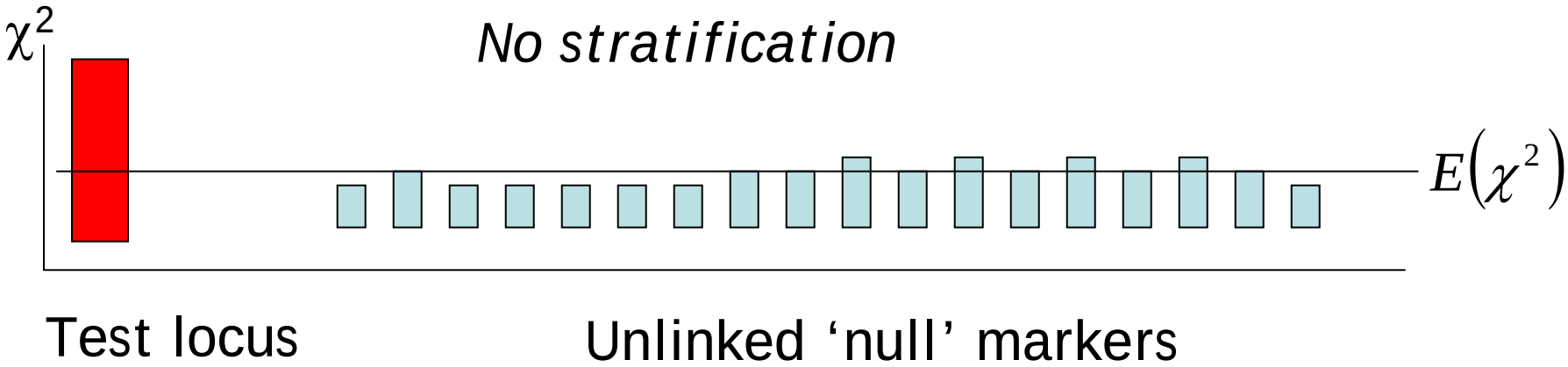



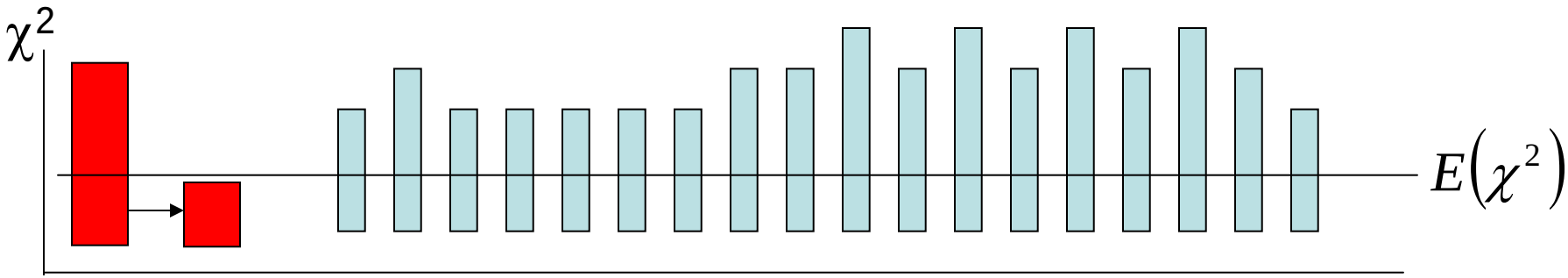

***Q4) What is the most associated SNP?***

# Q5) Evidence of stratification from genomic control?

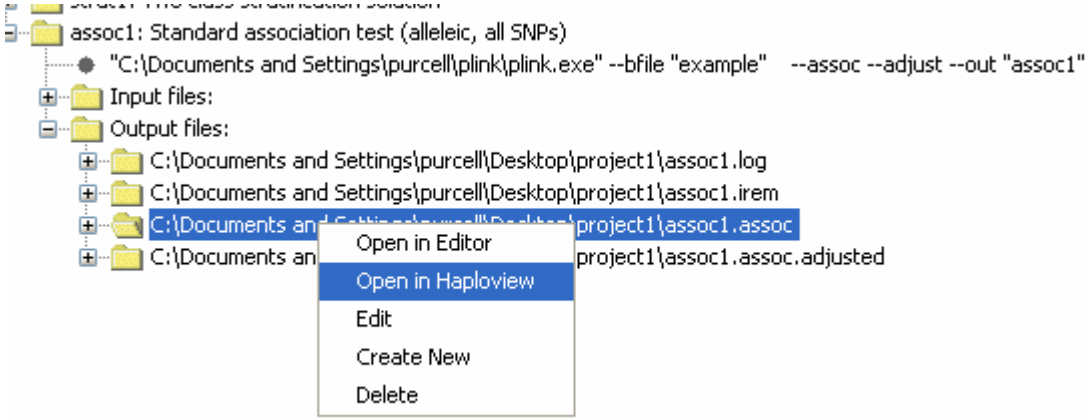


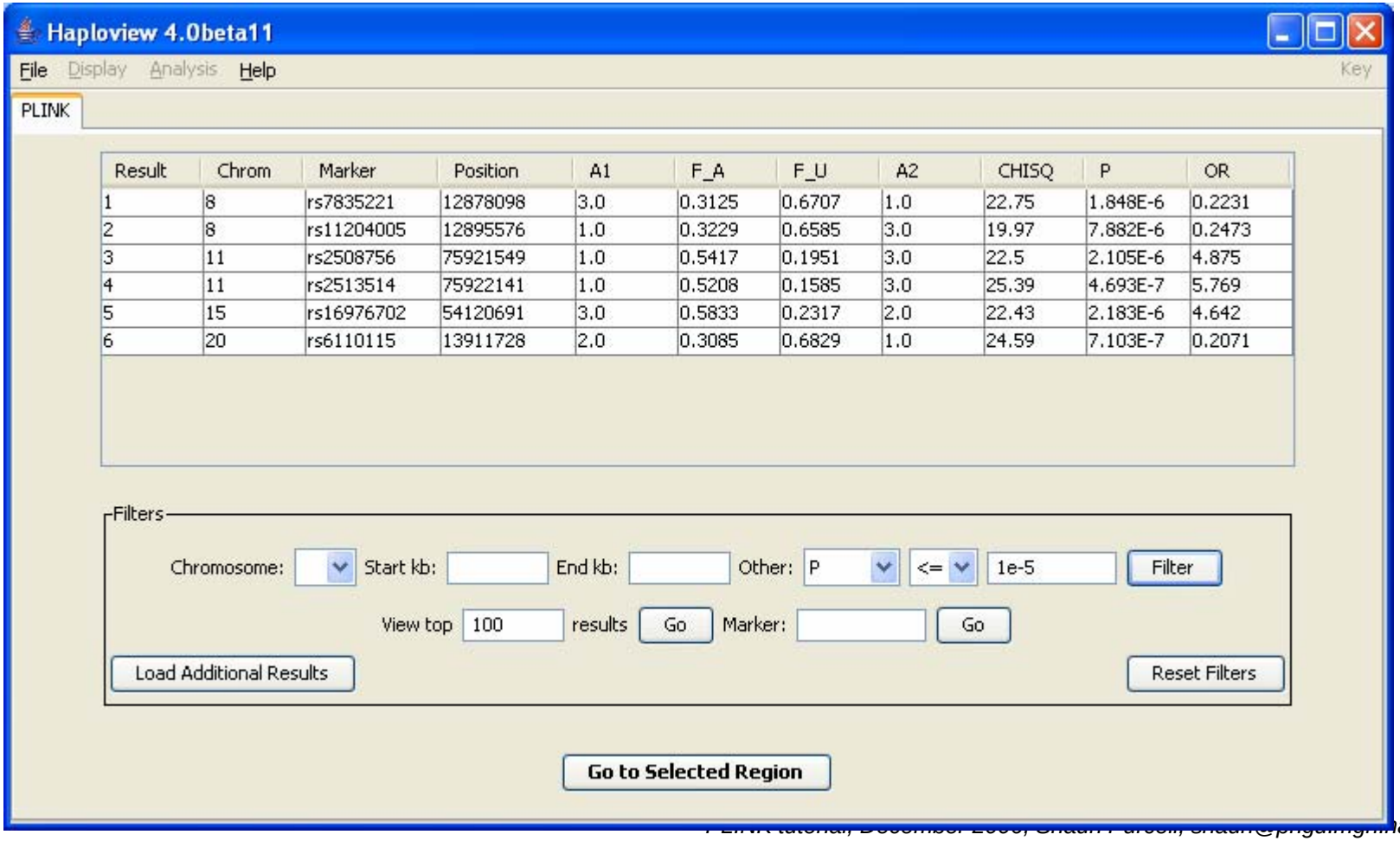
# Genomic control



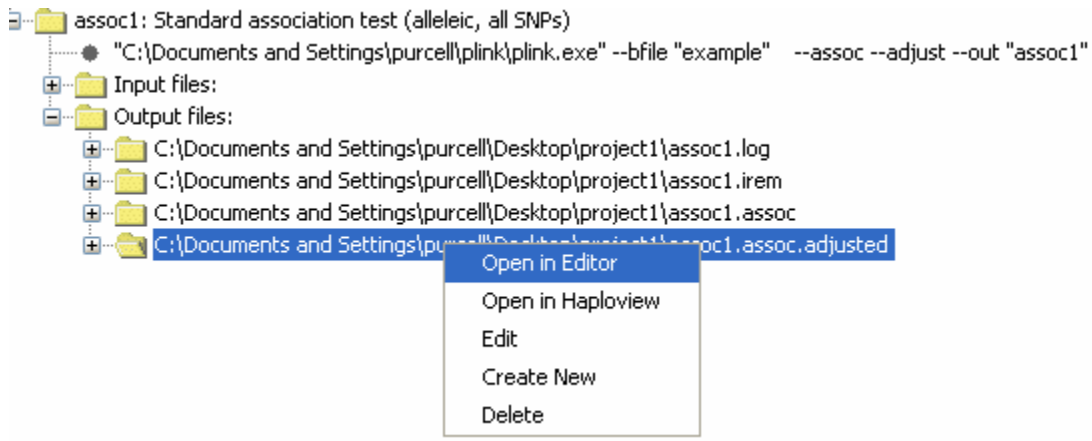



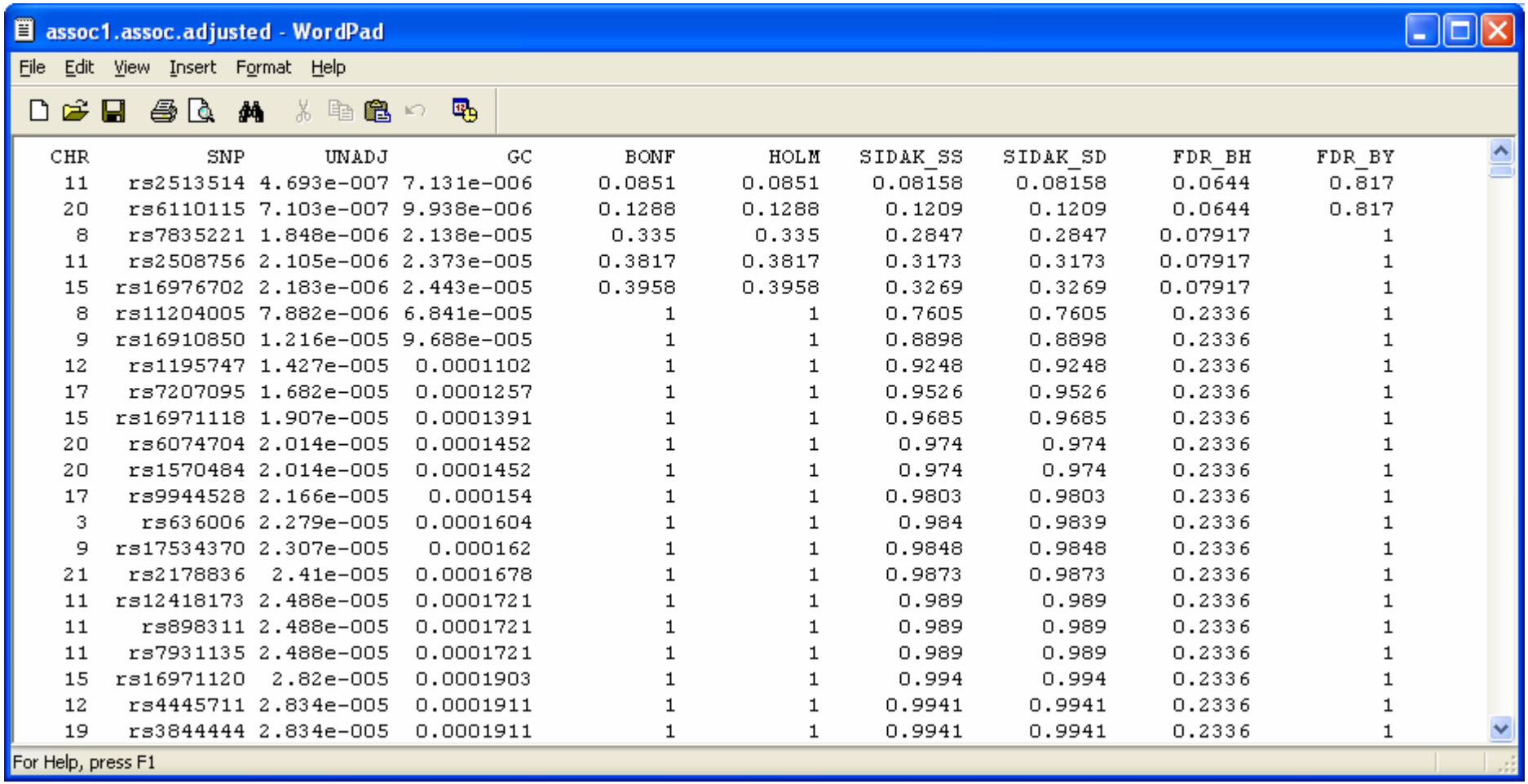

*Stratification → adjust test statistic*



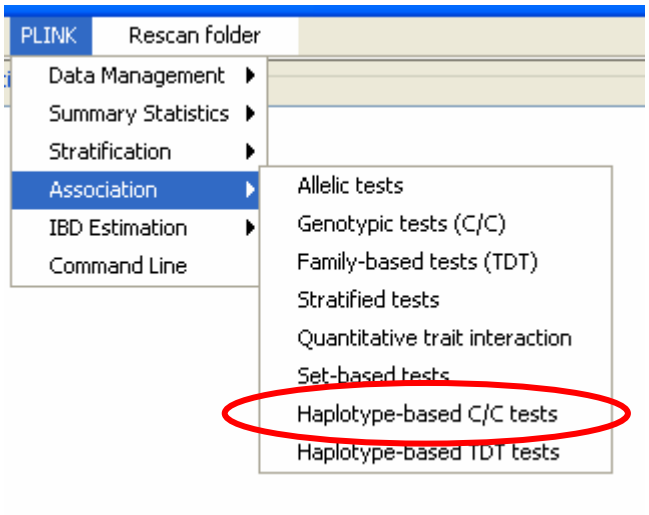

| Result | Chrom | Marker | Position | A1 | F_A | F_U | A2 | CHISQ | P | OR |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 8 | rs7835221 | 12878098 | 3.0 | 0.3125 | 0.6707 | 1.0 | 22.75 | 1.848E-6 | 0.2231 |
| 2 | 8 | rs11204005 | 12895576 | 1.0 | 0.3229 | 0.6585 | 3.0 | 19.97 | 7.882E-6 | 0.2473 |
| 3 | 11 | rs2508756 | 75921549 | 1.0 | 0.5417 | 0.1951 | 3.0 | 22.5 | 2.105E-6 | 4.875 |
| 4 | 11 | rs2513514 | 75922141 | 1.0 | 0.5208 | 0.1585 | 3.0 | 25.39 | 4.693E-7 | 5.769 |
| 5 | 15 | rs16976702 | 54120691 | 3.0 | 0.5833 | 0.2317 | 2.0 | 22.43 | 2.183E-6 | 4.642 |
| 6 | 20 | rs6110115 | 13911728 | 2.0 | 0.3085 | 0.6829 | 1.0 | 24.59 | 7.103E-7 | 0.2071 |

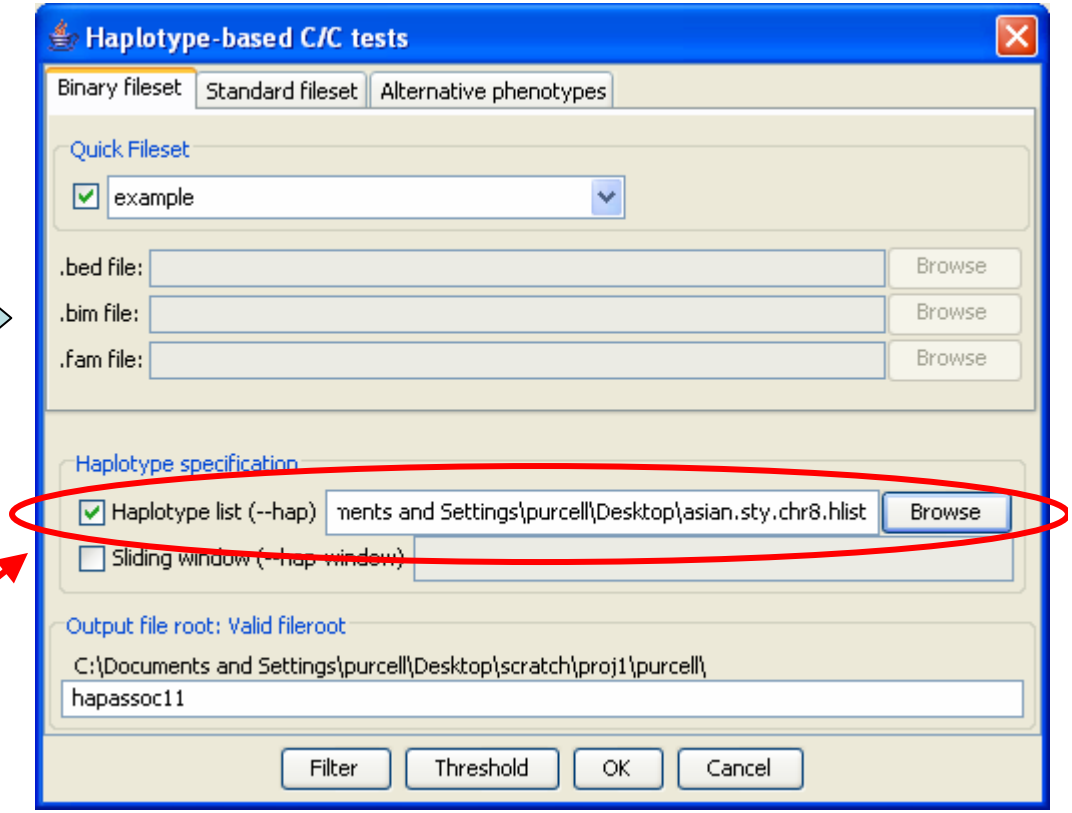



| CHR | SNP | UNADJ | GC | BONF | HOLM | SIDAK_SS | SIDAK_SD | FDR_BH | FDR_BY |
|---|---|---|---|---|---|---|---|---|---|
| 11 | rs2513514 | 4.693e-007 | 7.131e-006 | 0.0851 | 0.0851 | 0.08158 | 0.08158 | 0.0644 | 0.817 |
| 20 | rs6110115 | 7.103e-007 | 9.938e-006 | 0.1288 | 0.1288 | 0.1209 | 0.1209 | 0.0644 | 0.817 |
| 8 | rs7835221 | 1.848e-006 | 2.138e-005 | 0.335 | 0.335 | 0.2847 | 0.2847 | 0.07917 | 1 |
| 11 | rs2508756 | 2.105e-006 | 2.373e-005 | 0.3817 | 0.3817 | 0.3173 | 0.3173 | 0.07917 | 1 |
| 15 | rs16976702 | 2.183e-006 | 2.443e-005 | 0.3958 | 0.3958 | 0.3269 | 0.3269 | 0.07917 | 1 |
| 8 | rs11204005 | 7.882e-006 | 6.841e-005 | 1 | 1 | 0.7605 | 0.7605 | 0.2336 | 1 |
| 9 | rs16910850 | 1.216e-005 | 9.688e-005 | 1 | 1 | 0.8898 | 0.8898 | 0.2336 | 1 |
| 12 | rs1195747 | 1.427e-005 | 0.0001102 | 1 | 1 | 0.9248 | 0.9248 | 0.2336 | 1 |
| 17 | rs7207095 | 1.682e-005 | 0.0001257 | 1 | 1 | 0.9526 | 0.9526 | 0.2336 | 1 |
| 15 | rs16971118 | 1.907e-005 | 0.0001391 | 1 | 1 | 0.9685 | 0.9685 | 0.2336 | 1 |
| 20 | rs6074704 | 2.014e-005 | 0.0001452 | 1 | 1 | 0.974 | 0.974 | 0.2336 | 1 |
| 20 | rs1570484 | 2.014e-005 | 0.0001452 | 1 | 1 | 0.974 | 0.974 | 0.2336 | 1 |
| 17 | rs9944528 | 2.166e-005 | 0.000154 | 1 | 1 | 0.9803 | 0.9803 | 0.2336 | 1 |
| 3 | rs636006 | 2.279e-005 | 0.0001604 | 1 | 1 | 0.984 | 0.9839 | 0.2336 | 1 |
| 9 | rs17534370 | 2.307e-005 | 0.000162 | 1 | 1 | 0.9848 | 0.9848 | 0.2336 | 1 |
| 21 | rs2178836 | 2.41e-005 | 0.0001678 | 1 | 1 | 0.9873 | 0.9873 | 0.2336 | 1 |
| 11 | rs12418173 | 2.488e-005 | 0.0001721 | 1 | 1 | 0.989 | 0.989 | 0.2336 | 1 |
| 11 | rs898311 | 2.488e-005 | 0.0001721 | 1 | 1 | 0.989 | 0.989 | 0.2336 | 1 |
| 11 | rs7931135 | 2.488e-005 | 0.0001721 | 1 | 1 | 0.989 | 0.989 | 0.2336 | 1 |
| 15 | rs16971120 | 2.82e-005 | 0.0001903 | 1 | 1 | 0.994 | 0.994 | 0.2336 | 1 |
| 12 | rs4445711 | 2.834e-005 | 0.0001911 | 1 | 1 | 0.9941 | 0.9941 | 0.2336 | 1 |
| 19 | rs3844444 | 2.834e-005 | 0.0001911 | 1 | 1 | 0.9941 | 0.9941 | 0.2336 | 1 |

# Haplotype based association




***Specify a list of specific haplotype tests (*.hlist file)***

***Q4b) What is the most associated haplotype?***

# Specifying haplotype tests

***Specify specific haplotypes***

*Predicted*

| *ID* | *chr* | *cM* | *bp* | *alleles* |
|---|---|---|---|---|
| i_rs2906364 | 8 | 0 | 158484 | 1 2 |
| i_rs3750097 | 8 | 0 | 187042 | 1 2 |
| i_rs10105400 | 8 | 0 | 188546 | 1 2 |
| i_rs13258954 | 8 | 0 | 211039 | 1 2 |

… *etc* …

*Predictors*

| *Haplotype* | *SNPs (in data file)* | |
|---|---|---|
| 14 | rs7000519 | rs10488370 |
| 23 | rs2906334 | rs11988064 |
| 23 | rs2906334 | rs11988064 |
| 34 | rs13265571 | rs3008257 |

***Or, specify the locus (i.e. only specify predicting SNPs)***

```
* rs7000519 rs10488370
* rs2906334 rs11988064
* rs2906334 rs13265571 rs3008257
```

… *etc* …

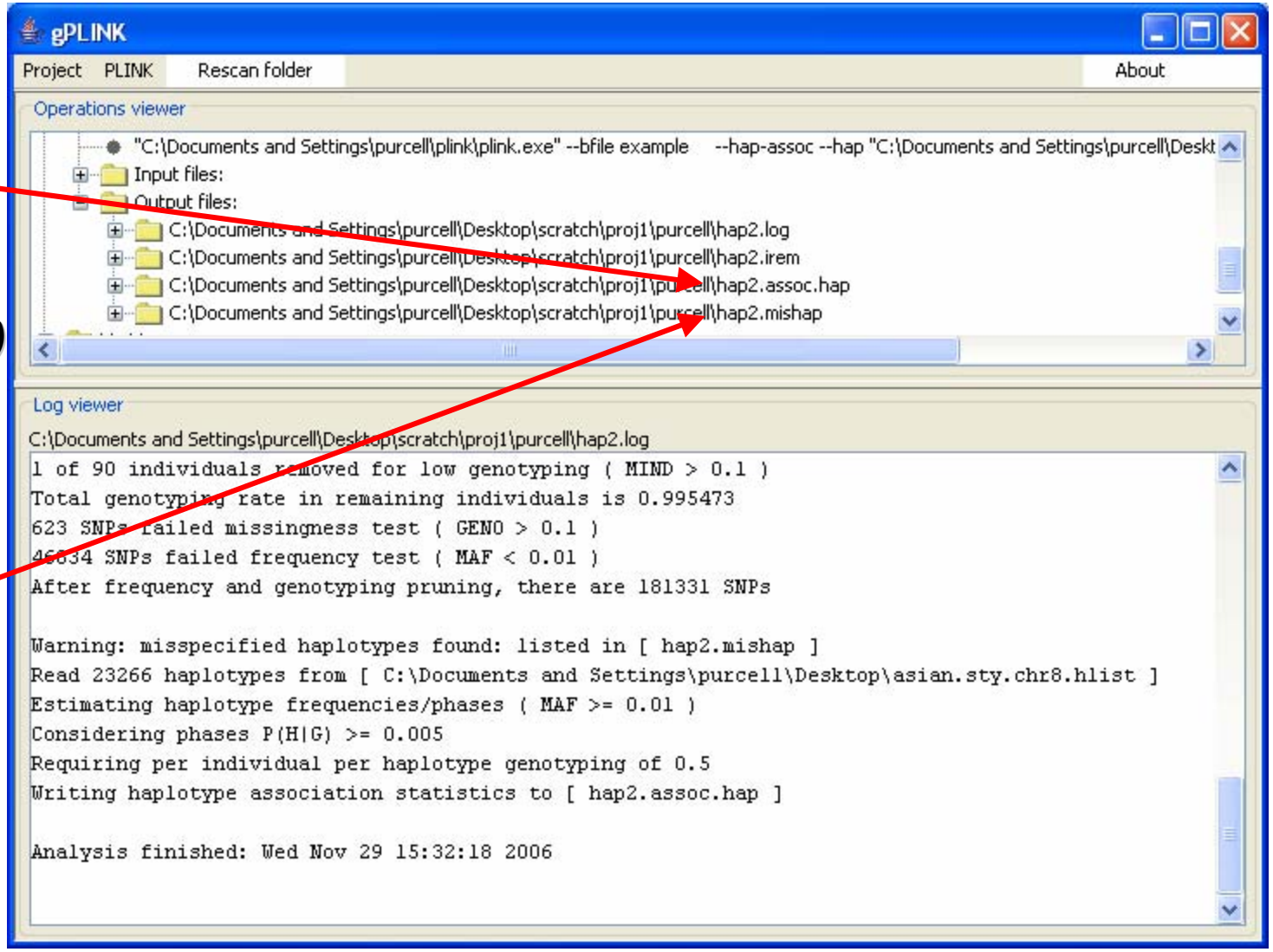
***Or, specifying a sliding window of fixed SNPs with:***

***e.g.*** **`--hap-window 4`**

# Haplotype-based tests

***Haplotype C/C association results (omnibus & haplotype-specific)***

***List of tests that could not be performed, e.g. if the predictor SNPs were removed in the filtering stage***



# Identity-by-state (IBS) sharing

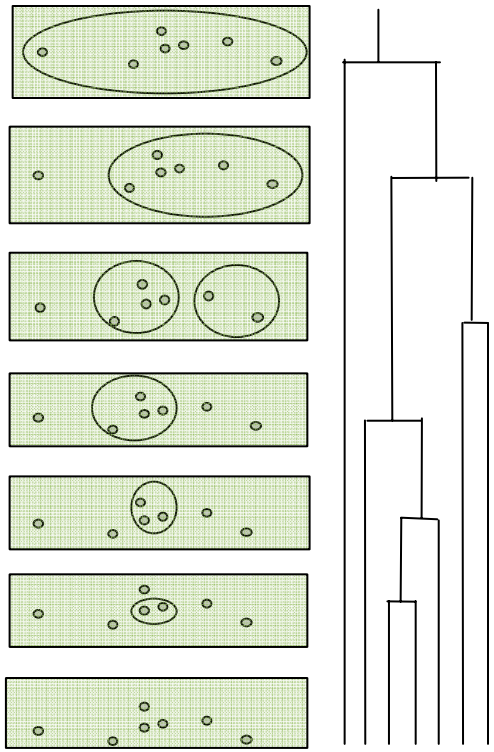
*Pair from same population*

| | | | | | |
|---|---|---|---|---|---|
| Individual 1 | A/**C** | G/**T** | **A**/**G** | A/A | **G**/**G** |
| Individual 2 | C/**C** | T/**T** | **A**/**G** | C/C | **G**/**G** |
| IBS | 1 | 1 | 2 | 0 | 2 |

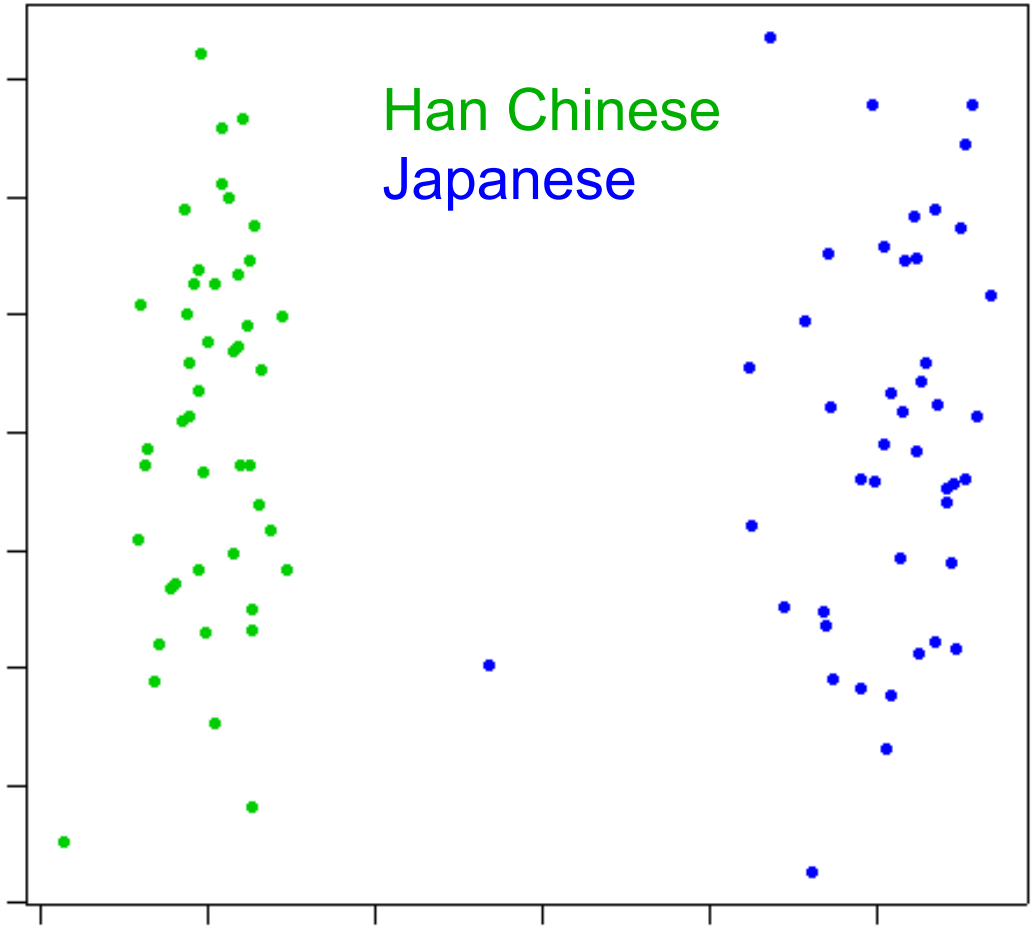
*Pair from different population*

| | | | | | |
|---|---|---|---|---|---|
| Individual 3 | A/**C** | G/G | A/A | A/A | G/**G** |
| Individual 4 | C/**C** | T/T | G/G | C/C | A/**G** |
| IBS | 1 | 0 | 0 | 0 | 1 |

# Empirical assessment of ancestry



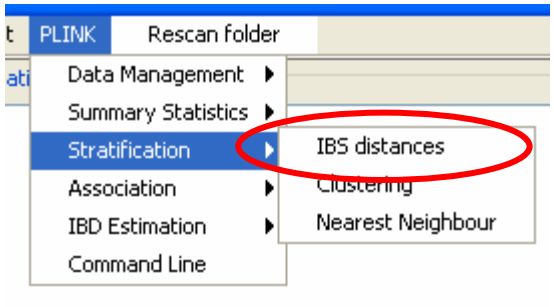
*Complete linkage IBS-based hierarchical clustering*

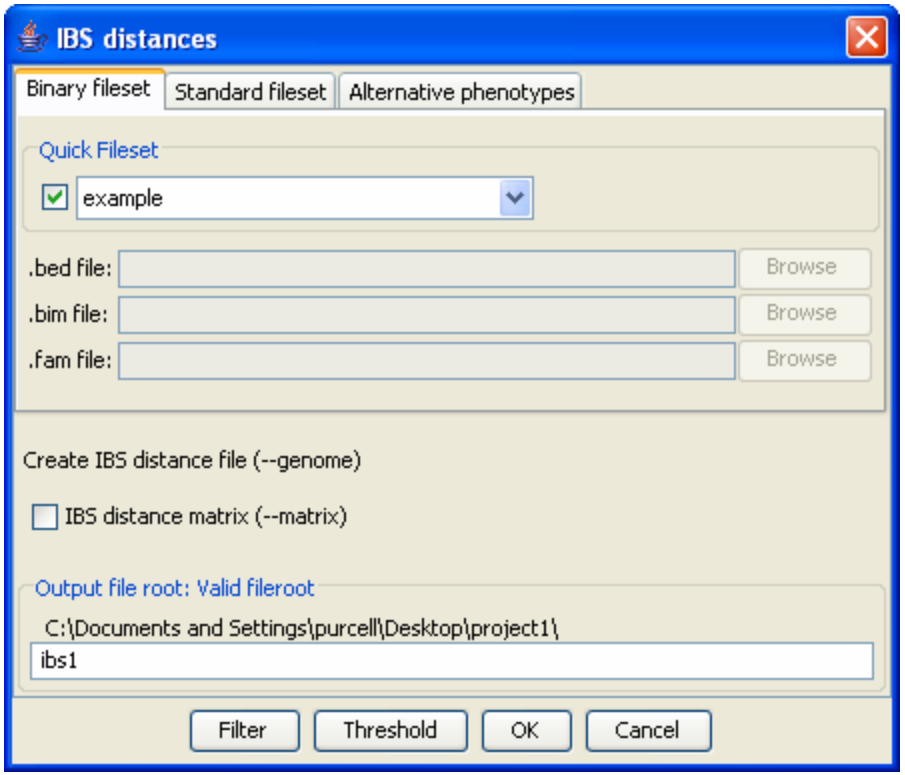


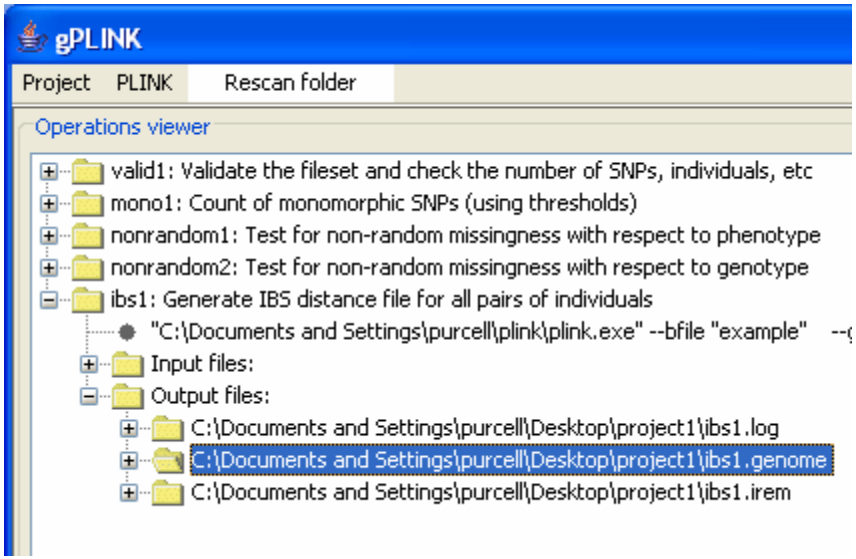

*Multidimensional scaling plot: ~10K random SNPs*

## *Q6) Use genotypes to cluster the sample into 2 subpopulations*

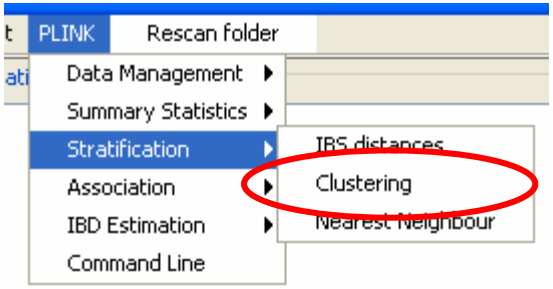
### *Step 1) Generate IBS distances for all pairs (may take a few minutes)*

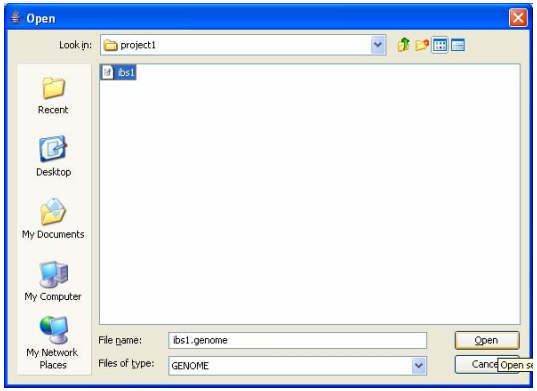


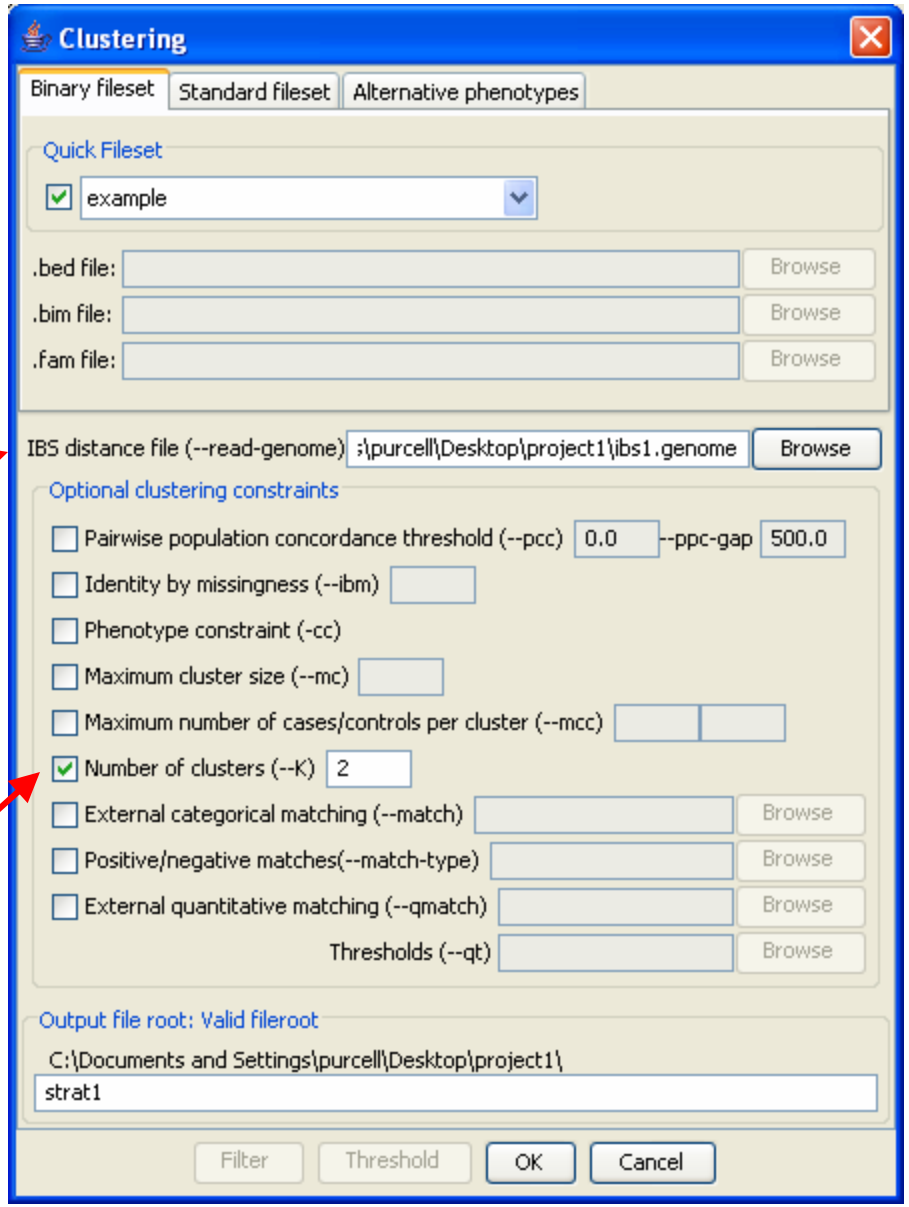



## *Step 2) Cluster individuals based on IBS distances and other constraints*

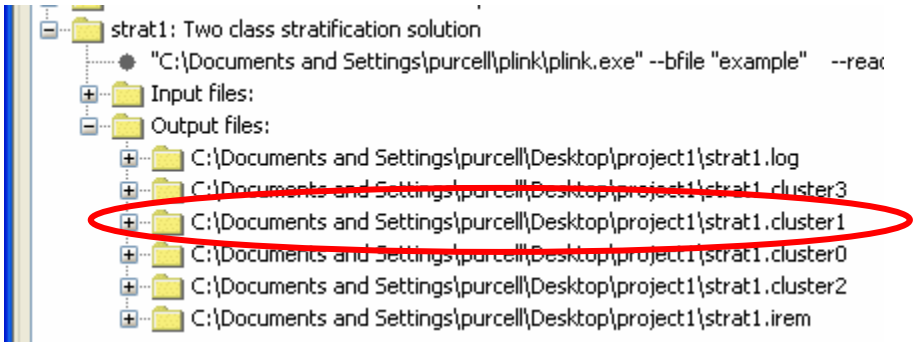


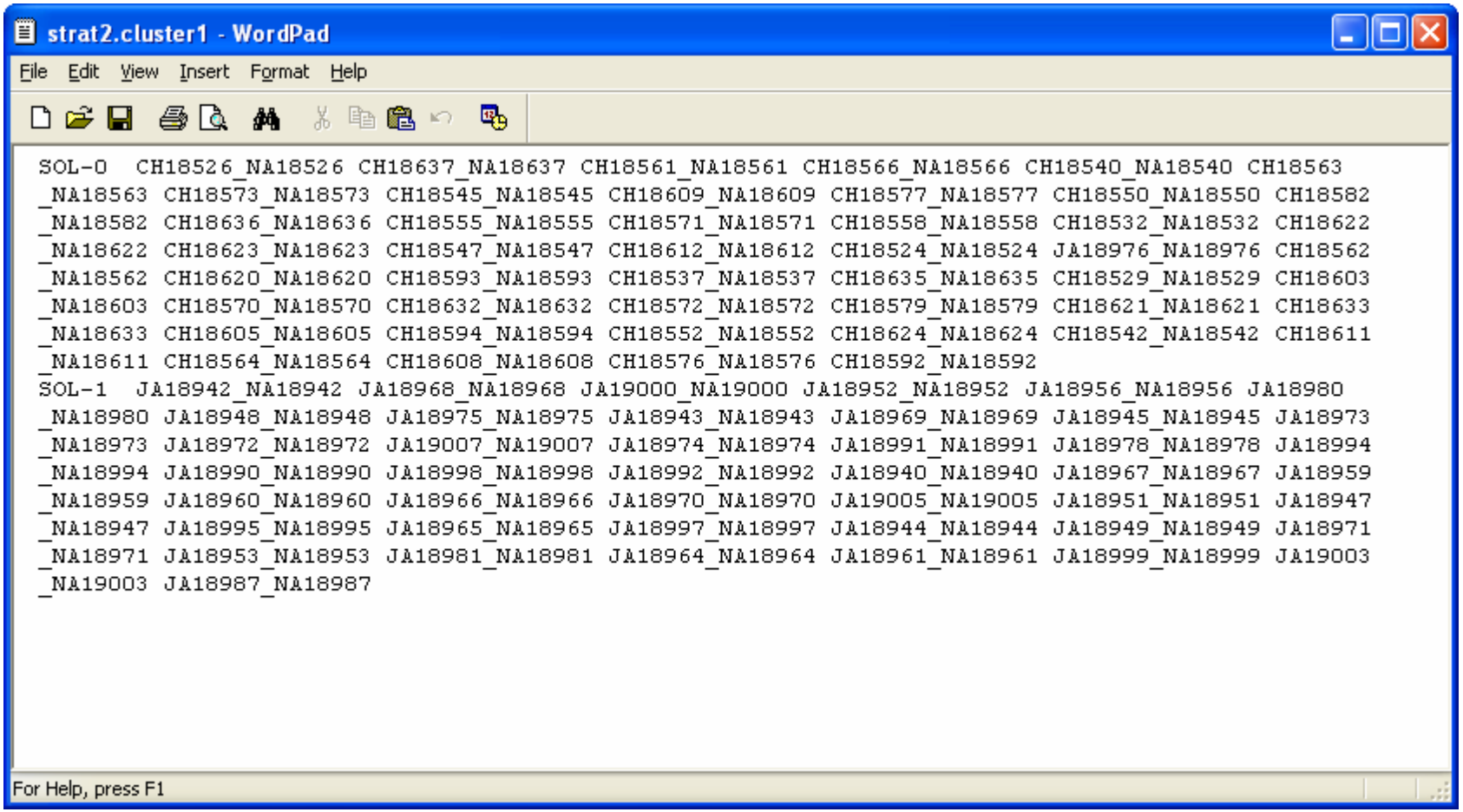



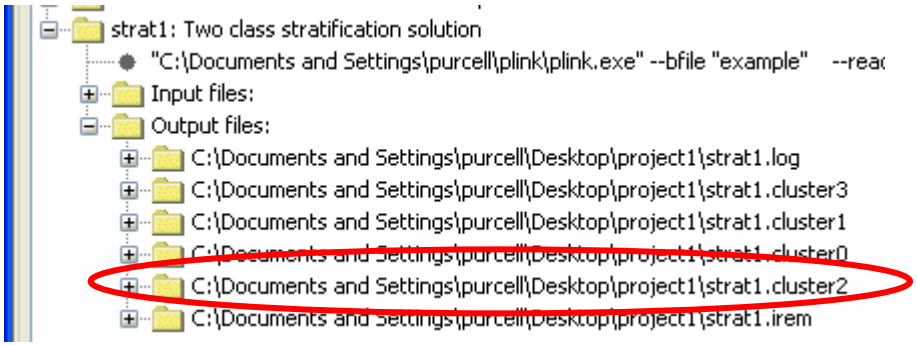

***Specify previously-generated IBS file (*.genome)***

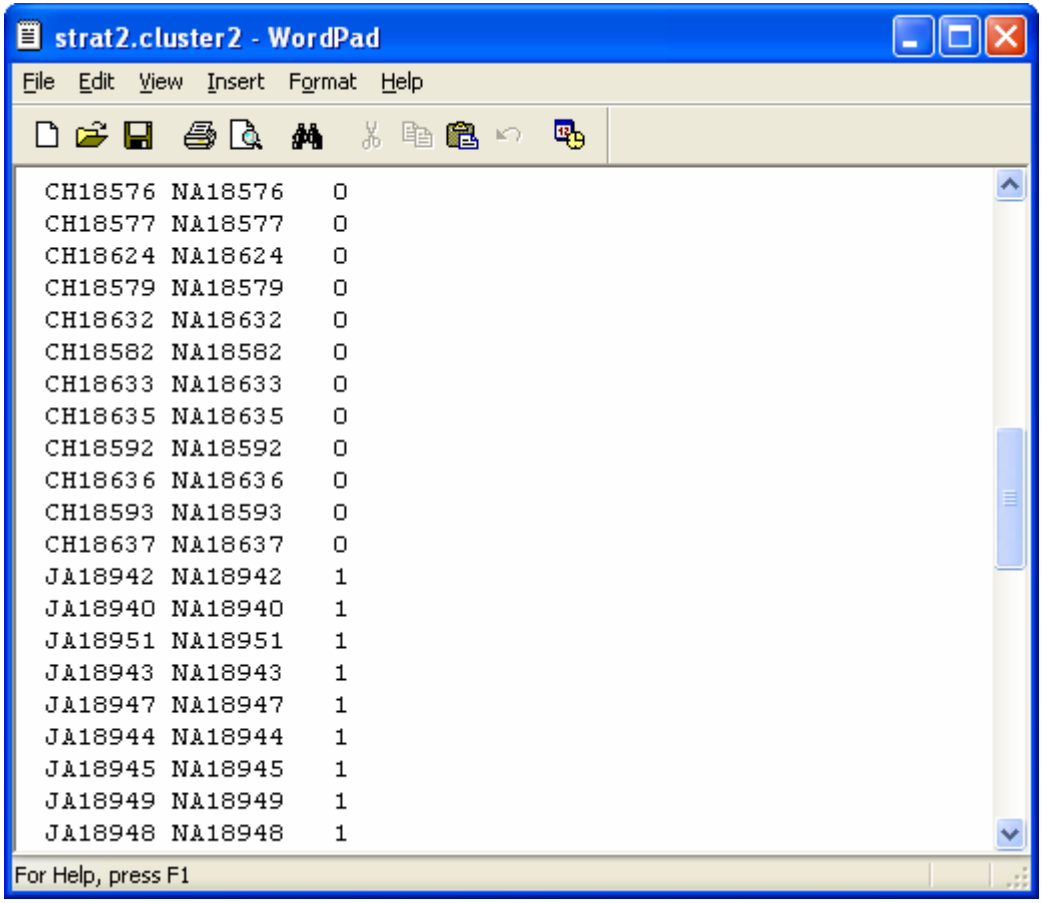



***Constrain cluster solution to two classes (K=2)***


strat1: Two class stratification solution
"C:\Documents and Settings\purcell\plink\plink.exe" --bfile "example" --reac
Input files:
Output files:
C:\Documents and Settings\purcell\Desktop\project1\strat1.log
C:\Documents and Settings\purcell\Desktop\project1\strat1.cluster3
C:\Documents and Settings\purcell\Desktop\project1\strat1.cluster1
C:\Documents and Settings\purcell\Desktop\project1\strat1.cluster0
C:\Documents and Settings\purcell\Desktop\project1\strat1.cluster2
C:\Documents and Settings\purcell\Desktop\project1\strat1.irem

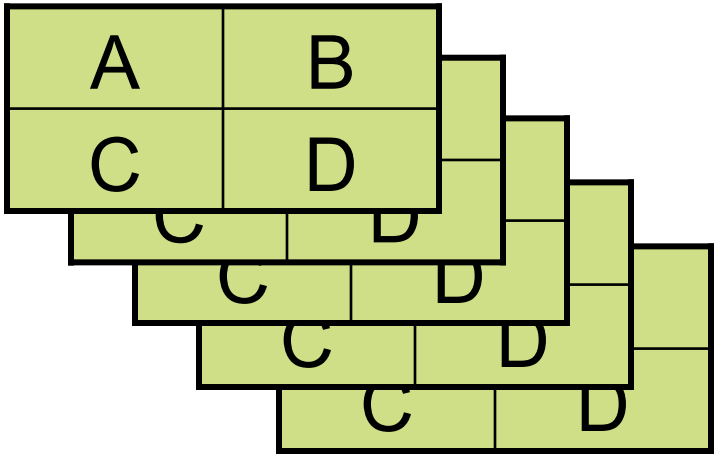

strat2.cluster1 - WordPad
File Edit View Insert Format Help
SOL-0 CH18526_NA18526 CH18637_NA18637 CH18561_NA18561 CH18566_NA18566 CH18540_NA18540 CH18563
_NA18563 CH18573_NA18573 CH18545_NA18545 CH18609_NA18609 CH18577_NA18577 CH18550_NA18550 CH18582
_NA18582 CH18636_NA18636 CH18555_NA18555 CH18571_NA18571 CH18558_NA18558 CH18532_NA18532 CH18622
_NA18622 CH18623_NA18623 CH18547_NA18547 CH18612_NA18612 CH18524_NA18524 JA18976_NA18976 CH18562
_NA18562 CH18620_NA18620 CH18593_NA18593 CH18537_NA18537 CH18635_NA18635 CH18529_NA18529 CH18603
_NA18603 CH18570_NA18570 CH18632_NA18632 CH18572_NA18572 CH18579_NA18579 CH18621_NA18621 CH18633
_NA18633 CH18605_NA18605 CH18594_NA18594 CH18552_NA18552 CH18624_NA18624 CH18542_NA18542 CH18611
_NA18611 CH18564_NA18564 CH18608_NA18608 CH18576_NA18576 CH18592_NA18592
SOL-1 JA18942_NA18942 JA18968_NA18968 JA19000_NA19000 JA18952_NA18952 JA18956_NA18956 JA18980
_NA18980 JA18948_NA18948 JA18975_NA18975 JA18943_NA18943 JA18969_NA18969 JA18945_NA18945 JA18973
_NA18973 JA18972_NA18972 JA19007_NA19007 JA18974_NA18974 JA18991_NA18991 JA18978_NA18978 JA18994
_NA18994 JA18990_NA18990 JA18998_NA18998 JA18992_NA18992 JA18940_NA18940 JA18967_NA18967 JA18959
_NA18959 JA18960_NA18960 JA18966_NA18966 JA18970_NA18970 JA19005_NA19005 JA18951_NA18951 JA18947
_NA18947 JA18995_NA18995 JA18965_NA18965 JA18997_NA18997 JA18944_NA18944 JA18949_NA18949 JA18971
_NA18971 JA18953_NA18953 JA18981_NA18981 JA18964_NA18964 JA18961_NA18961 JA18999_NA18999 JA19003
_NA19003 JA18987_NA18987
For Help, press F1


strat1: Two class stratification solution
"C:\Documents and Settings\purcell\plink\plink.exe" --bfile "example" --rea
Input files:
Output files:
C:\Documents and Settings\purcell\Desktop\project1\strat1.log
C:\Documents and Settings\purcell\Desktop\project1\strat1.cluster3
C:\Documents and Settings\purcell\Desktop\project1\strat1.cluster1
C:\Documents and Settings\purcell\Desktop\project1\strat1.cluster0
C:\Documents and Settings\purcell\Desktop\project1\strat1.cluster2
C:\Documents and Settings\purcell\Desktop\project1\strat1.irem

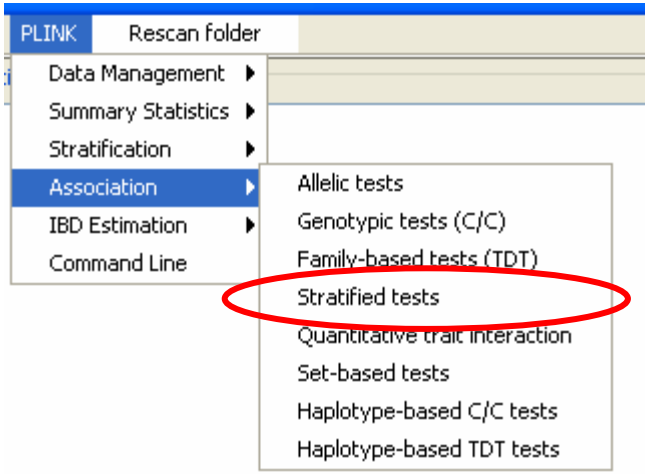

strat2.cluster2 - WordPad
File Edit View Insert Format Help
CH18576 NA18576 0
CH18577 NA18577 0
CH18624 NA18624 0
CH18579 NA18579 0
CH18632 NA18632 0
CH18582 NA18582 0
CH18633 NA18633 0
CH18635 NA18635 0
CH18592 NA18592 0
CH18636 NA18636 0
CH18593 NA18593 0
CH18637 NA18637 0
JA18942 NA18942 1
JA18940 NA18940 1
JA18951 NA18951 1
JA18943 NA18943 1
JA18947 NA18947 1
JA18944 NA18944 1
JA18945 NA18945 1
JA18949 NA18949 1
JA18948 NA18948 1
For Help, press F1

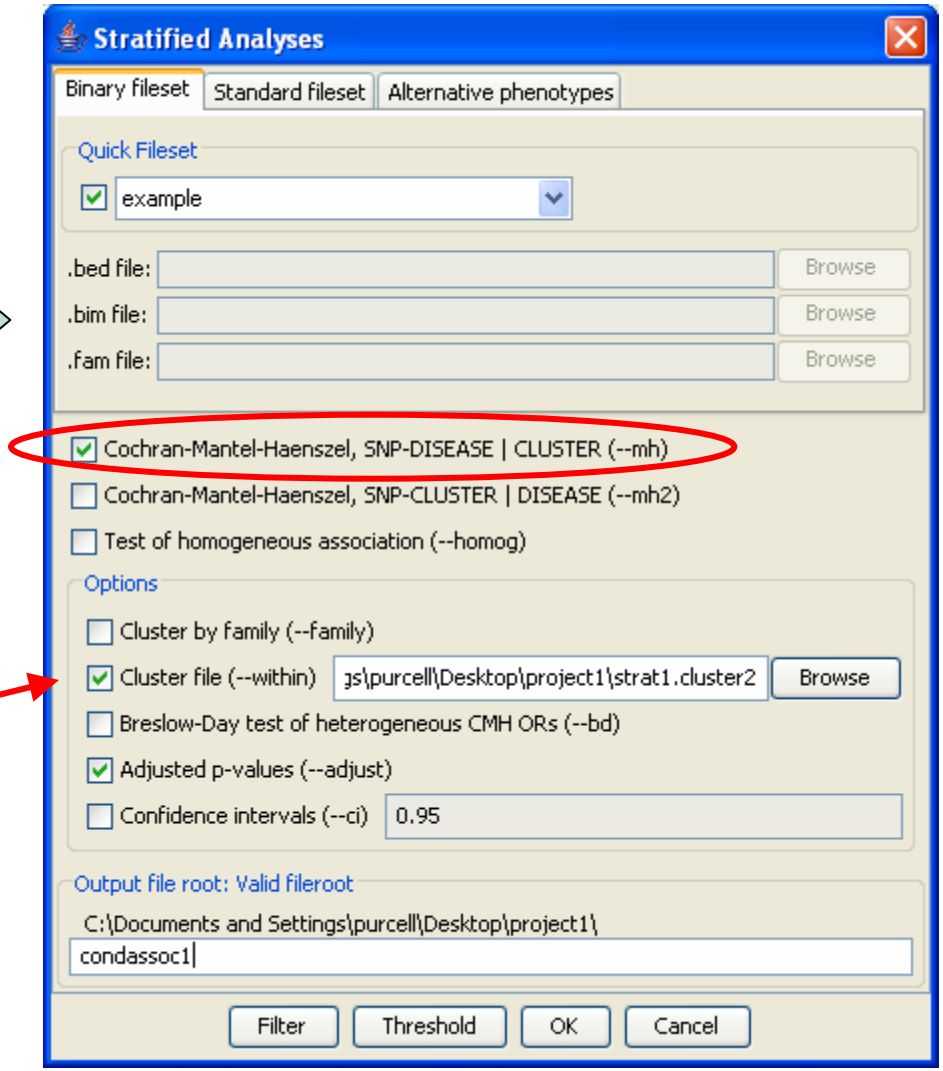
# Stratified analysis

- Cochran-Mantel-Haenszel test



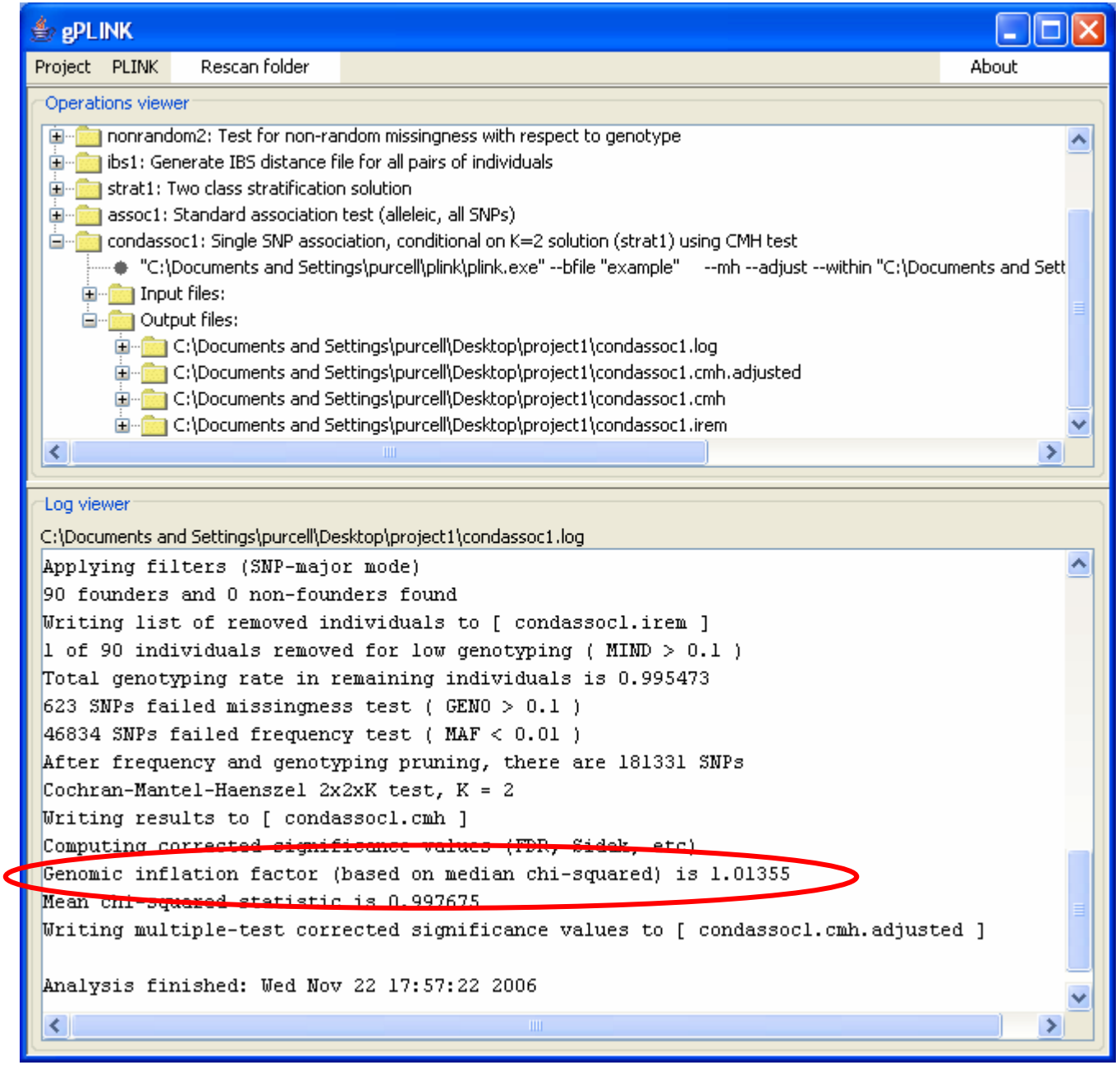

- Stratified 2×2×K tables



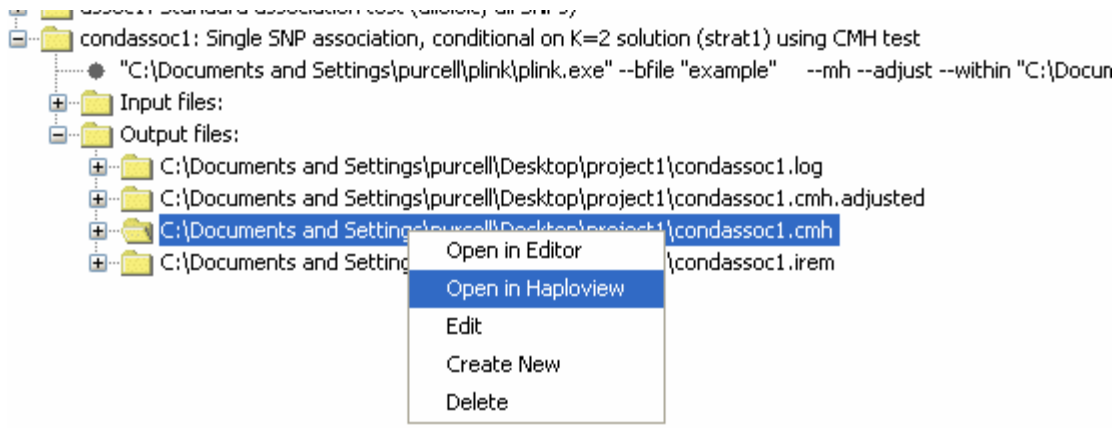



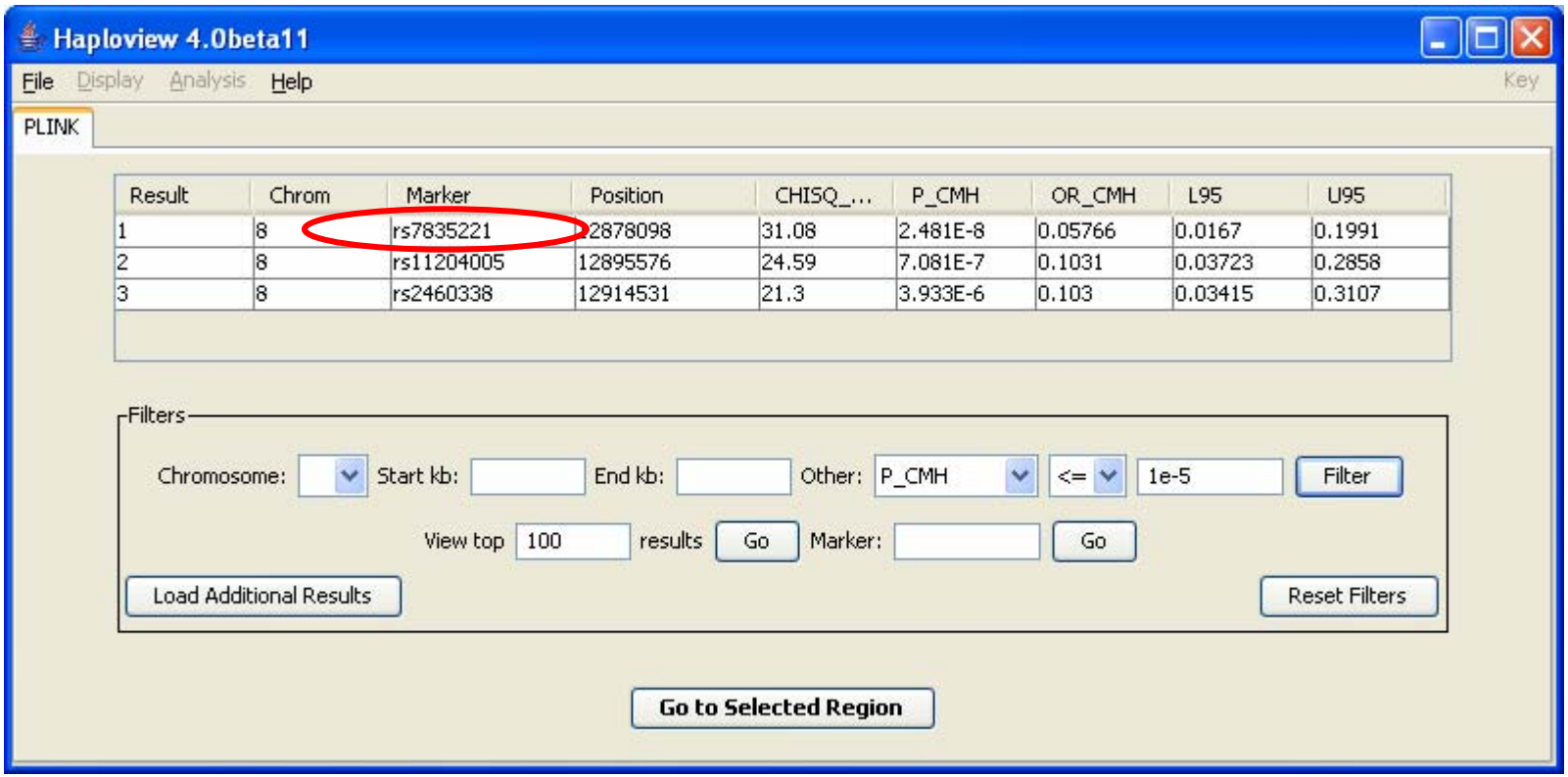

***Select the previously calculated *.cluster2 file. This "cluster file" has one line per individual***

# Q7) Evidence of stratification conditional on cluster solution?

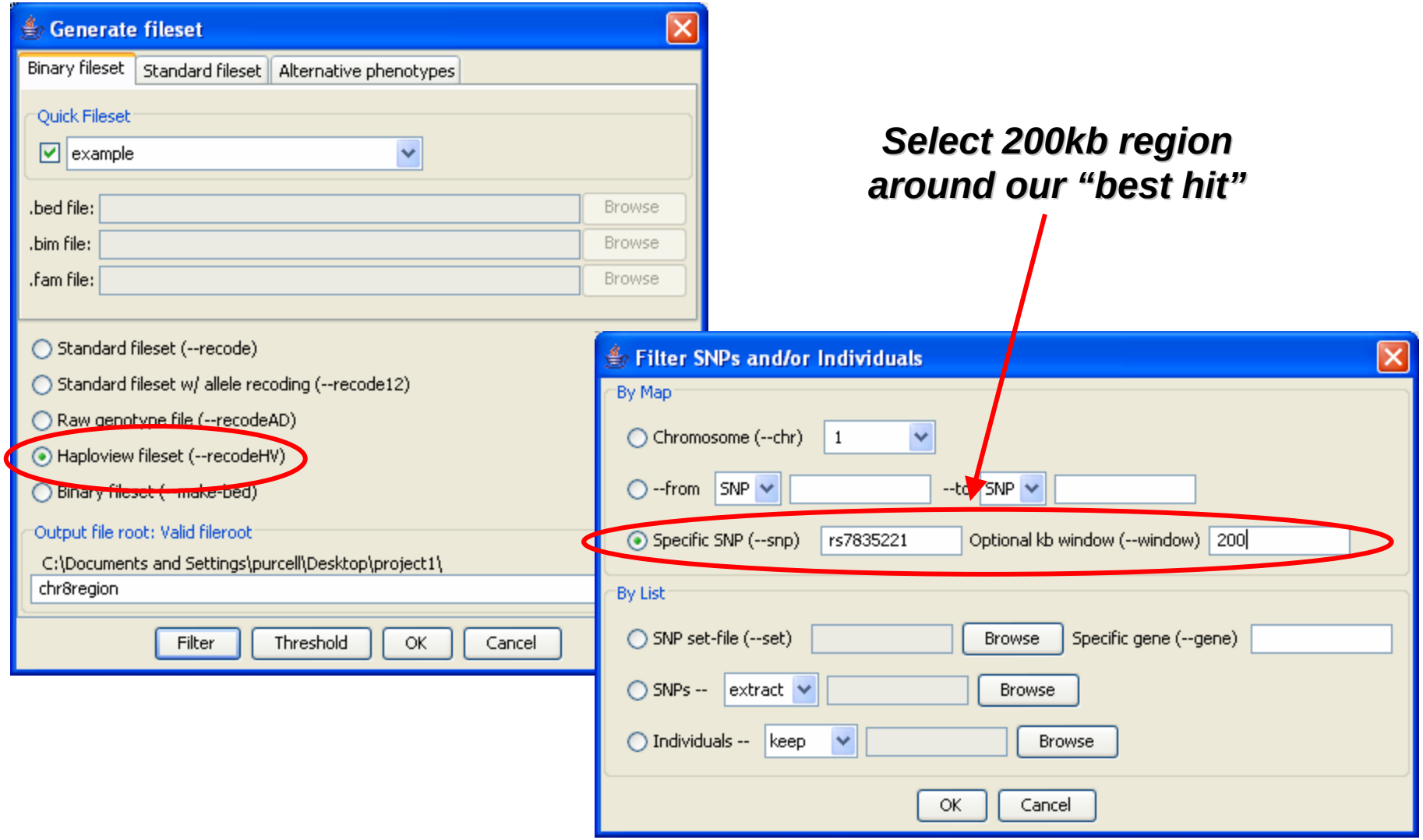


# Q8) What is the best SNP controlling for stratification?

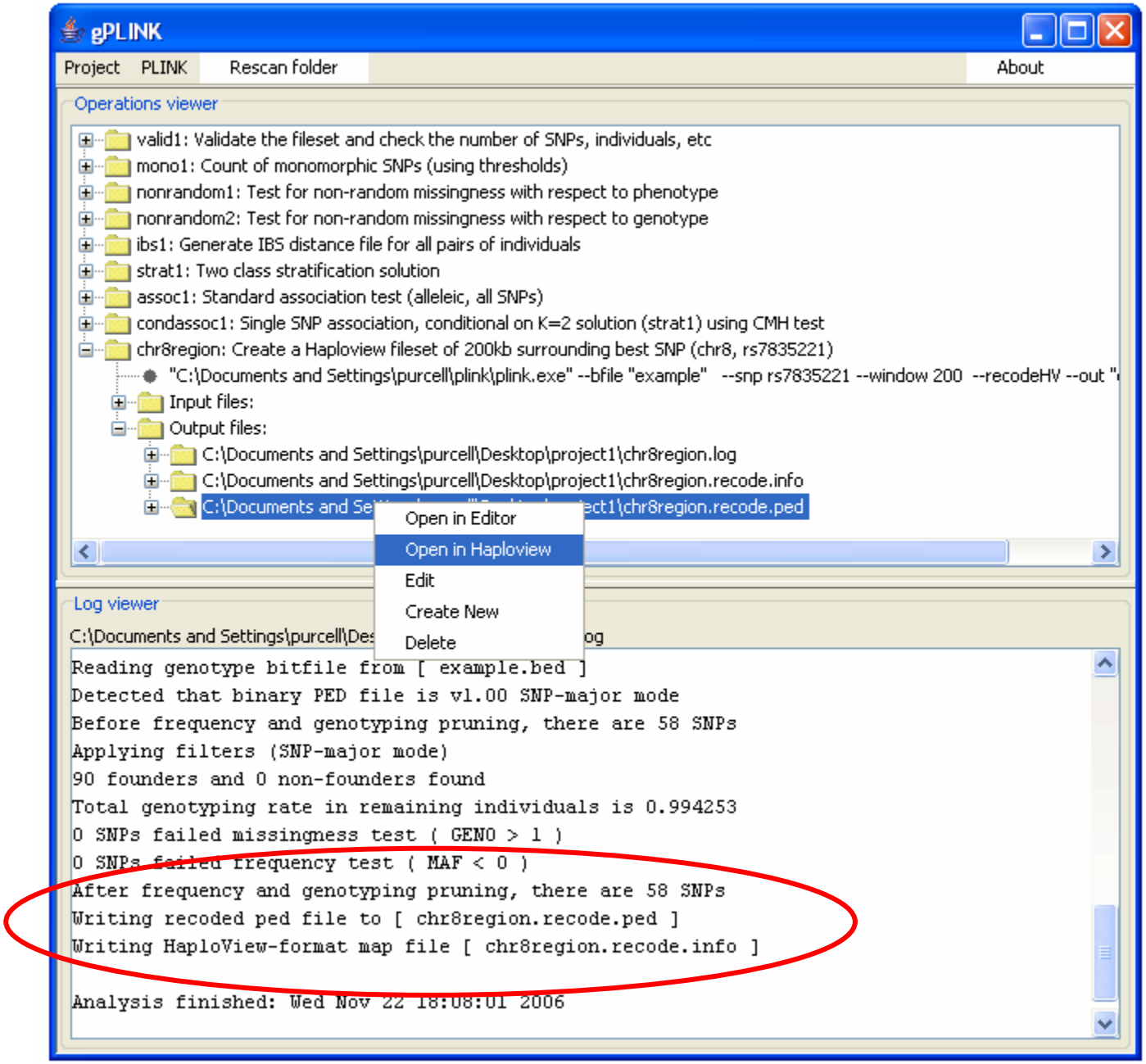


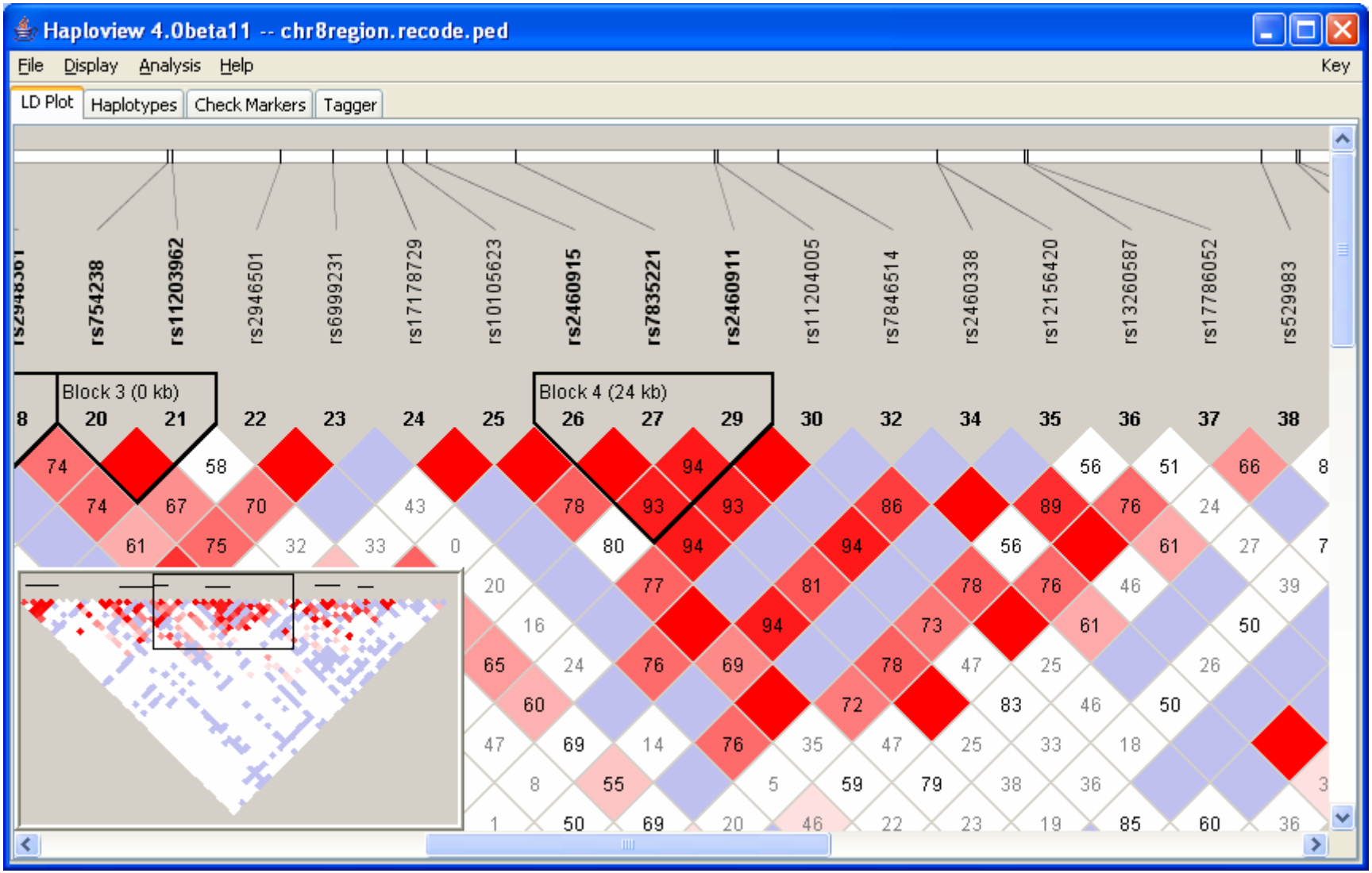



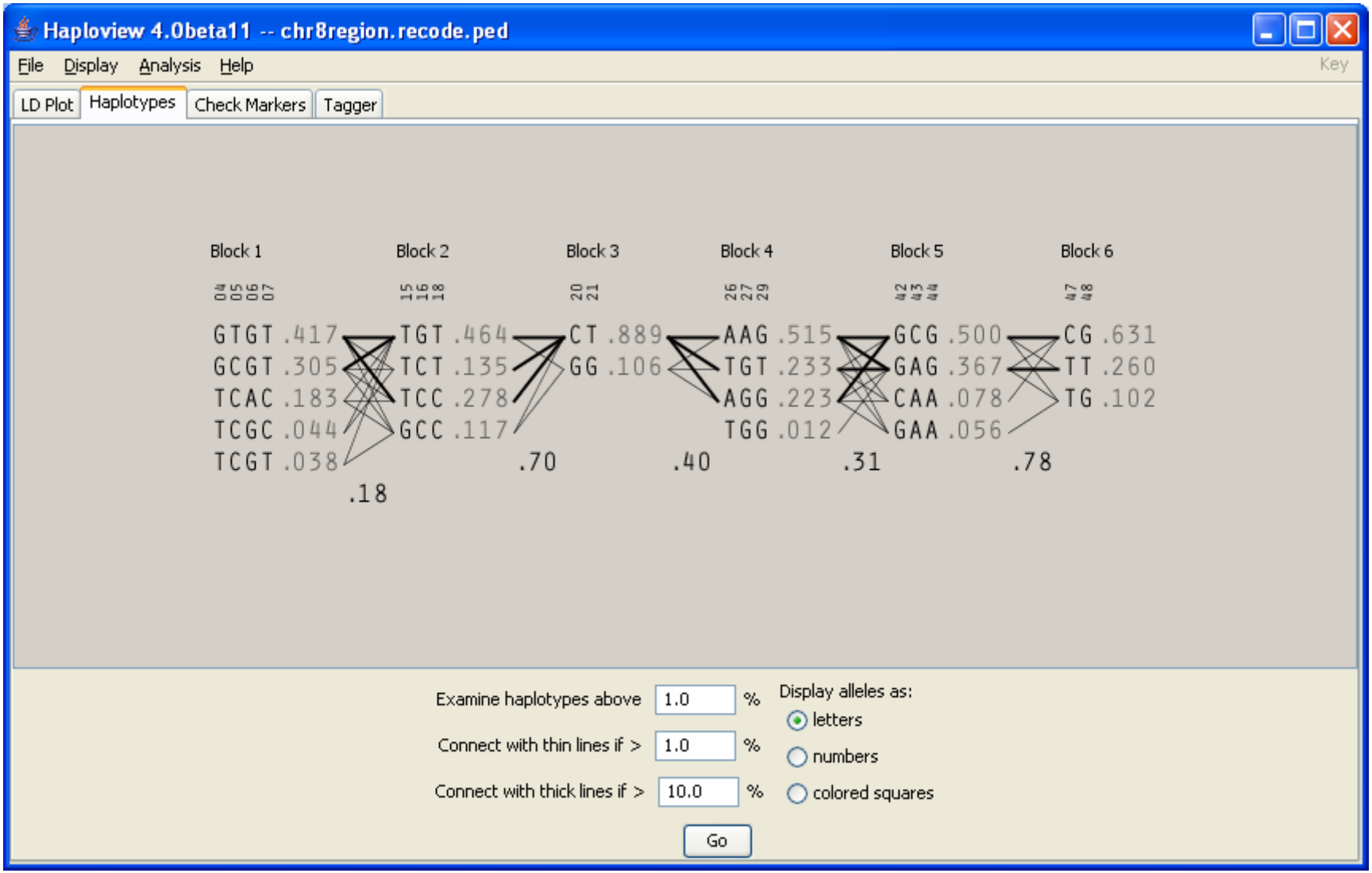
# Making a Haploview fileset

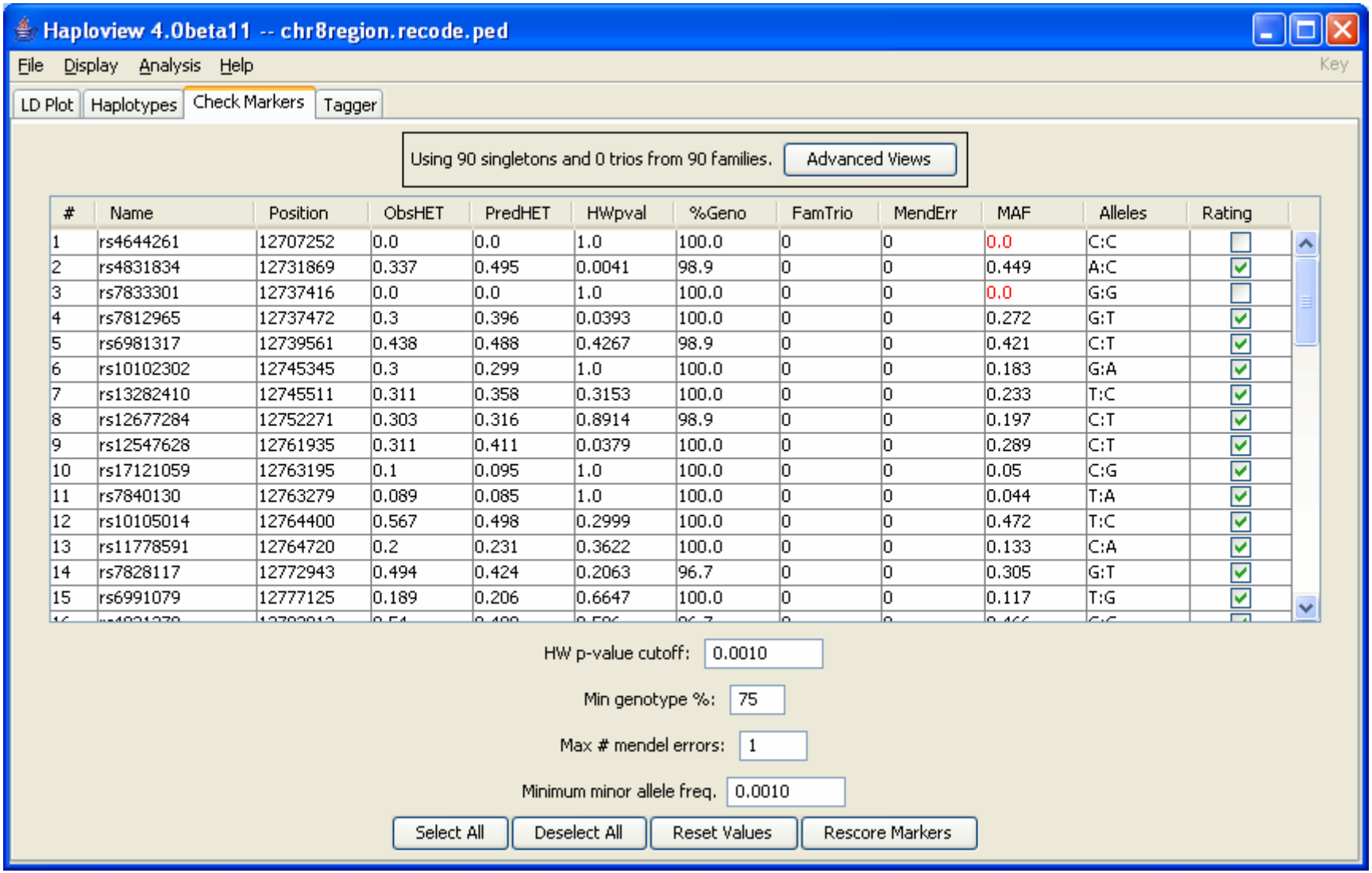



***Select 200kb region around our "best hit"***

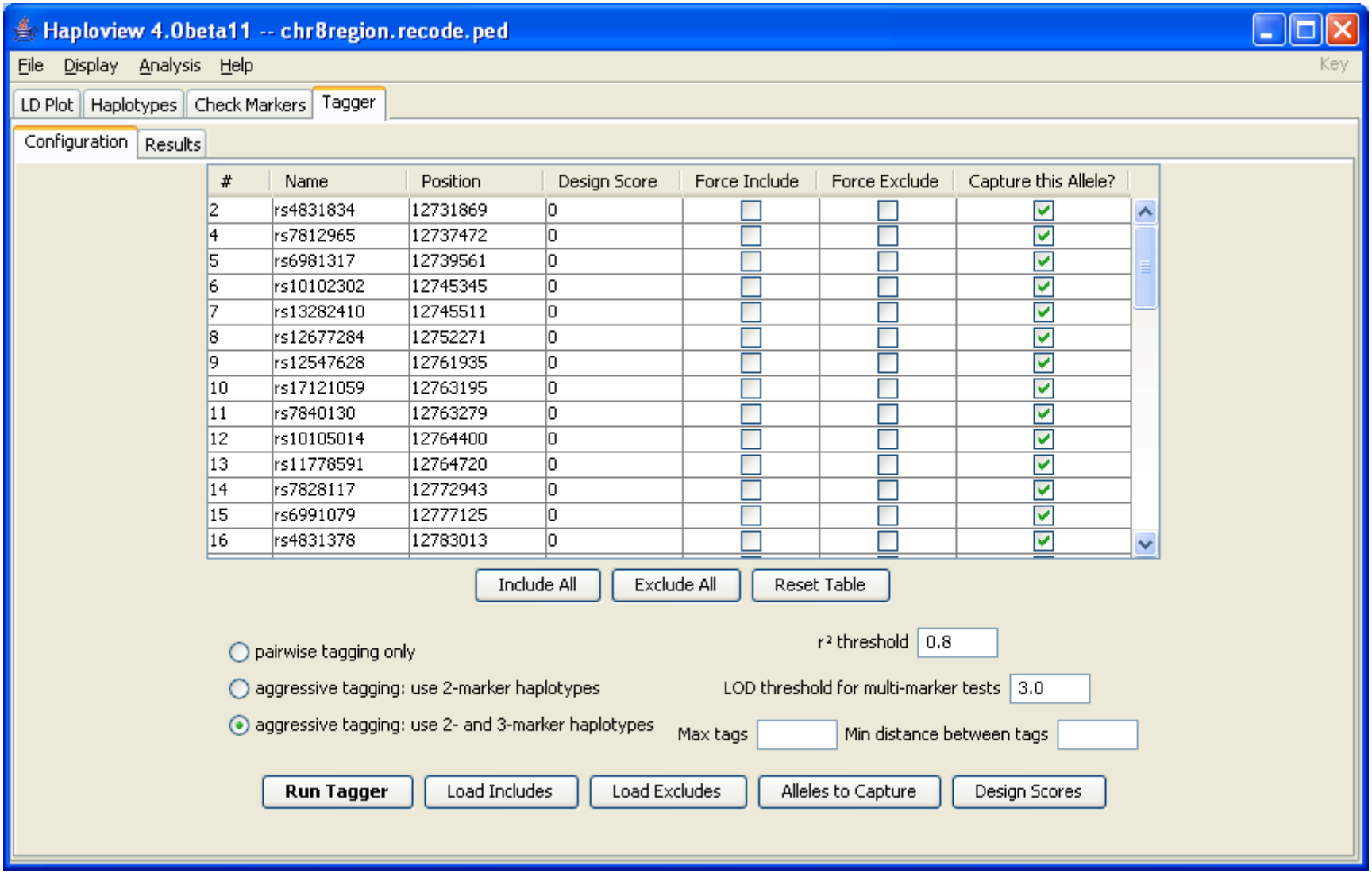

gPLINK
Project
PLINK
Rescan folder
About
Operations viewer
valid1: Validate the fileset and check the number of SNPs, individuals, etc
mono1: Count of monomorphic SNPs (using thresholds)
nonrandom1: Test for non-random missingness with respect to phenotype
nonrandom2: Test for non-random missingness with respect to genotype
ibs1: Generate IBS distance file for all pairs of individuals
strat1: Two class stratification solution
assoc1: Standard association test (alleleic, all SNPs)
condassoc1: Single SNP association, conditional on K=2 solution (strat1) using CMH test
chr8region: Create a Haploview fileset of 200kb surrounding best SNP (chr8, rs7835221)
"C:\Documents and Settings\purcell\plink\plink.exe" --bfile "example" --snp rs7835221 --window 200 --recodeHV --out "
Input files:
Output files:
C:\Documents and Settings\purcell\Desktop\project1\chr8region.log
C:\Documents and Settings\purcell\Desktop\project1\chr8region.recode.info
C:\Documents and Se
ect1\chr8region.recode.ped
Open in Editor
Open in Haploview
Edit
Create New
Delete
Log viewer
C:\Documents and Settings\purcell\De
og
Reading genotype bitfile from [ example.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 58 SNPs
Applying filters (SNP-major mode)
90 founders and 0 non-founders found
Total genotyping rate in remaining individuals is 0.994253
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 58 SNPs
Writing recoded ped file to [ chr8region.recode.ped ]
Writing HaploView-format map file [ chr8region.recode.info ]
Analysis finished: Wed Nov 22 18:08:01 2006


Haploview 4.0beta11 -- chr8region.recode.ped
File
Display
Analysis
Help
Key
LD Plot
Haplotypes
Check Markers
Tagger
rs754238
rs11203962
rs2946501
rs6999231
rs17178729
rs10105623
rs2460915
rs7835221
rs2460911
rs11204005
rs7846514
rs2460338
rs12156420
rs13260587
rs17786052
rs529983
Block 3 (0 kb)
Block 4 (24 kb)
8 20 21 22 23 24 25 26 27 29 30 32 34 35 36 37 38


Haploview 4.0beta11 -- chr8region.recode.ped
File Display Analysis Help
Key
LD Plot
Haplotypes
Check Markers
Tagger
Block 1
Block 2
Block 3
Block 4
Block 5
Block 6
04 05 06 07
15 16 18
20 21
26 27 29
42 43 44
47 48
GTGT .417
GCGT .305
TCAC .183
TCGC .044
TCGT .038
TGT .464
TCT .135
TCC .278
GCC .117
CT .889
GG .106
AAG .515
TGT .233
AGG .223
TGG .012
GCG .500
GAG .367
CAA .078
GAA .056
CG .631
TT .260
TG .102
.18
.70
.40
.31
.78
Examine haplotypes above 1.0 %
Connect with thin lines if > 1.0 %
Connect with thick lines if > 10.0 %
Display alleles as:
letters
numbers
colored squares
Go




| # | Name | Position | ObsHET | PredHET | HWpval | %Geno | FamTrio | MendErr | MAF | Alleles | Rating |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | rs4644261 | 12707252 | 0.0 | 0.0 | 1.0 | 100.0 | 0 | 0 | 0.0 | C:C | ☐ |
| 2 | rs4831834 | 12731869 | 0.337 | 0.495 | 0.0041 | 98.9 | 0 | 0 | 0.449 | A:C | ☑ |
| 3 | rs7833301 | 12737416 | 0.0 | 0.0 | 1.0 | 100.0 | 0 | 0 | 0.0 | G:G | ☐ |
| 4 | rs7812965 | 12737472 | 0.3 | 0.396 | 0.0393 | 100.0 | 0 | 0 | 0.272 | G:T | ☑ |
| 5 | rs6981317 | 12739561 | 0.438 | 0.488 | 0.4267 | 98.9 | 0 | 0 | 0.421 | C:T | ☑ |
| 6 | rs10102302 | 12745345 | 0.3 | 0.299 | 1.0 | 100.0 | 0 | 0 | 0.183 | G:A | ☑ |
| 7 | rs13282410 | 12745511 | 0.311 | 0.358 | 0.3153 | 100.0 | 0 | 0 | 0.233 | T:C | ☑ |
| 8 | rs12677284 | 12752271 | 0.303 | 0.316 | 0.8914 | 98.9 | 0 | 0 | 0.197 | C:T | ☑ |
| 9 | rs12547628 | 12761935 | 0.311 | 0.411 | 0.0379 | 100.0 | 0 | 0 | 0.289 | C:T | ☑ |
| 10 | rs17121059 | 12763195 | 0.1 | 0.095 | 1.0 | 100.0 | 0 | 0 | 0.05 | C:G | ☑ |
| 11 | rs7840130 | 12763279 | 0.089 | 0.085 | 1.0 | 100.0 | 0 | 0 | 0.044 | T:A | ☑ |
| 12 | rs10105014 | 12764400 | 0.567 | 0.498 | 0.2999 | 100.0 | 0 | 0 | 0.472 | T:C | ☑ |
| 13 | rs11778591 | 12764720 | 0.2 | 0.231 | 0.3622 | 100.0 | 0 | 0 | 0.133 | C:A | ☑ |
| 14 | rs7828117 | 12772943 | 0.494 | 0.424 | 0.2063 | 96.7 | 0 | 0 | 0.305 | G:T | ☑ |
| 15 | rs6991079 | 12777125 | 0.189 | 0.206 | 0.6647 | 100.0 | 0 | 0 | 0.117 | T:G | ☑ |


Haploview 4.0beta11 -- chr8region.recode.ped
File Display Analysis Help
Key
LD Plot
Haplotypes
Check Markers
Tagger
Configuration
Results
# Name Position Design Score Force Include Force Exclude Capture this Allele?
2 rs4831834 12731869 0
4 rs7812965 12737472 0
5 rs6981317 12739561 0
6 rs10102302 12745345 0
7 rs13282410 12745511 0
8 rs12677284 12752271 0
9 rs12547628 12761935 0
10 rs17121059 12763195 0
11 rs7840130 12763279 0
12 rs10105014 12764400 0
13 rs11778591 12764720 0
14 rs7828117 12772943 0
15 rs6991079 12777125 0
16 rs4831378 12783013 0
Include All
Exclude All
Reset Table
pairwise tagging only
aggressive tagging: use 2-marker haplotypes
aggressive tagging: use 2- and 3-marker haplotypes
r² threshold 0.8
LOD threshold for multi-marker tests 3.0
Max tags
Min distance between tags
Run Tagger
Load Includes
Load Excludes
Alleles to Capture
Design Scores

# In the remaining time (if any…)

- Extract as a new PLINK fileset just the single best SNP (rs7835221)

- Using this new file, attempt questions 9-14.
  - Here are some clues
    - 9) Summary statistics → Hardy Weinberg
    - 10) Standard association test, with an alternate phenotype
    - 11) Stratified association with Breslow-Day test
    - 12) You've already calculated these (i.e. `*.assoc`, `*.hwe`)
    - 13) This is already calculated also (i.e. `*.missing`)
    - 14) Use genotypic association test

*Consult the PLINK documentation (http://pngu.mgh.harvard.edu/purcell/plink/)*

# In summary

- We performed whole genome
  - summary statistics and QC
  - stratification analysis
  - conditional and unconditional association analysis

- We found a single SNP rs7835221 that…
  - is genome-wide significant
  - has similar frequencies and effects in Japanese and Chinese subpopulations
  - shows no missing or HW biases
  - is consistent with an allelic, dosage effect
  - has common T allele with strong protective effect ( ~0.05 odds ratio)

# Acknowledgements

*Haploview development*

*Julian Maller*

*Dave Bender*
*Jeff Barrett*
*Mark Daly*

*(g)PLINK development*

*Shaun Purcell*

*Kathe Todd-Brown*
*Ben Neale*
*Mark Daly*
*Pak Sham*