Chapter 16: Analyzing Genomewide Association Study Data: A Tutorial using PLINK

Disclaimer

Since the publication of *Statistical Genetics: Genemapping through linkage and association* the data set that was used for the examples in Chapter 16 has been relocated from the Queue portal at the Coriell Institute (https://queue.coriell.org/q) to the dbGaP database at the American National Institutes of Health http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap. The amyotrophic lateral sclerosis and neurologically normal controls (Schymick et al., 2007) are now obtained through application to dbGaP rather than through the Queue portal which granted immediate access to the data.

To assist readers who might find it difficult to gain access to the data, or need immediate access to a data set to learn how to use plink, we have created some simulated data which is available from the www.genemapping.org. Additional example data sets are available from http://pngu.mgh.harvard.edu/~purcell/plink/res.shtml#teach. These data were created based on the amyotrophic lateral sclerosis and neurologically normal controls (Schymick et al., 2007). The data were simulated using gene-dropping. This technique preserves missingness and allele frequencies. However, the simulation methods used to create these data do not preserve LD and Hardy-Weinburg patterns within the data, as such there are minor differences between the results obtained from these simulated data and the real data. All the practicals outlined in the chapter can be conducted with the simulated data. This file contains the first practical section of Chapter 16 which have been revised to include the new file names. Plink v1.04 for Windows was used for these analyses.

- The Editors: November 2008

Obtaining a GWAS dataset

The GWAS dataset that we will use to illustrate quality control and analytic principles can be obtained from www.genemapping.org. These are simulated GWAS data for 550K SNPs generated on Illumina chips based on controlled access datasets of cases with amyotrophic lateral sclerosis (ALS, OMIM=105400) and neurologically normal controls (Schymick et al., 2007).

To follow the practicals in this chapter you will need to download four files. The case data are contained in case_map.zip and case_pre.zip. The control data is located in control_map.zip and control_pre.zip. The zipped folders each contain 23 files: chr1-chr22 and chrX .

There are sets of .map files and .pre files for cases and controls. The .map files contain information about the SNP genetic markers and the .pre files contain genotypes for individual subjects. In these analyses, we will focus on just the autosomes.

Windows users: obtaining Linux-like tools

In order to manipulate the various GWAS input and output files, we suggest that Windows users obtain and install a package that provides the basic features of a Linux-like environment. The Cygwin package is freely available, powerful, easy to install, and can be obtained at: http://cygwin.com/setup.exe. From here forward in this tutorial, we assume that all readers have access to Linux-like command line tools.

Preparing the .map & .pre files for input into PLINK

Both sets of files require modification in order to be used by PLINK. To format the data properly, some facility with data manipulation is required. Teaching these skills is beyond the scope of this chapter; however, the following steps are required.

Key points

Read the PLINK documentation about file input formats. Always review the .log files created by PLINK. If any of the Linux commands below are unfamiliar, read the documentation. Do not sort or reorder the files as the order of SNPs in the .map file implicitly corresponds to the genotypes in the .pre files.

Unzip the four zip files. For Windows users, start Cygwin which opens a DOS-like terminal window. All users, use the “cd” command to change the current directory to the folder where you downloaded the four files. Uncompress the four files by issuing an “unzip” command for each file (e.g., “unzip case_map.zip”). Issue the “ls” command to see a listing of all files. These files contain all the raw data needed for GWAS data analysis. The next task is to reformat these files as “.ped” and “.map” files that PLINK can read in.

We start with the .map file, that should contain data on the SNPs genotyped for a particular sample (cases or controls) and chromosome, with one row per SNP. To see the first 10 rows for one file, issue the command “head case_chr1.snps“. There are eight fields in each file.

There are 46 files that need to be converted (chr1-22 and chrX for cases and controls. To make the appropriate .map file for input into PLINK, we need to delete columns 4-8, and switch the positions of columns 2 and 3. To do this we will use “gawk”. Issue the 46 commands shown in *Panel 16.1*. Using the “head” command, confirm that the new .map files are formatted correctly (there should be three fields – chromosome, SNP ID, and SNP base position). For example, the first line of chr22 in controls should read: 22 rs2027653 15298334 and there are 8,439 rows in this file.

Panel 16.1 Preparing the .map input files for cases and controls using “gawk”

```
gawk ' { print $1 , $3 , $2 } ' case_chr1.snps > case_chr1.map
…  (20 commands for chr2-21 not shown)
gawk ' { print $1 , $3 , $2 } ' case_chr22.snps > case_chr22.map
gawk ' { print $1 , $3 , $2 } ' case_chrX.snps > case_chrX.map
gawk ' { print $1 , $3 , $2 } ' control_chr1.snps > control_chr1.map
…  (20 commands for chr2-21 not shown)
gawk ' { print $1 , $3 , $2 } ' control_chr22.snps > control_chr22.map
gawk ' { print $1 , $3 , $2 } ' control_chrX.snps > control_chrX.map
```

Files of extension .pre describe subjects (and their families if present) and genotypes with one row per subject. As each row usually has thousands of genotypes, use the “less” command to page through the file (the spacebar advances a page, b goes back a page, and q quits). The 46 .pre files (chr1-22 and chrX for cases and controls) contain the subject ID, affection status (0=unknown, 1=unaffected, and 2=affected), and allele calls for each SNP separated by a space (e.g., for chr22 in controls, there are characters representing 2 x 8,439 alleles). Allele calls are given as A, C, G, or T with 0 indicating a missing value or no-call. These files are ready for input into PLINK.

Importing GWAS data into PLINK

The next task is to get the data into PLINK. Read the PLINK documentation for information about creating binary files. Two steps are required. First, the .map and .pre files have to be converted into the far more compact “binary” format (“--make-bed”). This process does not lose any information but allows PLINK to function far faster. This conversion must be done 46 times (chr1-22 and chrX for cases and controls). The PLINK commands for cases and controls are given in *Panel 16.2***.**

Panel 16.2 PLINK commands to make binary files for cases and controls

```
plink --ped case_chr1.pre --map case_chr1.map
      --map3 --no-fid --no-parents --no-sex --make-bed --out cc1
…  (20 commands for chr2-21 not shown)
plink --ped case_chr22.pre --map case_chr22.map
      --map3 --no-fid --no-parents --no-sex --make-bed --out cc22
plink --ped case_chrX.pre --map case_chrX.map
      --map3 --no-fid --no-parents --no-sex --make-bed --out ccX
plink --ped control_chr1.pre --map control_chr1.map
      --map3 --no-fid --no-parents --no-sex --make-bed --out cc1
…  (20 commands for chr2-21 not shown)
plink --ped control_chr22.pre --map control_chr22.map
      --map3 --no-fid --no-parents --no-sex --make-bed --out cc22
plink --ped control_chrX.pre --map control_chrX.map
      --map3 --no-fid --no-parents --no-sex --make-bed --out ccX
```

After these 46 commands have been issued, PLINK has made 138 files, or 2 groups (case/control) x 23 chromosomes x 3 PLINK files (* .bed, *.bim, and *.fam).

The "--map3" option tells PLINK that the .map files contain three fields instead of the default four. The commands "--no-fid --no-parents --no-sex" tell PLINK that the .pre files contain only the subject ID number and the affection code.

We noted an error for the chr17 SNP rs1124736 in controls where more than two alleles were observed. In practice, we would discuss this error with the genotyping facility and attempt to resolve it. For this exercise, remove the SNP from analysis by making a change to control_chr22.map – change line 13,786 to "17 rs1124736 -75527355" (making a base position negative removes that SNP from analysis).

Second, the binary files need to be merged into one set of data for quality control and statistical analyses of association. The PLINK command to merge the 132 individual files is given in *Panel 16.3*.

Panel 16.3. PLINK command to merge case and control binary files plus the contents of "files.txt".

```
Plink --bfile cc1 --merge-list files.txt --out v1gwas --make-bed
  Files.txt
    cc2.bed cc2.bim cc2.fam
    …  (19 lines not shown)
    cc22.bed cc22.bim cc22.fam
    ccX.bed ccX.bim ccX.fam
    con1.bed con1.bim con1.fam
    …  (20 lines not shown)
    con22.bed con22.bim con22.fam
    conX.bed conX.bim conX.fam
```

The command begins with the prefix for one file (arbitrarily selected as cc1); "files.txt" contains a list of the files to merge with the cc1 data 45 rows in total or 22 rows for cc2-ccX and 23 rows for con1-conX. Each row contains the names of the three PLINK binary files in the order shown above (.bed, .bim, and .fam).

For this tutorial, we require a sex for each subject. We will use the sex assignment determined by chrX SNP genotypes. The two PLINK commands in Panel 16.5 assign sex based on chrX SNP genotypes, and the two "gawk" and the "cat" commands make a new version of the v1gwas.fam file that now contains sex for all subjects (1=male and 2=female instead of 0=unassigned).

Panel 16.4 PLINK and Unix commands to create a new .fam file and to assign sex based on chrX genotypes.

```
plink --bfile ccX  --check-sex --mind 1 --out ccsex
plink --bfile conX --check-sex --mind 1 --out consex
gawk ' NR>1 { print $1,$2,0,0,$4,2 } ' ccsex.sexcheck > cc.fam
gawk ' NR>1 { print $1,$2,0,0,$4,1 } ' consex.sexcheck > con.fam
cat cc.fam con.fam > v1gwas.fam
```

Successful execution of these commands has created version 1 of the GWAS data for cases and controls. There are now three files: v1gwas.bed, v1gwas.bim, and v1gwas.fam that contain GWAS data for 541,327 SNPs (genotyping success rate 0.99632) on 547 individuals (276 cases, 271 controls, 306 males, 241 females). To save hard drive space you can now delete, backup, or compress/store all of the *.pre, *.map, cc*, and con* files.

Results from analyses of the simulated data

1. Identify/remove highly related samples using genomewide GWAS data.

Panel 16.5 Commands for identifying excessively similar individuals.

```
# Step 1: calculate pairwise genomewide identity-by-state (IBS)
plink --bfile v1gwas --genome --geno 0 --mind 1 --out ibsibd
# Step 2: find the pairs of subjects with excessive relatedness

gawk ' $8 > 0.8 ' ibsibd.genome
FID1    IID1    FID2     IID2            Z0  Z1      Z2  PI_HAT  PHE  IBS0   IBS1     IBS2
ND-6188 ND-6188 ND-6595  ND-6595          0   0       1       1    1     0      0   283021
ND-6805 ND-6805 ND-9185  ND-9185 5.478e-05   0  0.9999  0.9999    1     1      2   283018
ND-8469 ND-8469 ND-10260 ND-10260         0   0       1       1    1     0      0   283021
```

The second section of Panel 16.5 shows the results obtained from analysis of the ALS cases and neurologically normal controls. Running the gawk command on the simulated data will not yield any cases with a phi-hat >.8 and there is no need to remove any cases from the data. This is because the gene-dropping method used to simulate the data considered each snp in isolation and did not preserve the LD structure within the data. For the sake of yielding some data you may with to change the gawk command to: gawk ' $8 > 0.037 ' ibsibd.genome

2.Creating the analysis GWAS dataset.

Panel 16.6 PLINK command to apply default thresholds & create GWAS analysis dataset.

```
plink --bfile v1gwas --mind 0.05 --geno 0.01 --maf 0.01 --hwe 1e-3
      --out v2gwas --make-bed
```

In the simulated dataset no cases or controls are removed for excessive missing data and there are 273 cases (63.0% male) and 271 controls (48.3% male). Of 541,327 initial SNPs, 317 were excluded due to HWE deviations, 48,196 failed the missingness test, and 23,499 SNPs failed the frequency test – leaving a total of 483,588 SNPs in the analysis file. The genotyping success rate in these 547 unique individuals was 0.996.

3. Missingness

Panel 16.7 PLINK commands for analyzing missing data.

```
plink --bfile v2gwas --out v2gwasmiss ––missing
```

Results from the simulated data:

Individual missingness (v2gwasmiss.imiss)

FID	IID	MISS_PHENO	N_MISS	N_GENO	F_MISS
Case01	Case01	N	196	483588	0.0004053
Case02	Case02	N	132	483588	0.000273
Case03	Case03	N	102	483588	0.0002109
Case04	Case04	N	1719	483588	0.003555
Case05	Case05	N	344	483588	0.0007113

SNP missingness (v2gwasmiss.lmiss)

CHR	SNP	N_MISS	N_GENO	F_MISS
1	rs3934834	5	547	0.009141
1	rs3737728	0	547	0
1	rs6687776	0	547	0
1	rs9651273	1	547	0.001828

1	rs4970405	1	547	0.001828

4. Descriptive analyses of the case/control GWAS data

--test-missing

Panel 16.8 PLINK commands for testing missingness between cases and controls.

```
plink --bfile v2gwas --out v2gwasmisstest --test-missing
```

Results from the simulated data:

Tests for missingness by case-control status (v2gwasmisstest.missing)

CHR	SNP	F_MISS_A	F_MISS_U	P
1	rs3934834	0	0.01845	0.02929
1	rs3737728	0	0	1
1	rs6687776	0	0	1
1	rs9651273	0	0.00369	0.4954
1	rs4970405	0.003623	0	1

--freq

Panel 16.9 PLINK commands for generating allele frequencies.

```
plink –-bfile v2gwas --out v2gwasfreq –-freq
```

Results from the simulated data;

Allele frequency data (v2gwasfreq.frq)

CHR	SNP	A1	A2	MAF	NCHROBS
1	rs3934834	T	C	0.1679	1084
1	rs3737728	T	C	0.2706	1094
1	rs6687776	T	C	0.1344	1094
1	rs9651273	A	G	0.2866	1092

A1 and A2 refer to the minor and major alleles. MAF (minor allele frequency) is the number of occurrences of the minor allele divided by the number of non-missing chromosomes (NCHROBS, maximum value is two times the sample size).

--hardy

Panel 16.10 PLINK commands for generating genotype frequencies and HWE estimates.

```
plink --bfile v2gwas --out v2gwashwe --hardy
```

Results from the simulated data:

Tests of Hardy-Weinberg Equilibrium (v2gwashwe.hwe)

CHR	SNP	A1	A2	TEST	GENO	O(HET)	E(HET)	P
1	rs3934834	T	T	ALL	11/160/371	0.2952	0.2794	0.2203
1	rs3934834	T	T	AFF	5/82/189	0.2971	0.2778	0.3831
1	rs3934834	T	T	UNAFF	6/78/182	0.2932	0.2811	0.6619
1	rs3737728	T	T	ALL	43/210/294	0.3839	0.3947	0.5166
1	rs3737728	T	T	AFF	20/99/157	0.3587	0.3768	0.4257

There are three rows per SNP – for all subjects ("ALL"), cases ("AFF"), and controls ("UNAFF"). Because the simulated data were created via gene-dropping only 3403 snps in the case sample and 2997 snps in the control samples had HWE exact pvalues <.01.

5. Evaluate relatedness

Panel 16.11 Re-calculating genomewide autosomal IBS matrix using PLINK.

```
plink --bfile v1gwas --read-genome ibsibd.genome
      --remove dup_samples.txt --cluster --cc --ppc 1e-3
      --mds-plot 5 --out v2gwasstrat
```

Results from the simulated data;

Multidimensional scaling results (v2gwasstrat.mds)

FID	IID	SOL	C1	C2	C3	C4	C5
Case01	Case01	0	-0.0116226	0.0128164	-0.00194903	0.0140202	-0.00439348
Case02	Case02	0	-0.012478	0.0124638	0.00162505	0.00540692	0.00862621
Case03	Case03	1	-0.0126428	0.0174033	-0.00131193	0.0171144	-0.00560634
Case04	Case04	0	-0.0113282	0.00565285	-0.00502735	0.00424166	0.00835868
Case05	Case05	2	-0.0112742	-0.00489727	0.000230032	0.00483719	0.00433144

Because the simulated data were created via gene-dropping there is no evidence of complex relatedness or population stratification within the data

6. Association Analyses of GWAS Data

Single SNP analyses for case-control status.

Panel 16.12 PLINK commands for association analyses.

```
plink --bfile v2gwas --out v2gwasassoc --assoc

plink --bfile v2gwas --out v2gwasmodel --model --model-trend --adjust

Adjustment for multiple comparisons (v2gwasmodel.model.trend.adjusted)
     CHR        SNP      UNADJ          GC     BONF     HOLM SIDAK_SS SIDAK_SD FDR_BH FDR_BY
     10   rs4363506 4.568e-007 4.791e-007  0.2156   0.2156  0.1939  0.1939  0.2156       1
      5   rs5014235 7.823e-006 8.124e-006       1        1  0.9751  0.9751  0.9625       1
      9   rs7469576  9.07e-006 9.414e-006       1        1  0.9862  0.9862  0.9625       1
      5   rs2289599 1.129e-005 1.171e-005       1        1  0.9952  0.9952  0.9625       1
     …
```

Results from the simulated data:

Basic association information (v2gwasassoc.assoc)

CHR	SNP	BP	A1	F_A	F_U	A2	CHISQ	P	OR
1	rs3934834	995668	T	0.1667	0.1692	C	0.01218	0.9121	0.9822
1	rs3737728	1011277	T	0.2518	0.2897	C	1.986	0.1588	0.8253
1	rs6687776	1020427	T	0.1178	0.1513	C	2.645	0.1039	0.7487
1	rs9651273	1021402	A	0.2717	0.3019	G	1.211	0.2712	0.863
1	rs4970405	1038817	G	0.1018	0.1089	A	0.1435	0.7048	0.928

Tests for association (v2gwasmodel.model)

CHR	SNP	A1	A2	TEST	AFF	UNAFF	CHISQ	DF	P
1	rs3934834	T	C	GENO	5/82/189	6/78/182	0.1385	2	0.9331
1	rs3934834	T	C	TREND	92/460	90/442	0.01291	1	0.9095
1	rs3934834	T	C	ALLELIC	92/460	90/442	0.01218	1	0.9121
1	rs3934834	T	C	DOM	87/189	84/182	0.0002053	1	0.9886
1	rs3934834	T	C	REC	5/271	6/260	0.1343	1	0.714

Adjustment for multiple comparisons (v2gwasmodel.model.trend.adjusted)

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
2	rs2008831	2.126e-012	5.773e-007	1.028e-006	1.028e-006	1.028e-006	1.028e-006	1.028e-006	1.405e-005
10	rs12359379	8.107e-012	1.151e-006	3.921e-006	3.92e-006	3.92e-006	3.92e-006	1.96e-006	2.679e-005
14	rs1872020	7.08e-010	1.157e-005	0.0003424	0.0003424	0.0003423	0.0003423	0.0001027	0.001404
17	rs445507	9.414e-010	1.34e-005	0.0004552	0.0004552	0.0004551	0.0004551	0.0001027	0.001404
15	rs4887400	1.199e-009	1.519e-005	0.0005797	0.0005797	0.0005795	0.0005795	0.0001027	0.001404
2	rs4425131	1.274e-009	1.568e-005	0.0006163	0.0006163	0.0006161	0.0006161	0.0001027	0.001404
20	rs6028712	1.525e-009	1.72e-005	0.0007374	0.0007373	0.0007371	0.0007371	0.0001053	0.00144

The first PLINK command ("--assoc") is most useful in showing the allele frequency differences between cases and controls along with the allelic odds ratio (confidence intervals can also be printed).

The second PLINK command ("--model") yields five lines per SNP, one for each of the association tests in *Table 16.1*. The fields give the chromosome, SNP ID, the specific test (keywords are given in *Table 16.1*), the cell frequencies for cases ("AFF") and controls ("UNAFF"), and the statistical data (χ^2 value, df, and asymptotic p-value).

The PLINK "--model-trend --adjust" command adjusts the TREND *P*-values using genomic control, Bonferroni, Holm, Sidak, and false discovery rate corrections. The most significant single SNP in the simulated GWAS rs2008831 had a TREND *P*=2.126e-012 which did survive adjustment for multiple comparisons. In the true ALS GWAS data the most significant single SNP, rs4363506, had a TREND *P*=0.00000046 which did not survive adjustment for multiple comparisons. In line with the original text we will examine the additional data available for rs4363506.

Panel 16.13 shows how to retrieve results for the ASCII files created by PLINK. First, use the "fgrep" command to retrieve lines containing "rs4363506" from the missingness, HWE, and association results. The second "fgrep" command gives a quick-and-dirty look at the genomic context by showing 10 lines before and after the line containing rs4363506.

Panel 16.13 Commands to retrieve results.

```
fgrep -w rs4363506 v2*missing v2*hwe v2*assoc
```

Results from the simulated data:

```
v2gwasmisstest.missing: 10 rs4363506 0.003623 0.00738 0.6209
v2gwashwe.hwe: 10 rs4363506 C C ALL 89/268/187 0.4926 0.4838 0.7231
v2gwashwe.hwe: 10 rs4363506 C C AFF 58/139/78 0.5055 0.4974 0.9034
v2gwashwe.hwe: 10 rs4363506 C C UNAFF 31/129/109 0.4796 0.458 0.5061
v2gwasassoc.assoc: 10 rs4363506 129164492 C 0.4636 0.355 T 13.26 0.0002704 1.57
```

```
fgrep -w rs4363506 -C 10 v2*.assoc
```

Results from the simulated data:

								PVALUE	OR
10	rs10741272	129157196	A	0.03804	0.07749	G	7.841	0.005106	0.4708
10	rs11018124	129157632	G	0.09818	0.07778	A	1.412	0.2347	1.291
10	rs11018125	129157748	A	0.2174	0.2196	G	0.007514	0.9309	0.9874
10	rs1931725	129158516	A	0.1431	0.107	C	3.254	0.07124	1.394
10	rs9651385	129158528	G	0.4058	0.2915	A	15.72	7.354e-005	1.66
10	rs11018127	129159048	T	0.3877	0.4428	C	3.423	0.0643	0.7967
10	rs12776795	129161305	G	0.1286	0.1384	A	0.225	0.6353	0.9191
10	rs10765110	129161839	A	0.3279	0.2509	C	7.874	0.005016	1.456
10	rs10765113	129163062	A	0.01993	0.02768	G	0.7076	0.4003	0.7144
10	rs11018134	129163246	T	0.3551	0.3155	C	1.921	0.1657	1.194
10	rs4363506	129164492	C	0.4636	0.355	T	13.26	0.0002704	1.57
10	rs11018140	129166808	G	0.1377	0.07749	T	10.3	0.001333	1.901
10	rs11018141	129167540	G	0.2699	0.2269	T	2.706	0.1	1.259
10	rs4622172	129171660	T	0.3786	0.3506	C	0.93	0.3349	1.129
10	rs1931737	129171889	T	0.3641	0.262	G	13.26	0.0002718	1.613
10	rs10830099	129174354	C	0.3818	0.5148	T	19.49	1.01e-005	0.5821
10	rs10830100	129174402	C	0.529	0.4272	T	11.28	0.000783	1.506
10	rs4751427	129178172	T	0.1377	0.09963	C	3.781	0.05183	1.443
10	rs10830102	129179415	T	0.1667	0.1167	G	5.603	0.01793	1.514
10	rs7090712	129179547	C	0.3478	0.3838	T	1.523	0.2172	0.8564
10	rs2065674	129181355	T	0.3025	0.3745	C	6.334	0.01184	0.7244

7. Multi-marker analyses

Panel 16.14 PLINK commands for haplotype association using HapMap CEU definitions

(il550_CEU.hap).

```
plink --bfile v2gwas --hap ill550_CEU.hap --out v2gwashapassoc --hap-assoc
```

Results from the simulated data:

Haplotype association information (v2gwashapassoc.assoc.hap)

LOCUS	HAPLOTYPE	F_A	F_U	CHISQ	DF	P	SNPS
i_rs4074137	GA	0.6584	0.6202	1.728	1	0.1886	rs9651273|rs4970405
i_rs10907174	CC	0.247	0.2423	0.03236	1	0.8573	rs9729550|rs6685064
i_rs6603783	TC	0.009171	0.01268	0.3128	1	0.576	rs7515488|rs6675798
i_rs3766177	CT	0.352	0.295	4.025	1	0.04483	rs9439462|rs6603793
i_rs7531530	TT	0.334	0.2728	4.829	1	0.02799	rs819980|rs6603793

```
fgrep -w rs4363506 -C 5 v2*hap
```

Results from the simulated data:

i_rs885587	CT	0.5288	0.4159	13.93	1	0.0001896	rs10741271|rs10830099
i_rs11018123	AG	0.8818	0.9011	1.046	1	0.3064	rs11018124|rs10765113
i_rs10765108	AC	0.588	0.6441	3.637	1	0.05649	rs12776795|rs10765110
i_rs10765111	CT	0.4802	0.5717	9.175	1	0.002454	rs10765110|rs11018141
i_rs884436	TC	0.4721	0.5106	1.623	1	0.2026	rs11018141|rs4622172
i_rs949523	TG	0.3468	0.4812	20.28	1	6.7e-006	rs4363506|rs1931737
i_rs11018150	CT	0.2483	0.1941	4.655	1	0.03095	rs10765110|rs1931737
i_rs4751428	AT	0.5295	0.4132	14.73	1	0.0001238	rs2184376|rs7090712
i_rs1572236	GT	0.1756	0.1981	0.8995	1	0.3429	rs10830102|rs12784910
i_rs2065675	GG	0.7827	0.7654	0.4633	1	0.4961	rs10765113|rs12784910
i_rs2184379	TT	0.05502	0.01355	14.09	1	0.0001741	rs10830102|rs1931738

8. Single SNP analyses of quantitative traits.

Panel 16.15 shows the PLINK commands to conduct quantitative GWAS using the five MDS dimensions as separate dependent variables. The alternate phenotype file contains the family number and ID for each of the 544 subjects along with the five MDS dimensions. The first few lines from one of the five output files are also shown.

Panel 16.15. PLINK commands to conduct quantitative GWAS separately for 5 phenotypes supplied in an alternate phenotypes file.

```
plink --bfile v2gwas --out v2gwas_d1 --assoc --pheno mds.txt --mpheno 1
plink --bfile v2gwas --out v2gwas_d2 --assoc --pheno mds.txt --mpheno 2
plink --bfile v2gwas --out v2gwas_d3 --assoc --pheno mds.txt --mpheno 3
plink --bfile v2gwas --out v2gwas_d4 --assoc --pheno mds.txt --mpheno 4
plink --bfile v2gwas --out v2gwas_d5 --assoc --pheno mds.txt --mpheno 5
```

Results from the simulated data:

Alternate phenotype file (mds.txt)

FID	IID	C1	C2	C3	C4	C5
Case01	Case01	-0.0116226	0.0128164	-0.00194903	0.0140202	-0.00439348
Case02	Case02	-0.012478	0.0124638	0.00162505	0.00540692	0.00862621
Case03	Case03	-0.0126428	0.0174033	-0.00131193	0.0171144	-0.00560634
Case04	Case04	-0.0113282	0.00565285	-0.00502735	0.00424166	0.00835868
Case05	Case05	-0.0112742	-0.00489727	0.000230032	0.00483719	0.00433144
Case06	Case06	-0.0120745	0.000616837	-0.0206544	0.0127316	0.00853495

Example output (v2gwas_d1.qassoc)

CHR	SNP	BP	NMISS	BETA	SE	R2	T	P
1	rs3934834	995668	542	0.0002167	0.00102	8.354e-005	0.2124	0.8319
1	rs3737728	1011277	547	0.001038	0.0008178	0.002948	1.269	0.2049
1	rs6687776	1020427	547	0.001735	0.001086	0.004661	1.598	0.1107
1	rs9651273	1021402	546	0.0009449	0.0008457	0.00229	1.117	0.2644
1	rs4970405	1038817	546	0.0003601	0.001192	0.0001678	0.3022	0.7626
1	rs12726255	1039812	547	-0.0002016	0.001024	7.117e-005	-0.197	0.8439

As these data were created using gene-dropping the MDS dimensions do not localize to any particular genomic region.