

Generating Marker Summary Reports Using the *GeneticsBase*

Gregory Warnes
gregory_warnesurmc.rochester.edu,
Nitin Jain
nitin.jainpfizer.com

April 21, 2009

1 Introduction

This document demonstrates how to use the *GeneticsBase* (version 0.6.0) package to generate marker summary tables *for studies with a small number of markers*. It is written as a step-by-step tutorial. For additional details on each of the R functions utilized, please see the individual help pages

Note: The textual displays described here are not suitable for large numbers of markers. They are intended for reviewing detailed information on a small number of markers, such as those in candidate gene studies, or a small set of markers achieving a 'quality' or 'significance' cutoff from a larger set.

2 Example

2.1 Prepare phenotype data

The first step is to prepare the phenotype data. It may be in the form of a SAS dataset, SAS export file, comma-delimited text file (CSV), tab-delimited text file (TSV), or Microsoft Excel spreadsheet file (XLS). It should have one row per observation and one column per variable, and must contain a subject identifier variable that can be used to match observations with the corresponding genotype data.

2.2 Prepare genotype data

You also need to store the genetic call data in a file that can be read into R. *GeneticsBase* package accepts genotype data in a variety of formats:

- standard pedigree (ped) format.

a2m	apoe					
50103	5010004	5090005	5090004	2	2	1
2	3	4				
50103	5010005	5090005	5090004	2	2	1
1	3	4				
50105	5010049	5090021	5090022	2	2	1
1	4	4				
50105	5010070	5090020	5090019	1	2	1
1	3	4				

- **hapmap format** : The hapmap .ped format is a variant of the standard pedigree format. A portion of the first two lines of the hapmap file for chromosome 1 are shown below:

```
rs2298011 rs1320571 rs11721 rs4018608 rs6685064 rs604618 ...
1347 14 0 0 1 1 1 1 3 3 3 3 1 1 2 2 3 3 3 3 2 2 3 3 2 ....
```

- Pfizer format: First few lines of an example file in Pfizer's data format are shown below:

Locus	Gene	Marker	Locus Start	Project	Protocol	Sample ID	Donor ID	Genotype
A1A	C1556G	-1243	P234	103	1028022.1	1028022	G/G	
A1B	T127A	20141	P234	103	1028022.1	1028022	A/T	
A1B	T5094A	102358	P234	103	1028022.1	1028022	A/T	
A1A	C1556G	-1243	P234	103	1035130.1	1035130	G/G	

- Perlegen format: A portion of first two lines of data in the Perlegen format are shown below:

[illegible]

2.3 Load the genotype and phenotype data

In GeneticsBase, both genotype and phenotype data is loaded by a single function *readgenes*. This function *readGenes* has four primary arguments: *gfile*, *gformat*, *file*, and *pformat*. Arguments *gfile* and *pfile* are the names of files containing the genotype and phenotype data (respectively), and arguments *gformat* and *pformat* are the corresponding file formats for the genotype and phenotype data.

Various types of data can be loaded by readGenes function. Example files are provided in **data** subdirectory of the **GeneticsBase** package. To access these execute

```
> library(GeneticsBase)
> setwd(file.path(.path.package("GeneticsBase"), "data"))
```

Supported file formats include:

- fbat .ped format

The Alzheimer's example dataset is stored in the Fbat variant of the .ped Pedigree Format. As it does not include phenotype data, we only use the `genotypep` filename and file type arguments:

```
> ALZH <- readGenes(gfile = "ALZH.ped", gformat = "fbat.ped")
```

The CAMP example dataset is from from the ‘Childhood Asthma Management Program (CAMP)’ and incudes both genotype and phenotype information. It can be loaded by:

```
> CAMP <- readGenes(gfile = "CAMP.ped", gformat = "fbat.ped", pfile = "CAMPZ.phe",
+                   pformat = "fbat.phe")
```

- HAPMAP .ped format

A subset of the data for the International HapMap project is available in the hapmap example data set. This file can be loaded via:

```
> hapmapchr1 <- readGenes(gfile = "hapmapchr1.ped", gformat = "hapmap")
```

- Pfizer genetics data format

```
> PfizerExample <- readGenes.pfizer("PfizerExample.txt", format = "Listing")
```

- Perlegen data format

```
> PerlegenExample <- readGenes("PerlegenExample.txt", gformat = "perlegen")
```

2.4 Reviewing the data

For the purpose of this example, we are going to use CAMP data set, which can be loaded manually as shown in the previous sub-section, or via

```
> library(GeneticsBase)
> data(CAMP)
```

```
Reading 8 markers and 2011 subjects from `CAMP.ped` ...
generating 'geneSet' object...
```

```
Successfully read the pedigree file `CAMP.ped`.
```

```
Number of Markers: 8
Number of Subjects: 2011
Number of Families: 651
```

```
Reading 12 vars from `CAMPZ.phe` ... Done.
```

```
Number of Phenotype Variables: 12
Number of Observations      : 2011
```

Now you can see a brief summary of the data that was loaded by simply entering the name of the object on a line by itself:

```
> CAMP
```

```
geneSet object
-----
```

```
Number of Markers:      8
Number of Observations: 2011
```

```
Sample variables: family, pid, father, mother, sex, affected, zposfevp,
```

```
zposfvcp, zlog
```

```
Genetic data:
```

	1.1900	1.1667	1.978	2.1391	2.1988	2.109		649.1837	650.1736	650.1908
m709	1/1	1/1	1/1	1/1	1/1	1/1	...	1/1	1/1	1/1
m654	1/1	1/1	1/1	1/1	2/1	1/1	...	2/1	<NA>	1/1
m47	1/1	1/2	1/2	1/2	2/2	1/2	...	2/2	1/1	1/2
p46	2/2	2/2	2/2	2/2	1/2	2/2	...	1/2	2/2	2/2
p79	2/2	2/1	2/1	2/1	1/1	2/1	...	1/1	<NA>	2/1
p252	2/2	1/2	1/2	1/2	1/2	<NA>	...	1/2	2/2	1/2
p491	1/1	1/1	1/1	1/1	1/1	1/1	...	1/1	<NA>	1/1
p523	1/1	1/2	1/2	1/2	1/2	1/2	...	1/2	1/1	1/2

	650.1675	651.568	651.1725
m709	1/1	1/1	1/1
m654	1/1	2/2	2/1
m47	1/2	2/2	2/2
p46	1/2	1/1	1/2
p79	2/1	1/1	1/1
p252	1/2	2/2	1/2
p491	1/1	1/1	1/1
p523	1/2	1/1	1/2

Warning messages:

```
1: geneSet Object has 121 observations. Only first and last 6 displayed
in: .local(object, ...)
```

The phenotype data can be extracted from the CAMP data object using the `sampleInfo` command:

```
> pdata <- sampleInfo(CAMP)
> summary(pdata)
```

family	pid	father	mother
Min. : 1.0	Min. : 1.0	Min. : 0.0	Min. : 0.0
1st Qu.:165.0	1st Qu.: 503.5	1st Qu.: 0.0	1st Qu.: 0.0
Median :327.0	Median :1006.0	Median : 0.0	Median : 0.0
Mean :326.9	Mean :1006.0	Mean : 340.9	Mean : 367.7
3rd Qu.:489.0	3rd Qu.:1508.5	3rd Qu.: 521.0	3rd Qu.: 637.5
Max. :651.0	Max. :2011.0	Max. :2009.0	Max. :2010.0

sex	affected	zposfevp	zposfvcp
Min. :1.000	Min. :0.0000	Min. : -3.234e+00	Min. : -2.880e+00
1st Qu.:1.000	1st Qu.:0.0000	1st Qu.: -6.790e-01	1st Qu.: -6.250e-01
Median :1.000	Median :0.0000	Median : 1.000e-03	Median : -1.600e-02
Mean :1.453	Mean :0.7041	Mean : 1.431e-05	Mean : -3.433e-05
3rd Qu.:2.000	3rd Qu.:2.0000	3rd Qu.: 6.275e-01	3rd Qu.: 6.005e-01
Max. :2.000	Max. :2.0000	Max. : 4.021e+00	Max. : 4.041e+00

...

2.5 Generate the tables

We can generate a variety of summary tables on our genetics data.

- Allele information

```
> alleleSummary(CAMP)
```

Gene	Marker	Position	Group	Allele	Count	Freq	CI-Lower	CI-Upper
ALL	m709	?	ALL	1	3904	0.998	0.996	0.999
			ALL	2	8	0.002	0.001	0.004
	m654	?	ALL	1	2491	0.638	0.623	0.654
			ALL	2	1411	0.362	0.346	0.377

m47	?	ALL	1	1417	0.369	0.354	0.384
		ALL	2	2427	0.631	0.616	0.646
p46	?	ALL	1	1557	0.401	0.385	0.416
		ALL	2	2329	0.599	0.584	0.615
p79	?	ALL	1	2407	0.626	0.611	0.642

...

- Genotype information

```
> genotypeSummary(CAMP)
```

Gene	Marker	Position	Group	Genotype	Count	Freq	CI-Lower	CI-Upper	Expected
?	m709	?	ALL	1/1	1948	0.996	0.993	0.998	1948.008
				1/2	8	0.004	0.002	0.007	7.984
				2/2	0	0.000			0.008
				NA	55				
?	m654	?	ALL	1/1	826	0.423	0.401	0.445	795.115
				1/2	839	0.430	0.408	0.452	900.769
				2/2	286	0.147	0.131	0.162	255.115
				NA	60				
?	m47	?	ALL	1/1	276	0.144	0.128	0.159	261.172
				1/2	865	0.450	0.428	0.472	894.656
				2/2	781	0.406	0.384	0.428	766.172

...

- Marker information
- Linkage disequilibrium, text view

```
> ld <- LD(CAMP)
```

```
> ld
```

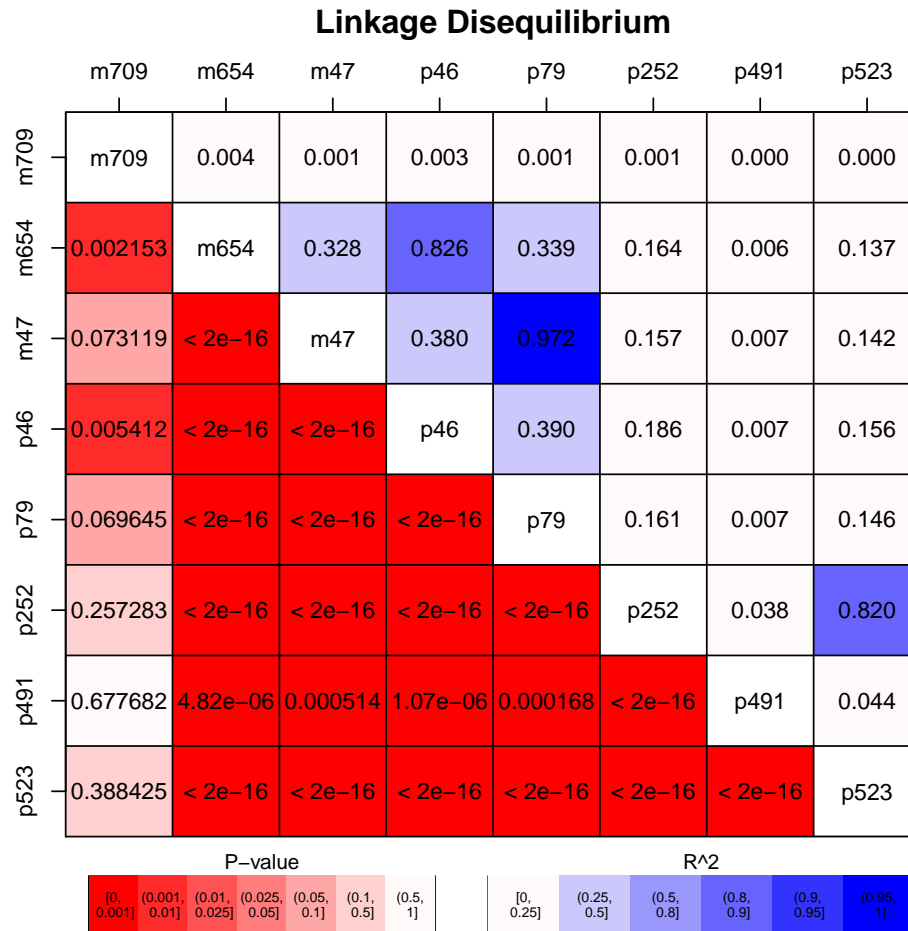
```
-----  
Pairwise LD
```

	m709	m654	m47	p46	p79	p252	p491	p523
m709 D		0.001	-0.001	0.001	-0.001	0.000	0.000	0.000
m709 D'		1.000	0.997	1.000	0.998	0.972	1.000	0.858
m709 Corr.		0.061	-0.035	0.056	-0.035	-0.025	-0.005	-0.019
m709 R ²		0.004	0.001	0.003	0.001	0.001	0.000	0.000
LD X ²		9.000	3.000	7.000	3.000	1.000	0.000	0.000
P-value		0.00215	0.0731	0.00541	0.0696	0.257	0.678	0.388
m709 LOD		2.044	0.697	1.680	0.715	0.279	0.038	0.162
m709 n		2011	2011	2011	2011	2011	2011	2011
m654 D			-0.133	0.214	-0.135	-0.081	-0.004	-0.071

...

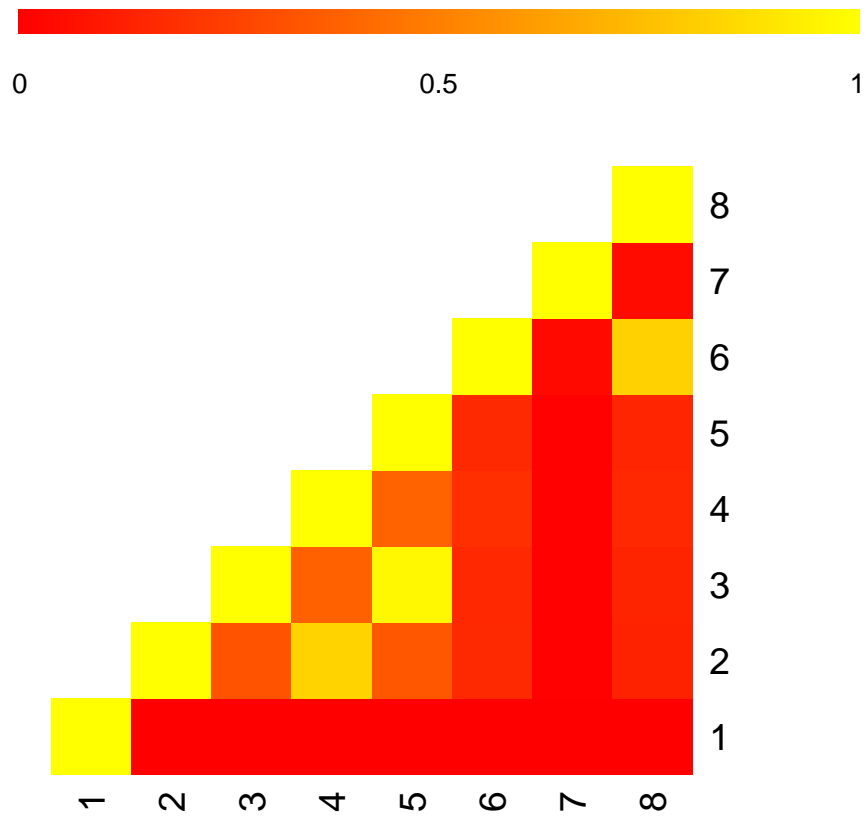
- Linkage disequilibrium, matrix plot

```
> plot(ld)
```



- Linkage disequilibrium, graphical view using LDView

```
> r2 <- t(ld@"R^2")  
> diag(r2) <- 1  
> LDView(r2, labRow = markerNames(CAMP))
```



3 Generating tables for inclusion in reports

To make it simple to include the summary tables in written reports, they can be written to files in a variety of formats, including plain text, html, and LaTeX.

3.1 Plain text files

```
> aS <- alleleSummary(CAMP)
> txt(aS, file = "CAMP_alleleSummary.txt")
```

3.2 LaTeX files

```
> aS <- alleleSummary(CAMP)
> latex(aS)
```

	Gene	Marker	Position	Group	Allele	Count	Freq	CI-Lower	CI-Upper
1	ALL	m709	?	ALL	1	3904	0.998	0.996	0.999
2				ALL	2	8	0.002	0.001	0.004
3									
4		m654	?	ALL	1	2491	0.638	0.623	0.654
5				ALL	2	1411	0.362	0.346	0.377
6									
7		m47	?	ALL	1	1417	0.369	0.354	0.384
8				ALL	2	2427	0.631	0.616	0.646
9									
10		p46	?	ALL	1	1557	0.401	0.385	0.416
11				ALL	2	2329	0.599	0.584	0.615
12									
13		p79	?	ALL	1	2407	0.626	0.611	0.642
14				ALL	2	1435	0.374	0.358	0.389
15									
16		p252	?	ALL	1	811	0.222	0.209	0.236
17				ALL	2	2835	0.778	0.764	0.791
18									
19		p491	?	ALL	1	3865	0.989	0.986	0.992
20				ALL	2	43	0.011	0.008	0.014
21									
22		p523	?	ALL	1	3148	0.804	0.791	0.816
23				ALL	2	768	0.196	0.184	0.209
24									

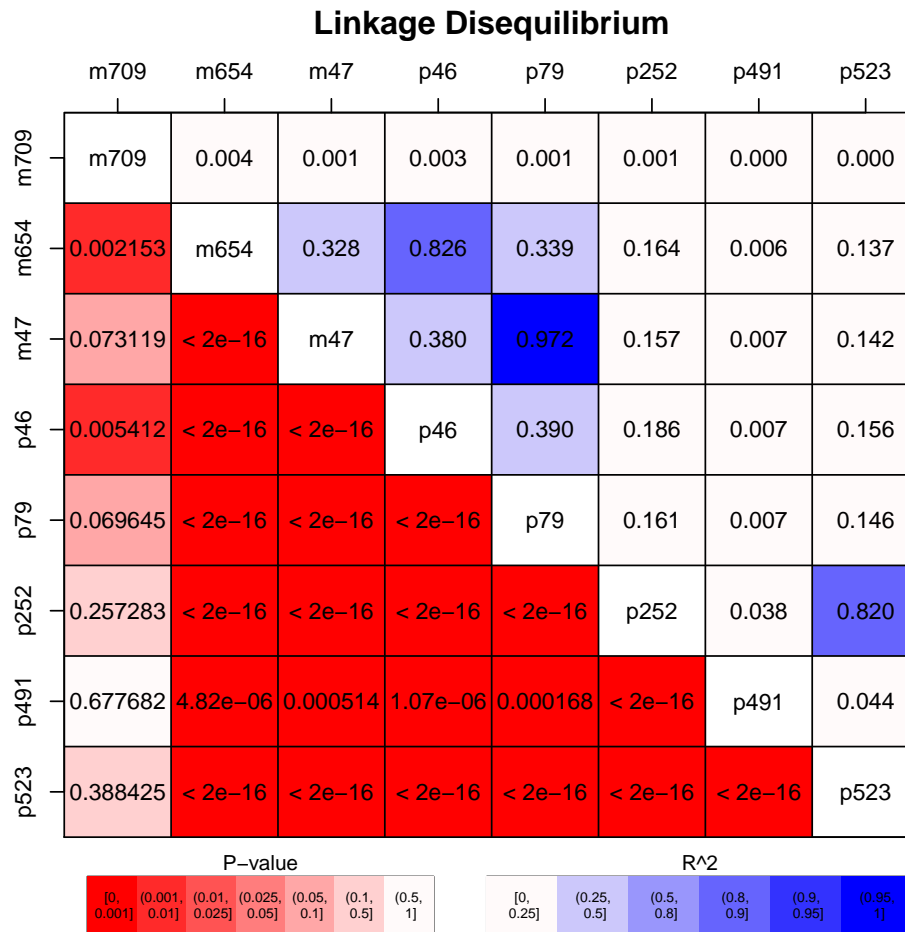

```
> gs <- genotypeSummary(CAMP[-2, ])
> latex(gs)
```

	Gene	Marker	Position	Group	Genotype	Count	Freq	CI-Lower	CI-Upper	Expected	Obs-Exp	HWE X ²	P-value
1	?	m709	?	ALL	1/1	1948	0.996	0.993	0.998	1948.008	-0.008	0.008	1
2					1/2	8	0.004	0.002	0.007	7.984	0.016		
3					2/2	0	0.000			0.008	-0.008		
4					NA	55							
5													
6	?	m47	?	ALL	1/1	276	0.144	0.128	0.159	261.172	14.828	2.112	0.144
7					1/2	865	0.450	0.428	0.472	894.656	-29.656		
8					2/2	781	0.406	0.384	0.428	766.172	14.828		
9					NA	89							
10													
11	?	p46	?	ALL	1/1	328	0.169	0.152	0.186	311.921	16.079	2.308	0.127
12					1/2	901	0.464	0.442	0.486	933.158	-32.158		
13					2/2	714	0.367	0.346	0.389	697.921	16.079		
14					NA	68							
15													
16	?	p79	?	ALL	1/1	774	0.403	0.381	0.425	753.989	20.011	3.807	0.0487
17					1/2	859	0.447	0.425	0.470	899.023	-40.023		
18					2/2	288	0.150	0.134	0.166	267.989	20.011		
19					NA	90							
20													
21	?	p252	?	ALL	1/1	107	0.059	0.048	0.070	90.198	16.802	5.177	0.0265
22					1/2	597	0.327	0.306	0.349	630.605	-33.605		
23					2/2	1119	0.614	0.591	0.636	1102.198	16.802		
24					NA	188							
25													
26	?	p491	?	ALL	1/1	1911	0.978	0.971	0.984	1911.237	-0.237	0.242	1
27					1/2	43	0.022	0.016	0.029	42.527	0.473		
28					2/2	0	0.000			0.237	-0.237		
29					NA	57							
30													
31	?	p523	?	ALL	1/1	1274	0.651	0.630	0.672	1265.309	8.691	1.552	0.234
32					1/2	600	0.306	0.286	0.327	617.381	-17.381		
33					2/2	84	0.043	0.034	0.052	75.309	8.691		
34					NA	53							
35													

```
> ld <- LD(CAMP)
> latex(ld)
```

	m709	m654	m47	p46	p79	p252	p491	p523
m709 D	0.001	-0.001	0.001	-0.001	0.000	0.000	0.000	0.000
m709 D'	1.000	0.997	1.000	0.998	0.972	1.000	0.858	
m709 Corr.	0.061	-0.035	0.056	-0.035	-0.025	-0.005	-0.019	
m709 \$R^2\$	0.004	0.001	0.003	0.001	0.001	0.000	0.000	
LD \$\chi^2\$	9	3	7	3	1	0	0	
P-value	0.002153	0.073119	0.005412	0.069645	0.257283	0.677682	0.388425	
m709 LOD	2.044	0.697	1.680	0.715	0.279	0.038	0.162	
m709 n	2011	2011	2011	2011	2011	2011	2011	
m654 D		-0.133	0.214	-0.135	-0.081	-0.004	-0.071	
m654 D'		0.992	0.978	0.997	0.994	1.000	0.994	
m654 Corr.		-0.572	0.909	-0.582	-0.405	-0.080	-0.370	
m654 \$R^2\$		0.328	0.826	0.339	0.164	0.006	0.137	
LD \$\chi^2\$		965	2727	1040	468	20	416	
P-value		< 2e-16	< 2e-16	< 2e-16	< 2e-16	4.82e-06	< 2e-16	
m654 LOD		209.756	592.178	226.003	101.725	4.540	90.541	
m654 n		2011	2011	2011	2011	2011	2011	
m47 D			-0.146	0.230	-0.080	-0.004	-0.072	
m47 D'			0.991	0.991	0.966	0.998	0.987	
m47 Corr.			-0.617	0.986	-0.396	-0.082	-0.376	
m47 \$R^2\$			0.380	0.972	0.157	0.007	0.142	
LD \$\chi^2\$			1138	3462	413	12	395	
P-value			< 2e-16	< 2e-16	< 2e-16	0.000514	< 2e-16	
m47 LOD			247.311	751.833	89.707	2.620	85.947	
m47 n			2011	2011	2011	2011	2011	
p46 D				-0.148	-0.088	-0.004	-0.077	
p46 D'				0.993	0.980	1.000	0.983	
p46 Corr.				-0.624	-0.432	-0.086	-0.395	
p46 \$R^2\$				0.390	0.186	0.007	0.156	
LD \$\chi^2\$				1185	498	23	433	
P-value				< 2e-16	< 2e-16	1.07e-06	< 2e-16	
p46 LOD				257.333	108.232	5.169	94.059	
p46 n				2011	2011	2011	2011	
p79 D					-0.081	-0.004	-0.073	
p79 D'					0.972	0.999	1.000	
p79 Corr.					-0.401	-0.082	-0.382	
p79 \$R^2\$					0.161	0.007	0.146	
LD \$\chi^2\$					432	14	430	
P-value					< 2e-16	0.000168	< 2e-16	
p79 LOD					93.840	3.073	93.524	
p79 n					2011	2011	2011	
p252 D						0.008	0.148	
p252 D'						1.000	0.990	
p252 Corr.						0.196	0.905	
p252 \$R^2\$						0.038	0.820	
LD \$\chi^2\$						77	2088	
P-value						< 2e-16	< 2e-16	
p252 LOD						16.742	453.461	
p252 n						2011	2011	
p491 D							0.009	
p491 D'							1.000	
p491 Corr.							0.210	
p491 \$R^2\$							0.044	
LD \$\chi^2\$							88	
P-value							< 2e-16	
p491 LOD							19.116	
p491 n							2011	
p523 D								
p523 D'								
p523 Corr.								
p523 \$R^2\$								
LD \$\chi^2\$								
P-value								
p523 LOD								
p523 n								

```
> plot(ld)
```



3.3 HTML files

3.4 Graphics files

As usual, plots can be generated in any format R supports.

We can also output everything all at once to a set of files, encoded as plain text (format="print"), html (format="html"), or LaTeX (format="latex"):

```
> PGtables(CAMP, filename = "CAMP", sep = "_", format = "html")
```

```
Creating CAMP_alleleSummary.html ...
Creating CAMP_genotypeSummary.html ...
Creating CAMP_LD.html ...
Creating CAMP_LD.pdf ...
Done.
```

which creates a set of html and a PDF files in the current directory.

Figure 1: HTML allele summary table

	Gene	Marker	Position	Group	Allele	Count	Freq	CI-Lower	CI-Upper
1	ALL	m709	?	ALL	1	2534	0.998	0.997	1.000
2				ALL	2	4	0.002	0.000	0.003
3				ALL	NA	0			
4									
5		m654	?	ALL	1	1630	0.647	0.629	0.666
6				ALL	2	888	0.353	0.334	0.371
7				ALL	NA	0			
8									
9		m47	?	ALL	1	924	0.371	0.352	0.390
10				ALL	2	1564	0.629	0.610	0.648
11				ALL	NA	0			
12									
13		p46	?	ALL	1	990	0.395	0.376	0.414
14				ALL	2	1516	0.605	0.586	0.624
15				ALL	NA	0			
16									
17		p79	?	ALL	1	1556	0.625	0.607	0.644
18				ALL	2	932	0.375	0.356	0.393
19				ALL	NA	0			
20									
21		p252	?	ALL	1	546	0.231	0.214	0.247
22				ALL	2	1822	0.769	0.753	0.786
23				ALL	NA	0			
24									
25		p491	?	ALL	1	2499	0.989	0.985	0.993
26				ALL	2	27	0.011	0.007	0.015
27				ALL	NA	0			
28									
29		p523	?	ALL	1	2031	0.799	0.783	0.814
30				ALL	2	511	0.201	0.186	0.217
31				ALL	NA	0			
32									

Confidence intervals width is 95%, computed using the exact quantiles for the binomial distribution.

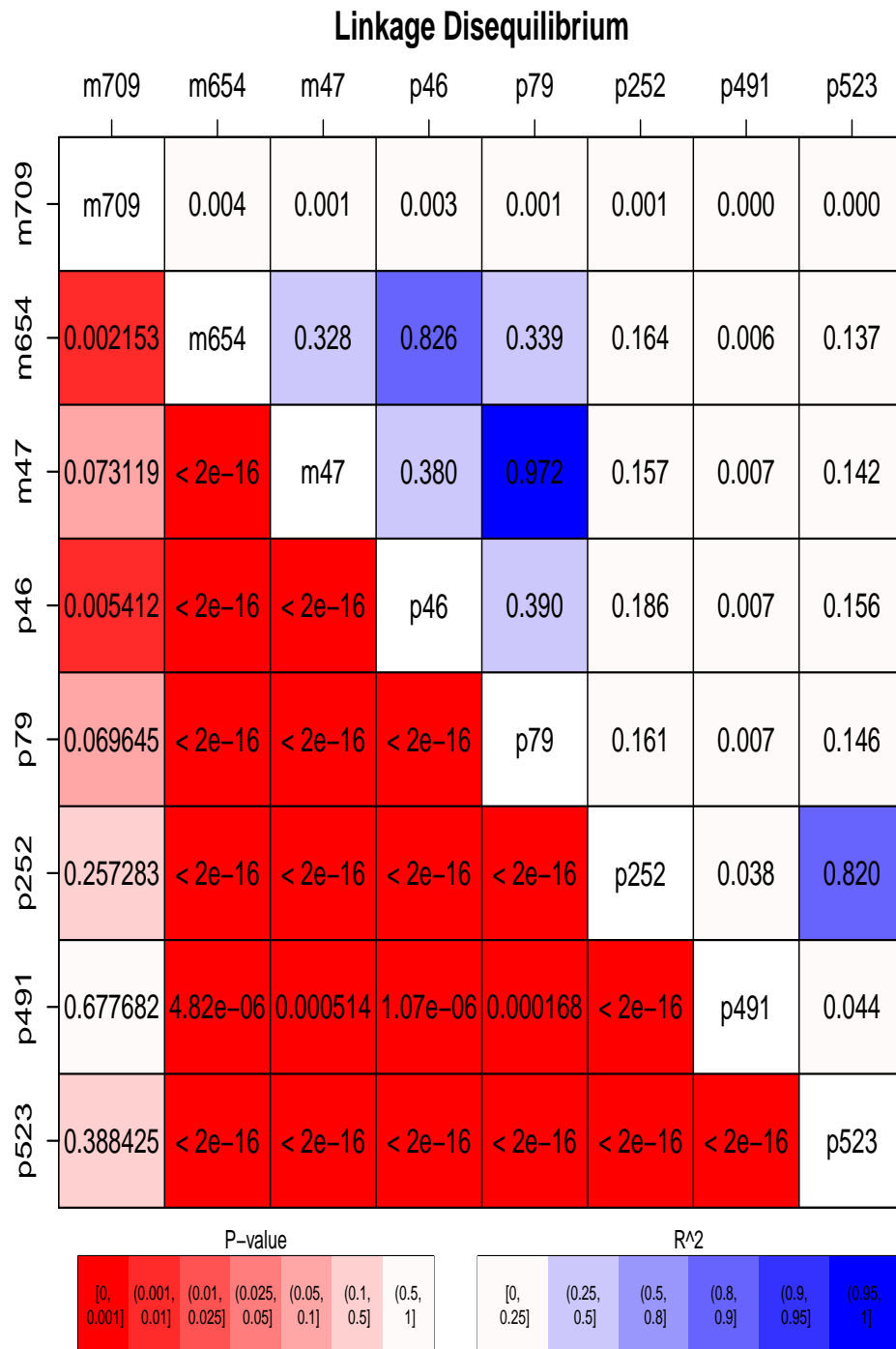
Figure 2: HTML genotype summary table

	Gene	Marker	Position	Group	Genotype	Count	Freq	CI-Lower	CI-Upper	Expected	Obs-Exp	HWE X^2	P-value
1	?	m709	?	ALL	1/1	1265	0.997	0.994	0.999	1265.003	-0.003	0.003	1
2					1/2	4	0.003	0.001	0.006	3.994	0.006		
3					2/2	0	0.000			0.003	-0.003		
4					NA	0							
5													
6	?	m654	?	ALL	1/1	536	0.426	0.399	0.453	527.581	8.419	1.080	0.323
7					1/2	558	0.443	0.416	0.471	574.837	-16.837		
8					2/2	165	0.131	0.113	0.150	156.581	8.419		
9					NA	0							
10													
11	?	m47	?	ALL	1/1	171	0.137	0.119	0.157	171.579	-0.579	0.005	0.955
12					1/2	582	0.468	0.441	0.496	580.842	1.158		
13					2/2	491	0.395	0.367	0.422	491.579	-0.579		
14					NA	0							
15													
16	?	p46	?	ALL	1/1	197	0.157	0.137	0.178	195.551	1.449	0.029	0.906
17					1/2	596	0.476	0.448	0.504	598.899	-2.899		
18					2/2	460	0.367	0.341	0.394	458.551	1.449		
19					NA	0							
20													
21	?	p79	?	ALL	1/1	488	0.392	0.365	0.420	486.563	1.437	0.030	0.905
22					1/2	580	0.466	0.439	0.494	582.875	-2.875		
23					2/2	176	0.141	0.122	0.161	174.563	1.437		
24					NA	0							
25													
26	?	p252	?	ALL	1/1	68	0.057	0.045	0.071	62.947	5.053	0.685	0.419
27					1/2	410	0.346	0.319	0.373	420.106	-10.106		
28					2/2	706	0.596	0.568	0.624	700.947	5.053		
29					NA	0							
30													
31	?	p491	?	ALL	1/1	1236	0.979	0.970	0.987	1236.144	-0.144	0.147	1
32					1/2	27	0.021	0.013	0.030	26.711	0.289		
33					2/2	0	0.000			0.144	-0.144		
34					NA	0							

Figure 3: HTML linkage disequilibrium table

	m709	m654	m47	p46	p79	p252	p491	p523
m709 D		0.001	-0.001	0.001	-0.001	0.000	0.000	0.000
m709 D'		1.000	0.999	0.999	0.999	0.047	1.000	0.151
m709 Corr.		0.055	-0.031	0.050	-0.031	0.004	-0.004	0.012
m709 R^2		0.003	0.001	0.002	0.001	0.000	0.000	0.000
LD X^2		4	2	3	2	0	0	0
P-value		0.039830	0.127371	0.069309	0.125046	0.895257	0.767100	0.702909
m709 LOD		0.917	0.505	0.716	0.511	0.004	0.019	0.032
m709 n		1303	1303	1303	1303	1303	1303	1303
m654 D			-0.130	0.212	-0.132	-0.082	-0.004	-0.070
m654 D'			0.996	0.978	0.996	0.991	1.000	0.990
m654 Corr.			-0.563	0.905	-0.570	-0.405	-0.077	-0.366
m654 R^2			0.317	0.818	0.325	0.164	0.006	0.134
LD X^2			588	1698	619	300	12	255
P-value			< 2e-16	< 2e-16	< 2e-16	< 2e-16	0.000391	< 2e-16
m654 LOD			127.716	368.786	134.516	65.332	2.730	55.408
m654 n			1303	1303	1303	1303	1303	1303
m47 D				-0.143	0.231	-0.083	-0.004	-0.075
m47 D'				0.992	0.991	0.966	0.998	1.000
m47 Corr.				-0.608	0.985	-0.407	-0.081	-0.388
m47 R^2				0.370	0.970	0.166	0.007	0.150
LD X^2				699	2213	268	7	273
P-value				< 2e-16	< 2e-16	< 2e-16	0.007251	< 2e-16
m47 LOD				151.847	480.561	58.250	1.566	59.344
m47 n				1303	1303	1303	1303	1303
p46 D					-0.145	-0.090	-0.004	-0.078
p46 D'					0.993	0.985	1.000	0.991
p46 Corr.					-0.615	-0.439	-0.085	-0.400
p46 R^2					0.379	0.192	0.007	0.160
LD X^2					719	335	14	282
P-value					< 2e-16	< 2e-16	0.000113	< 2e-16
p46 LOD					156.342	72.789	3.237	61.413
p46 n					1303	1303	1303	1303
n79 D						-0.083	-0.004	-0.075

Figure 4: Linkage disequilibrium plot



4 Subsetting by Group

The `alleleSummary` and `genotypeSummary` functions also allow you to create tables which show the summary information separated out by a grouping variable, which must be discrete “factor” variables (such as Sex).

To accomplish this, add the argument `by=Sex` to the function call. For example:

```
> alleleSummary(CAMP, by = "sex")
```

Gene	Marker	Position	Group	Allele	Count	Freq	CI-Lower	CI-Upper
ALL	m709	?	1	1	3904	0.998	0.996	0.999
			1	2	8	0.002	0.001	0.004
			2	1	3904	0.998	0.996	0.999
			2	2	8	0.002	0.001	0.004
	m654	?	1	1	2491	0.638	0.623	0.654
			1	2	1411	0.362	0.346	0.377
			2	1	2491	0.638	0.623	0.654
			2	2	1411	0.362	0.346	0.377
	m47	?	1	1	1417	0.369	0.354	0.384
			1	2	2427	0.631	0.616	0.646
			2	1	1417	0.369	0.354	0.384
			2	2	2427	0.631	0.616	0.646
	p46	?	1	1	1557	0.401	0.385	0.416
			1	2	2329	0.599	0.584	0.615
			2	1	1557	0.401	0.385	0.416
			2	2	2329	0.599	0.584	0.615
	p79	?	1	1	2407	0.626	0.611	0.642
			1	2	1435	0.374	0.358	0.389
			2	1	2407	0.626	0.611	0.642
			2	2	1435	0.374	0.358	0.389
	p252	?	1	1	811	0.222	0.209	0.236
			1	2	2835	0.778	0.764	0.791
			2	1	811	0.222	0.209	0.236
			2	2	2835	0.778	0.764	0.791
	p491	?	1	1	3865	0.989	0.986	0.992
			1	2	43	0.011	0.008	0.014
			2	1	3865	0.989	0.986	0.992
			2	2	43	0.011	0.008	0.014
	p523	?	1	1	3148	0.804	0.791	0.816

1	2	768	0.196	0.184	0.209
2	1	3148	0.804	0.791	0.816
2	2	768	0.196	0.184	0.209

Footer:

Confidence intervals width is 95%, computed using
the exact quantiles for the binomial
distribution.

This will display a table within a separate block within each marker for each level of the variable **Sex**.

To control whether the summary table for entire data in addition to individual factor levels, add `includeOverall=TRUE` or `includeOverall=FALSE` (the default) as appropriate.

A Example R script

```
> library(GeneticsBase)
> data(CAMP)
> PGtables(CAMP, filename = "test", format = "html")
> PGtables(CAMP, filename = "test", format = "latex")
```

References

- Warnes GR. “The Genetics Package: Utilities for handling genetic data” *R News*, Volume 3, Issue 1, June 2003.
- Warnes GR. “genetics”, a package for handling marker-based genetic data within the open-source statistical package “R”. The package includes function to compute allele frequencies, use genetic markers in statistical models, estimate disequilibrium, and test for departure from Hardy-Weinberg equilibrium.
<http://cran.us.r-project.org/src/contrib/PACKAGES.html#genetics>, 2002-