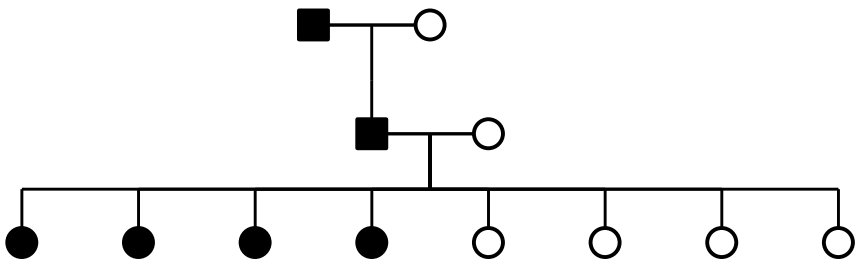
Parametric Multipoint Linkage Analysis Exercises with GENEHUNTER

Exercise 1

In the previous exercises the LINKMAP program of LINKAGE was used to perform multipoint linkage analysis. The problem was encountered that not all markers could be analyzed simultaneously, but instead a sliding window of markers had to be used to analyze the data. The number of markers that can be analyzed with LINKMAP decreases with increasing number of alleles at each locus. The limiting factor is Maxfac which is the product of the number of alleles at each locus (number of alleles_disease locus X number of alleles_marker1 X … number of alleles_marker N). For this exercise the GENHUNTER program will be used to perform the analysis.

For the multipoint LINKMAP exercises the datafile **MULTDIS1.DAT** was provided. This file will also be used for the GENEHUNTER exercises. The same pedigree data will not be analyzed, since this is too time concerning to be carried out within the realm of the course. Please note that it would not be possible to analyze the two pedigrees shown on pg 127 (chapter 17) using GENEHUNTER since maxbit for each pedigree is 2(21)-5=37. The pedigrees would have to be divided up into multiple pedigrees in order to perform the analysis with the GENEHUNTER program.

A pedigree file is provided which is called **md1.pre**. This file contains two autosomal dominant pedigrees with the structure shown below. Note that for GENEHUNTER pre makeped files must be used.



What is maxbit for this pedigree _____?

The datafile **MULTDIS1.DAT** must be edited. Below is the unedited file that was provided.

```
 9 0 0 5  << NO. OF LOCI, RISK LOCUS, SEXLINKED (IF 1) PROGRAM
 0 0.0 0.0 0  << MUT LOCUS, MUT RATE, HAPLOTYPE FREQUENCIES (IF 1)
  1  2  3  4  5  6  7  8  9
1   2  << AFFECTION, NO. OF ALLELES
 0.999900 0.000100   << GENE FREQUENCIES
 1 << NO. OF LIABILITY CLASSES
 0.0000 1.0000 1.0000 << PENETRANCES
3   2  << ALLELE NUMBERS, NO. OF ALLELES
 0.500000 0.500000   << GENE FREQUENCIES
3   2  << ALLELE NUMBERS, NO. OF ALLELES
 0.500000 0.500000   << GENE FREQUENCIES
3   2  << ALLELE NUMBERS, NO. OF ALLELES
 0.500000 0.500000   << GENE FREQUENCIES
3   2  << ALLELE NUMBERS, NO. OF ALLELES
 0.500000 0.500000   << GENE FREQUENCIES
3   2  << ALLELE NUMBERS, NO. OF ALLELES
 0.500000 0.500000   << GENE FREQUENCIES
3   2  << ALLELE NUMBERS, NO. OF ALLELES
 0.500000 0.500000   << GENE FREQUENCIES
3   2  << ALLELE NUMBERS, NO. OF ALLELES
 0.500000 0.500000   << GENE FREQUENCIES
3   2  << ALLELE NUMBERS, NO. OF ALLELES
 0.500000 0.500000   << GENE FREQUENCIES
 0 0  << SEX DIFFERENCE, INTERFERENCE (IF 1 OR 2)
 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 0.10000 << RECOMBINATION VALUES
 1 0.10000 0.45000 << REC VARIED, INCREMENT, FINISHING VALUE
```

A name has been given for each marker and the disease locus. Notice each marker and the disease locus was given a name. In addition the comment <<<RECOMINATION VALUES has been removed from the file. The order of the disease locus (1) and markers on the map is 1-2-3-4-5-6-7-8. The recombination fraction between the disease locus (1) and marker 1 (locus 2) is 0.5 and recombination fraction between markers 1 and 2 (loci 2 and 3) is 0.075; the same recombination fraction separates markers 2 and 3, 4 and 5, 5 and 6, 6 and 7, 7 and 8, 8 and 9. The recombination fraction between markers 3 and 4 is 0.225. Edit recombination fractions between loci in accordance with this information. Rename the datafile to **datain.dat**.

```
 9 0 0 5  << NO. OF LOCI, RISK LOCUS, SEXLINKED (IF 1) PROGRAM
 0 0.0 0.0 0  << MUT LOCUS, MUT RATE, HAPLOTYPE FREQUENCIES (IF 1)
  1  2  3  4  5  6  7  8  9
1   2  # disease
 0.999900 0.000100   << GENE FREQUENCIES
 1 << NO. OF LIABILITY CLASSES
 0.0000 1.0000 1.0000 << PENETRANCES
3   2  # marker1
 0.500000 0.500000   << GENE FREQUENCIES
3   2  # marker2
 0.500000 0.500000   << GENE FREQUENCIES
3   2  # marker3
 0.500000 0.500000   << GENE FREQUENCIES
3   2  # marker4
 0.500000 0.500000   << GENE FREQUENCIES
3   2  # marker5
 0.500000 0.500000   << GENE FREQUENCIES
3   2  # marker6
 0.500000 0.500000   << GENE FREQUENCIES
3   2  # marker7
 0.500000 0.500000   << GENE FREQUENCIES
3   2  # marker8
 0.500000 0.500000   << GENE FREQUENCIES
 0 0  << SEX DIFFERENCE, INTERFERENCE (IF 1 OR 2)
 0.50000 0.07500 0.07500 0.22500 0.07500 0.07500 0.07500 0.07500
 1 0.10000 0.45000 << REC VARIED, INCREMENT, FINISHING VALUE
```

It is now time to run the GENEHUNTER program. It is invoked by typing **gh** at the prompt.

> gh

In order that the session can be recorded (a file created with everything that appears on the screen) the photo command is given with the output name for the data file. Name the output file **mult.out**.

> photo mult.out

To obtain post script files with the output of the results in graphical format give the following command. For this exercise four post script files will be made. Two post-script files will contain the haplotype data for the two pedigree drawings (default names 1.ps and 2.ps). Another postscript file will be made the graphically displays the LOD scores (default name lod_plot) and the fourth postscript file will graphically display the information contents (default name info.content.ps). Later you will be given the option of either maintaining the default names or providing new names for the postscript files.

> ps on

Set the max bit to the lowest value that can be used for these pedigrees.

>mb 15

The datafile must be loaded using the load command.

> load datain.dat

Have the map distances reported using the haldane map function. GENEHUNTER will convert recombination fractions to map distances using either the Haldane or Kosambi map function.

> map function haldane

To set the number of increments at which the lod score will be calculated between each marker give the increment command. For this exercise the number of increments will be set to 5 (thus, lod score will be calculated at marker points and in 4 equally-spaced points between markers).

> increment steps 5

To calculate lod scores off the map markers use the off end command. For this exercise calculate lod scores an additional 10 cM at either end of the map of markers.

> off end 10

By turning the haplo switch on GENEHUNTER will reconstruct haplotypes for the two pedigrees. The haplotypes with can be found in a file called haplo.dump

> haplo on

For this exercise only the parametric lod scores will be calculated. In order to perform the analysis the command **analysis lod** is given.

> analysis lod

The pedigree data needs to be scanned.

> scan md1.pre

GENEHUNTER now begins to analyze the pedigrees. It first analyzes the first pedigree and then the second pedigree. In order to obtain totals for the lod scores for both pedigrees the total stat command must be run. It will ask for you to enter the names for the postscript files. The default file names lod_plot.ps and info_content.ps can be used

> total stat

To calculate two-point lod score at every marker position (which is equivalent to using MLINK with zero recombination between marker and trait locus), turn on the "single point" option:

> single point on

Now the pedigree data needs to be re-scanned

> scan md1.pre

In order to obtain totals of two-point lod scores for the pedigrees, run the "total stat" command:

> total stat

Questions

1.) What is the maximum lod score _____and at what map position did it occur _____?

2.) How large of a genetic interval does the disease locus map to based upon a 3-unit support interval _____, and a 1-unit support interval _____?

3.) What is maximum two-point lod score _____ and at which marker did it occur _____?

Exercise 2 – GENEHUNTER Linkage Admixture/Heterogeneity.

Three pedigrees have been collected for disease X with the same pedigree structure shown in exercise 1. The pedigree data file is **mdhet.pre** and the parameter file is **datain.dat.** Note that this is the same parameter file that was used for exercises 1 and 2.

The analysis will be carried out in a very similar fashion to Exercise 1, however, now you will test for linkage in the presence of admixture.

Invoke GENEHUNTER by typing **gh** at the prompt.

> gh

In order that the session is recorded (a file made with everything that appears on the screen) the photo command is given with the output name for the data file. Name the output file **multhet.out**.

> photo multhet.out

To obtain post script files with the output in graphical format give the following command.

> ps on

Set the max bit to the lowest value that can be used for these pedigrees.

>mb 15

The parameter file must be loaded using the load command.

> load datain.dat

Use the haldane map function

> map function haldane

Set the number of increment where the lod score will be calculated to 5.

> increment steps 5

To calculate lod scores off the map markers use the off end command. For this exercise calculate lod scores an additional 10 cM at either end of the map of markers.

> off end 10

In order to perform the analysis the command **analysis lod** is given.

> analysis lod

The pedigree data needs to be scanned.

> scan mdhet.pre

GENEHUNTER now begins to analyze the pedigrees. In order to obtain totals for the lod scores for the three pedigrees and perform the admixture test the **total stat het** command must be run.

It will ask for you to enter the names for the postscript files. Name the post script files lod_plot_het.ps and info_content_het.ps.

> total stat het

To calculate two-point lod score at every marker position (which is equivalent to using MLINK with zero recombination between marker and trait locus), turn on the "single point" option:

> single point on

Now the pedigree data needs to be re-scanned

> scan mdhet.pre

In order to obtain totals of two-point lod scores for the pedigrees, run the "total stat" command:

> total stat het

Questions

1.) What is the maximum multipoint lod score under linkage homogeneity_____and at which position did it occur _____?

2.) What is the maximum multipoint lod score when testing for linkage allowing for admixture _____ and at what maps distance did it occur _____?

3.) What is the estimate for alpha_____?

4.) What is the maximum two-point lod score when testing for linkage allowing for admixture _____ and at what marker did it occur _____?

5.) What is the estimate for alpha when performing 2-point analysis _____?

6.) Do you suspect that at least one additional disease locus is responsible for this phenotype _____?

RESULTS

Exercises 1

What is maxbit for this pedigree _15_?

Output from GENEHUNTER

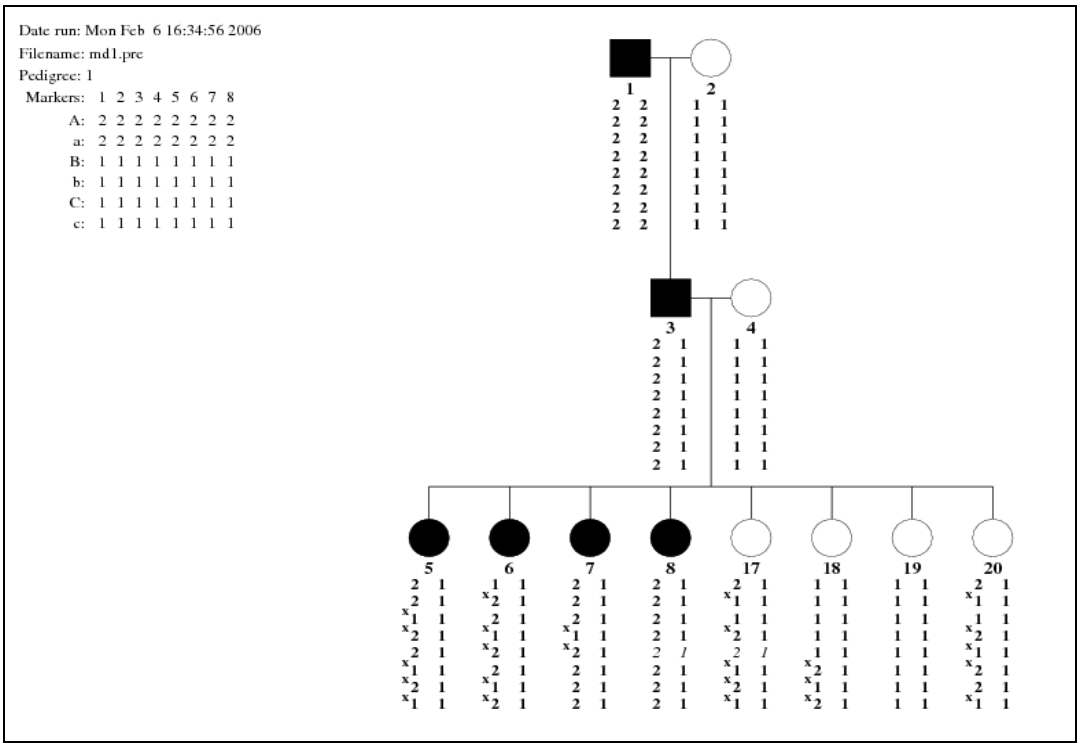
```
npl:10> total stat
position LOD_score   information
 -10.00  -0.850927    0.299422
  -8.00  -1.206380    0.330446
  -6.00  -1.699790    0.366105
  -4.00  -2.446769    0.407946
  -2.00  -3.816244    0.459083
   0.00  -INFINITY    0.533333
   1.63  -0.959843    0.432243
   3.25   0.416398    0.397071
   4.88   1.119999    0.397071
   6.50   1.446105    0.432243
   8.13  -INFINITY    0.533333
   9.75   1.735805    0.503284
  11.38   1.783713    0.492474
  13.00   1.607813    0.492474
  14.63   1.134318    0.503284
  16.25  -INFINITY    0.533333
  22.23   2.884113    0.332020
  28.21   2.827617    0.263013
  34.19   2.307021    0.263013
  40.17   1.100891    0.332020
  46.14  -INFINITY    0.533333
  47.77   1.018437    0.415741
  49.39   2.517078    0.371308
  51.02   3.394301    0.365018
  52.64   4.019040    0.395972
  54.27   4.506913    0.494320
  55.89   4.194369    0.395972
  57.52   3.760754    0.365018
  59.15   3.116853    0.371308
  60.77   1.969636    0.415741
  62.40  -INFINITY    0.533333
  64.02  -0.551490    0.384883
  65.65   0.149821    0.333469
  67.27   0.149821    0.333469
  68.90  -0.551490    0.384883
  70.52  -INFINITY    0.533333
  72.15   1.727238    0.361202
  73.77   2.553221    0.301668
  75.40   2.729121    0.301668
  77.02   2.328725    0.361202
  78.65  -INFINITY    0.533333
  80.65  -0.418188    0.459083
  82.65   0.349574    0.407946
  84.65   0.744949    0.366105
  86.65   0.989290    0.330446
  88.65   1.151962    0.299422

npl:13> total stat
position LOD_score   information
marker1  -INFINITY    0.533333
marker2  -INFINITY    0.533333
marker3  -INFINITY    0.533333
marker4  -INFINITY    0.533333
marker5   3.612360    0.400000
marker6  -INFINITY    0.533333
marker7  -INFINITY    0.533333
marker8  -INFINITY    0.533333
```

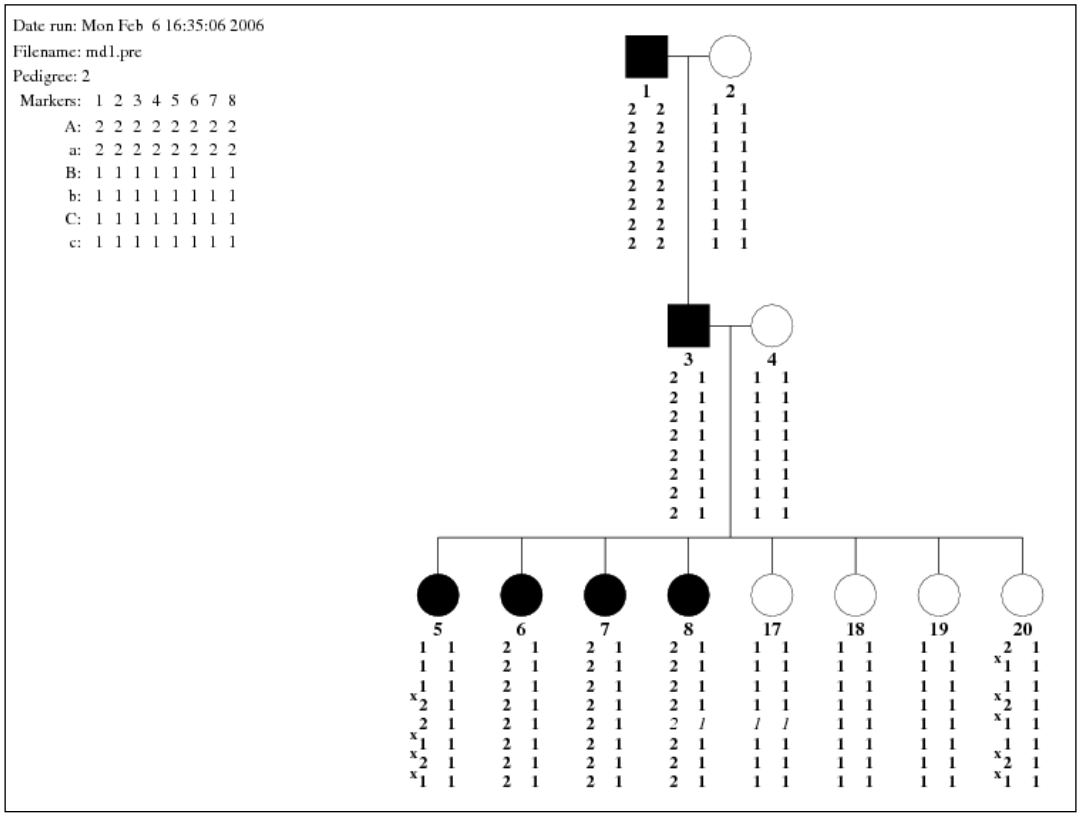
output haplo.dump

```
*****  1  3.475
1     0     0     2     2   2   2   2   2   2   2   2
                        2   2   2   2   2   2   2   2
2     0     0     1     1   1   1   1   1   1   1   1
                        1   1   1   1   1   1   1   1
4     0     0     1     1   1   1   1   1   1   1   1
                        1   1   1   1   1   1   1   1
3     1     2     2     2   2   2   2   2   2   2   2
                        1   1   1   1   1   1   1   1
5     3     4     2     2   2   1   2   2   1   2   1
                        1   1   1   1   1   1   1   1
6     3     4     2     1   2   2   1   2   2   1   2
                        1   1   1   1   1   1   1   1
7     3     4     2     2   2   2   1   2   2   2   2
                        1   1   1   1   1   1   1   1
8     3     4     2     2   2   2   2   2   2   2   2
                        1   1   1   1   1   1   1   1
17    3     4     1     2   1   1   2   2   1   2   1
                        1   1   1   1   1   1   1   1
18    3     4     1     1   1   1   1   1   2   1   2
                        1   1   1   1   1   1   1   1
19    3     4     1     1   1   1   1   1   1   1   1
                        1   1   1   1   1   1   1   1
20    3     4     1     2   1   1   2   1   2   2   1
                        1   1   1   1   1   1   1   1
*****  2  3.475
1     0     0     2     2   2   2   2   2   2   2   2
                        2   2   2   2   2   2   2   2
2     0     0     1     1   1   1   1   1   1   1   1
                        1   1   1   1   1   1   1   1
4     0     0     1     1   1   1   1   1   1   1   1
                        1   1   1   1   1   1   1   1
3     1     2     2     2   2   2   2   2   2   2   2
                        1   1   1   1   1   1   1   1
5     3     4     2     1   1   1   2   2   1   2   1
                        1   1   1   1   1   1   1   1
6     3     4     2     2   2   2   2   2   2   2   2
                        1   1   1   1   1   1   1   1
7     3     4     2     2   2   2   2   2   2   2   2
                        1   1   1   1   1   1   1   1
8     3     4     2     2   2   2   2   2   2   2   2
                        1   1   1   1   1   1   1   1
17    3     4     1     1   1   1   1   1   1   1   1
                        1   1   1   1   1   1   1   1
18    3     4     1     1   1   1   1   1   1   1   1
                        1   1   1   1   1   1   1   1
19    3     4     1     1   1   1   1   1   1   1   1
                        1   1   1   1   1   1   1   1
20    3     4     1     2   1   1   2   1   1   2   1
                        1   1   1   1   1   1   1   1
```

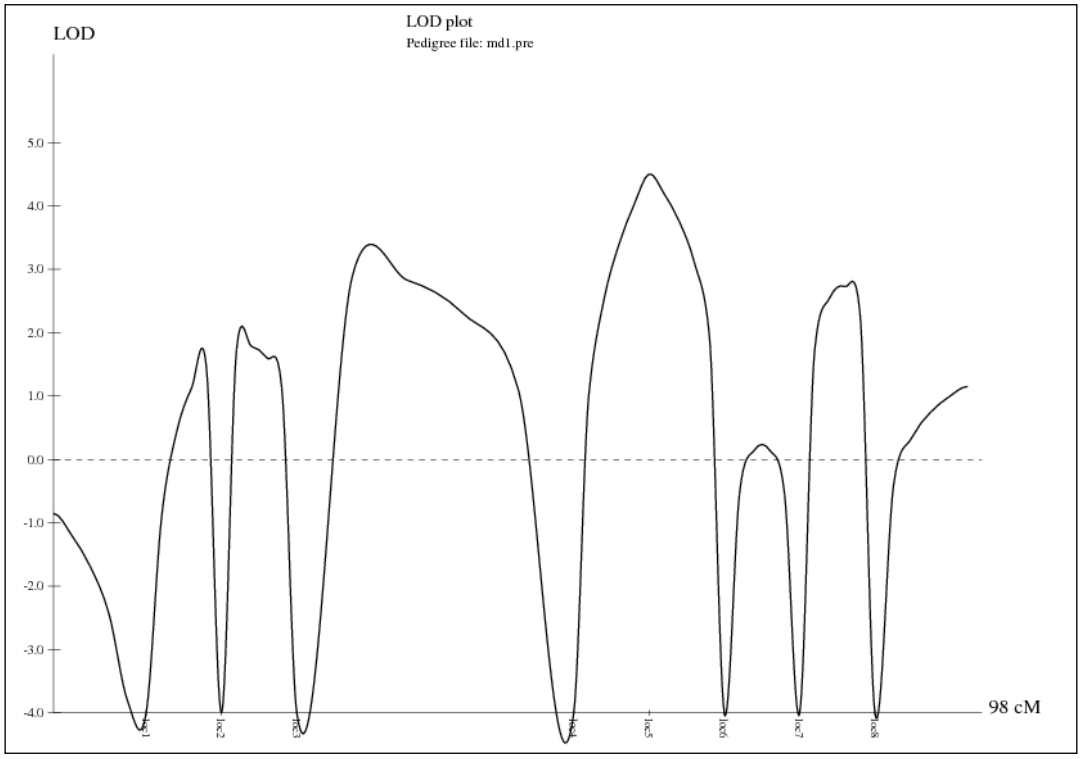
Postscript file showing haplotypes for pedigree 1



Postscript file displaying the haplotypes for pedigree 2

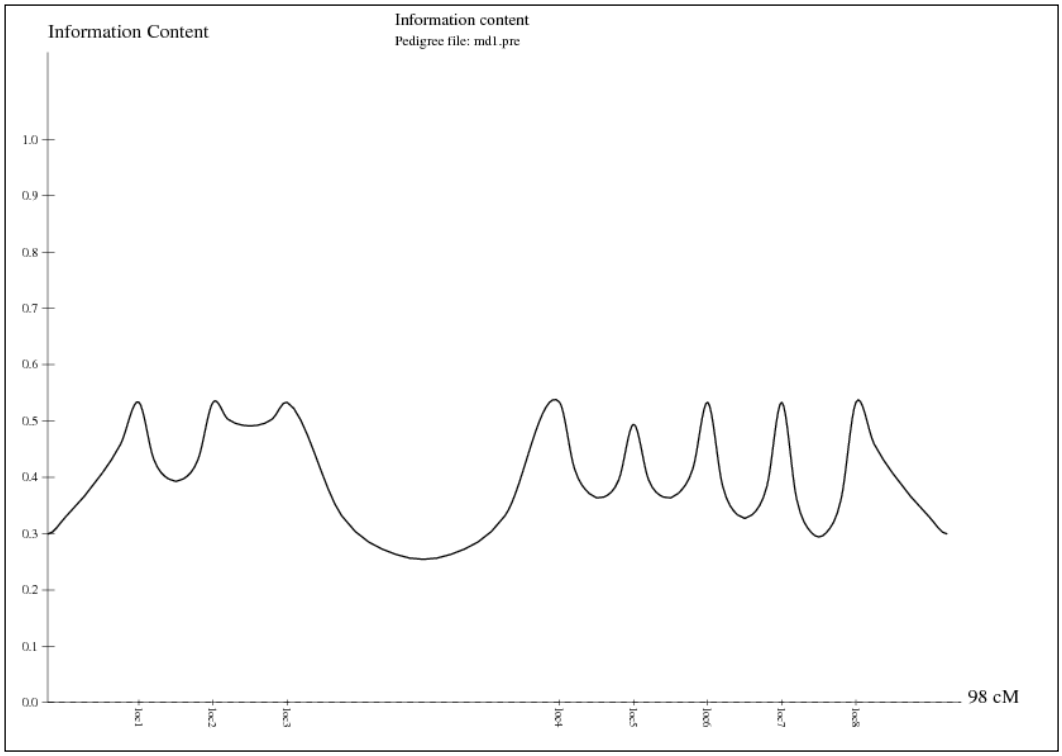


LOD score plot – multipoint analysis of pedigrees 1 and 2



\

Information content plot for families 1 and 2



Answers to questions for exercise 1

1.) What is the maximum lod score __4.51__ and at what map position did it occur _54.27__?
2.) How large of a genetic interval does the disease locus map to based upon a 3-unit support interval 70.52 cM__ and a 1-unit support interval 8.13 cM_? Note for a more accurate support interval use more increments. It is should be noted that the 3-unit support interval is disjointed and those regions with a LOD score < 1.5 are not included in the 3-unit support interval (e.g. those positions where the LOD score goes to –infinity).
3.) What is maximum two-point lod score __3.61_ and at which marker did it occur __5__? Only the lod scores at theta=0.0 are reported. It is possible that a significant LOD score could be missed if it occurred at a theta value other than 0.0.

Exercies 2

Results

```
npl:10> total stat het
position LOD_score  (alpha, HLOD)     information
 -10.00  -2.778557 (0.0000, 0.0000)     0.299422
  -8.00  -3.456323 (0.0000, 0.0000)     0.330446
  -6.00  -4.383239 (0.0000, 0.0000)     0.366105
  -4.00  -5.767410 (0.0000, 0.0000)     0.407946
  -2.00  -8.272908 (0.0000, 0.0000)     0.459083
   0.00  -INFINITY (0.0000, 0.0000)     0.533333
   1.63  -8.183665 (0.0000, 0.0000)     0.448030
   3.25  -5.979609 (0.3176, 0.4664)     0.418272
   4.88  -5.100108 (0.3632, 0.9767)     0.418272
   6.50  -5.176230 (0.3656, 1.3390)     0.448030
   8.13  -INFINITY (0.3307, 1.5823)     0.533333
   9.75  -4.789963 (0.3772, 1.5421)     0.511177
  11.38  -4.215155 (0.4004, 1.4468)     0.503074
  13.00  -4.391055 (0.3982, 1.2731)     0.503074
  14.63  -5.391450 (0.3673, 0.9512)     0.511177
  16.25  -INFINITY (0.0000, 0.0000)     0.533333
  22.23   0.757859 (0.6585, 2.0778)     0.352766
  28.21   1.211387 (0.6660, 2.0400)     0.290564
  34.19   0.690791 (0.6313, 1.5595)     0.290564
  40.17  -1.025363 (0.3765, 0.8458)     0.352766
  46.14  -INFINITY (0.0000, 0.0000)     0.533333
  47.77  -3.228603 (0.3559, 0.9366)     0.437028
  49.39  -1.078391 (0.6101, 1.7594)     0.401096
  51.02  -0.025268 (0.6554, 2.5801)     0.396903
  52.64   0.373487 (0.6627, 3.1952)     0.423849
  54.27  -INFINITY (0.6647, 3.6802)     0.507325
  55.89   0.548816 (0.6638, 3.3691)     0.423849
  57.52   0.341185 (0.6614, 2.9389)     0.396903
  59.15  -0.478616 (0.6519, 2.3069)     0.401096
  60.77  -2.277404 (0.5625, 1.2708)     0.437028
  62.40  -INFINITY (0.0000, 0.0000)     0.533333
  64.02  -4.895097 (0.3177, 0.8036)     0.400670
  65.65  -3.666886 (0.3261, 0.9743)     0.354670
  67.27  -3.666886 (0.3261, 0.9743)     0.354670
  68.90  -4.895097 (0.3177, 0.8036)     0.400670
  70.52  -INFINITY (0.0000, 0.0000)     0.533333
  72.15  -2.616369 (0.5223, 1.0819)     0.384883
  73.77  -1.263487 (0.6338, 1.7658)     0.333469
  75.40  -1.087587 (0.6441, 1.9288)     0.333469
  77.02  -2.014882 (0.6280, 1.5481)     0.384883
  78.65  -INFINITY (0.0000, 0.0000)     0.533333
  80.65  -3.175824 (0.1509, 0.0433)     0.459083
  82.65  -1.572896 (0.2882, 0.2139)     0.407946
  84.65  -0.716131 (0.3782, 0.3568)     0.366105
  86.65  -0.162818 (0.4667, 0.4838)     0.330446
  88.65   0.225776 (0.5536, 0.6008)     0.299422
```

```
npl:14> total stat het
position LOD_score  (alpha, HLOD)     information
marker1  -INFINITY (0.0000, 0.0000)     0.533333
marker2  -INFINITY (0.3307, 1.5823)     0.533333
marker3  -INFINITY (0.0000, 0.0000)     0.533333
marker4  -INFINITY (0.0000, 0.0000)     0.533333
marker5  -INFINITY (0.6614, 2.7899)     0.444444
marker6  -INFINITY (0.0000, 0.0000)     0.533333
marker7  -INFINITY (0.0000, 0.0000)     0.533333
marker8  -INFINITY (0.0000, 0.0000)     0.533333
```

Answers for exercise 2

1.) What is the maximum multipoint lod score under linkage homogeneity 1.21 and at which position did it occur 28.21 ?
2.) What is the maximum multipoint lod score when testing for linkage allowing for admixture 3.68 and at what maps distance did it occur 54.27 cM ?
3.) What is the estimate for alpha 0.66 ?
4.) What is the maximum two-point lod score when testing for linkage allowing for admixture 2.79 and at what marker did it occur 5 ?
5.) What is the estimate for alpha when performing 2-point analysis 0.66 ?
6.) Do you suspect that at least one additional disease locus is responsible for this phenotype yes?