Practical Session

Linkage analysis &
Homozygosity mapping

Practical session: content of the presentation

- Introduction
- Superlink SNP Online (GUI)
- Merlin (command line)
- HomozygosityMapper

Introduction

Linkage: mapping the location of *disease-causing loci* by identifying genetic markers that are co-inherited with a phenotype of interest

→ Types of information

	Superlink	Merlin
Markers	SNP file	DAT file & MAP file
Pedigree	PED file	PED file & MODEL file

→ Algorithm + computer

→ Interpretation of results

Superlink SNP Online

http://cbl-hap.cs.technion.ac.il/superlink-snp/

+ Powerful
+ User-friendly (GUI)
+/- External cloud

BIOINFORMATICS ORIGINAL PAPER

Vol. 29 no. 2 2013, pages 197–205
doi:10.1093/bioinformatics/bts658

Genetics and population analysis

Advance Access publication November 18, 2012

A system for exact and approximate genetic linkage analysis of SNP data in large pedigrees

Mark Silberstein[1,2], Omer Weissbrod[2,*], Lars Otten[3], Anna Tzemach[2], Andrei Anisenia[2,4], Oren Shtark[2], Dvir Tuberg[2], Eddie Galfrin[2], Irena Gannon[2], Adel Shalata[5,6,7], Zvi U. Borochowitz[5,8], Rina Dechter[3], Elizabeth Thompson[9] and Dan Geiger[2]

[1]Department of Computer Science, Technion-Israel Institute of Technology, Haifa 32000, Israel, [2]Department of Computer Science, University of Texas at Austin, Austin, TX 78712-0500, USA, [3]Donald Bren School of Information and Computer Sciences, UC Irvine, CA 92697-3435, USA, [4]Department of Computer Science, University of Ottawa, Ottawa, Canada K1S 0S1, [5]The Simon Winter Institute for Human Genetics, Bnai-Zion Medical Center, Haifa, 31048, Israel, [6]Research and Development Center, The Galilee Society, Shefa-Amr 20200, Israel, [7]Holy Family Hospital, Nazareth 16100, Israel, [8]The Rappaport Faculty of Medicine and Research Institute, Technion-Israel Institute of Technology, Haifa 32000, Israel and [9]Department of Statistics, University of Washington, Seattle, WA 98195-4322, USA

Associate Editor: Jeffrey Barrett

Superlink SNP Online: SNP file

HumanCytoSNP-12 BeadChip on HiScan instrument (Illumina)

Name	Chr	Position	LIN_1014_121657.Log R Ratio	LIN_1014_121657.B Allele Freq	LIN_1014_121657.GType
rs12103	1	1247494	-0,10565	0,4831263	AB
rs4970432	1	1254136	-0,08407687	0,00136925	AA
rs167405	1	1259784	-0,02384759	0,9981409	BB
rs34860204	1	1266041	-0,06988116	1	BB
rs35424002	1	1269986	0,01927868	0,9894595	BB
cnvi23177506	1	1271687	0,09490499	0,471527	NC
cnvi23177516	1	1272187	-0,2455748	0,1008989	NC
rs307372	1	1272497	0,07114677	1	NC
rs307371	1	1273116	-0,04399384	1	BB
cnvi23177549	1	1273837	0,01389528	0,9962418	NC
rs28475450	1	1274350	-0,002288754	1	BB
rs11583882	1	1274807	-0,03383686	1	BB
rs12037363	1	1275264	-0,1528965	0,008648425	AA
rs12735892	1	1275425	-0,3370034	1	BB
rs3855955	1	1276077	-0,2520254	0,5154034	AB
rs3855956	1	1276410	-0,1058069	0,01371952	AA
cnvi23177611	1	1276937	-0,1732782	0,00446967	NC

300 000 genotypes/patient

Cannot directly be used in Superlink/Merlin

→ conversion

Center for Medical Genetics

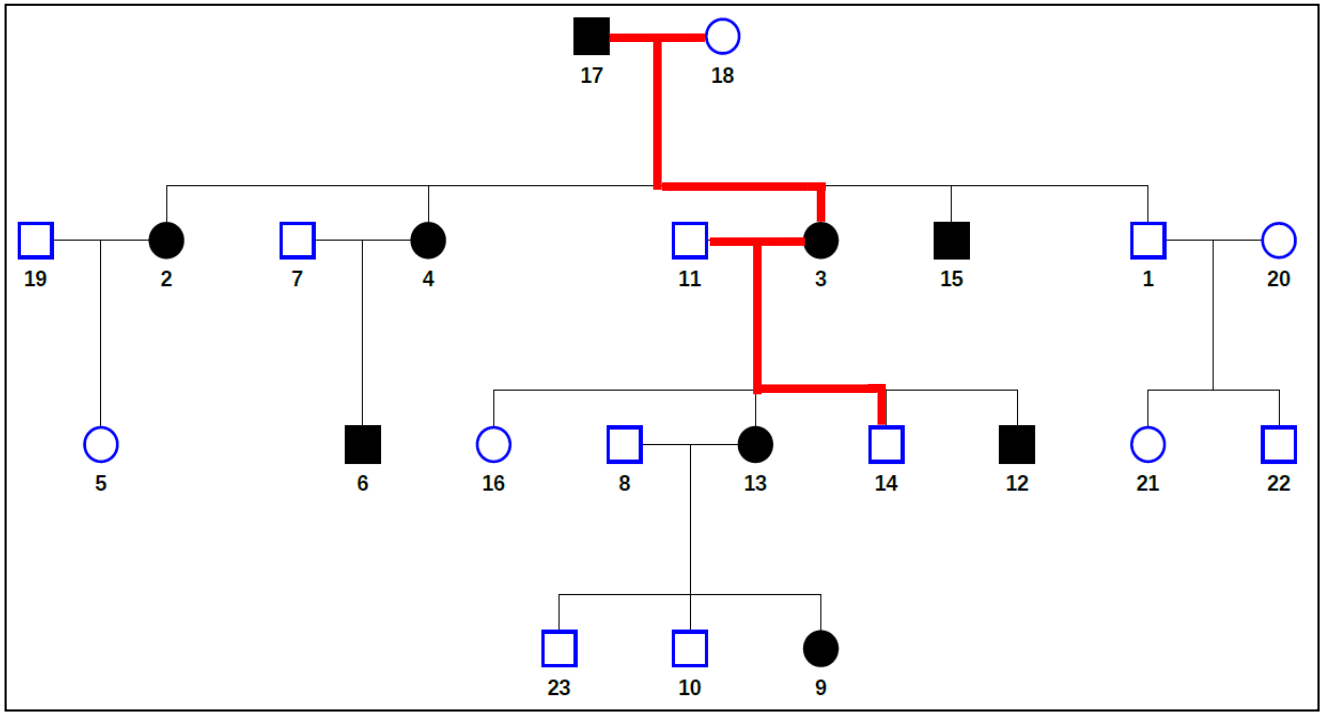
Superlink SNP Online: SNP file

Superlink SNP Online SNP file = 1 data file for whole pedigree

SNP	Chromosome	Position	HOZ_3022_13809	HOZ_3022_13810	HOZ_3022_13811	HOZ_3022_13812	HOZ_3022_13813	HOZ_3022_13814
rs4442317	1	1106784	AA	AA	AA	AA	AA	AB
rs12092254	1	1113121	BB	BB	BB	BB	BB	BB
rs6668667	1	1114668	BB	BB	BB	BB	BB	BB
rs3813204	1	1121014	AA	AA	AB	AA	AB	AB
rs4314833	1	1122024	BB	BB	AB	BB	AB	AB
rs12060422	1	1129920	BB	BB	BB	BB	BB	BB
rs9729550	1	1135242	BB	BB	AB	BB	AB	AB
rs11466681	1	1141387	BB	BB	BB	BB	BB	BB
rs34945898	1	1147024	BB	BB	BB	BB	BB	BB
rs12036216	1	1153113	BB	BB	BB	BB	BB	BB
rs3813199	1	1158277	BB	BB	BB	BB	BB	BB
rs11260562	1	1165310	BB	BB	BB	BB	BB	BB
rs7528416	1	1171249	AA	AA	AA	AA	AA	AA
rs715643	1	1172907	BB	BB	BB	BB	BB	BB
rs12093154	1	1178925	BB	BB	BB	BB	BB	BB
rs6692115	1	1186747	NC	NC	NC	NC	NC	NC
rs7524470	1	1192515	AA	AA	AA	AA	AA	AA
rs6704013	1	1197591	BB	BB	BB	BB	BB	BB
rs4018608	1	1203938	AA	AA	AB	AA	AB	AB
rs12073590	1	1205155	AA	AA	AA	AA	AA	AA
rs6689813	1	1210963	AA	AA	AB	AA	AB	AB

Conversion via Perl/Python/…

Superlink SNP Online: PED file



1	17	18	1	1
5	19	2	2	1
7	0	0	1	1
8	0	0	1	1
10	8	13	1	1
11	0	0	1	1
14	11	3	1	1
16	11	3	2	1
18	0	0	2	1
19	0	0	1	1
20	0	0	2	1
21	1	20	2	1
22	1	20	1	1
23	8	13	1	1
2	17	18	2	2
3	17	18	2	2
4	17	18	2	2
6	7	4	1	2
9	8	13	2	2
12	11	3	1	2
13	11	3	2	2
15	17	18	1	2
17	0	0	1	2

Simplified PED-format (notepad)

Column 1: id

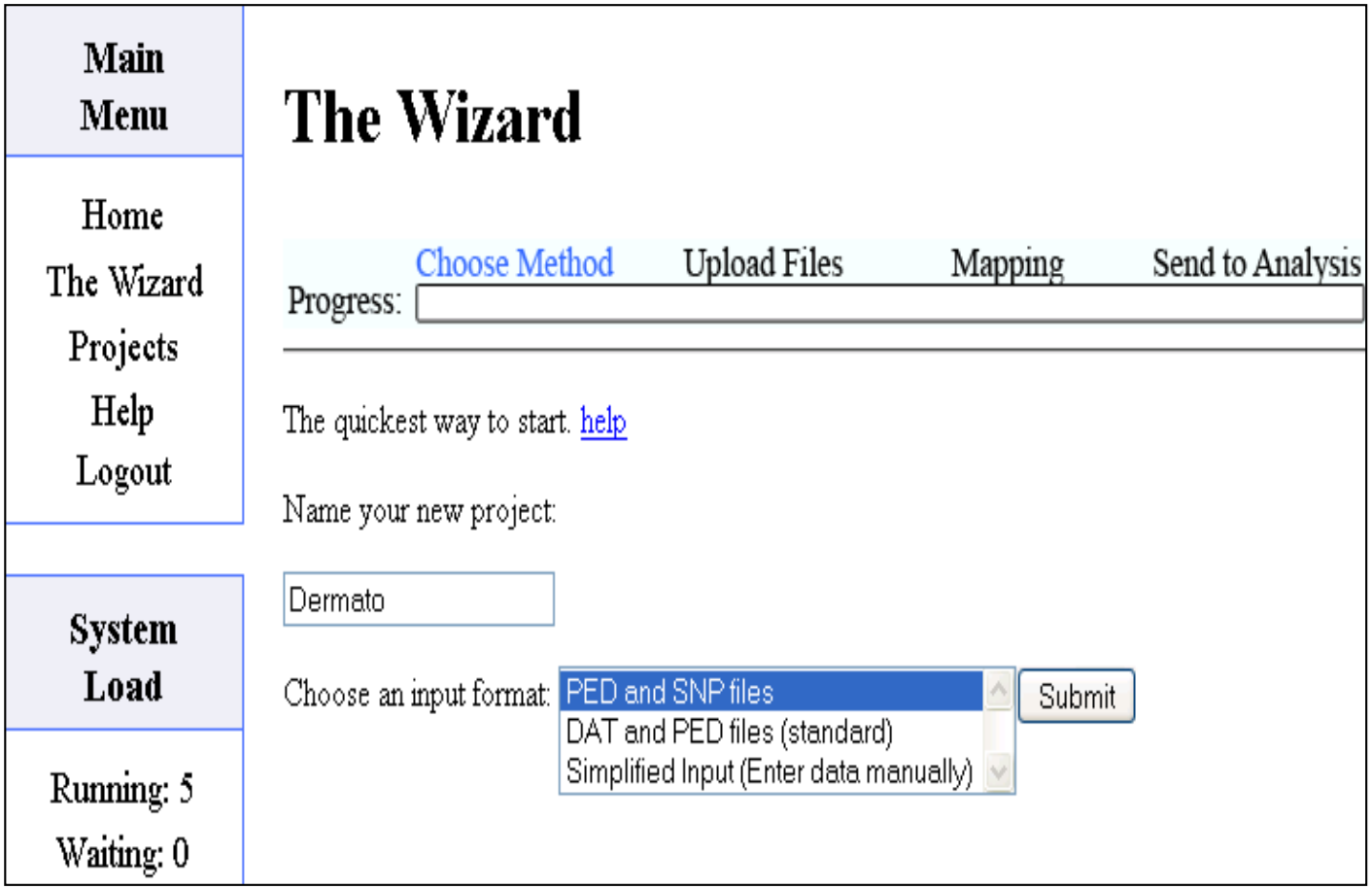
Column 2: id of the father

Column 3: id of the mother

Column 4: sex (1=Male, 2=Female)

Column 5: status (0=Unknown, 1=unaffected, 2=affected)

Superlink SNP Online: wizard



Choose input format

Superlink SNP Online: wizard

Upload data

SNP and PED files format chosen.

Submit data for project Dermato:

Snp

Upload your snp file(s): Bladeren... Geen bestand geselecteerd. View example | Download example

Pedigree

○ Upload your ped file: Bladeren... Geen bestand geselecteerd. View example | Download example | ped file format

◉ Pedigree in simplified input format:

id father mother sex status

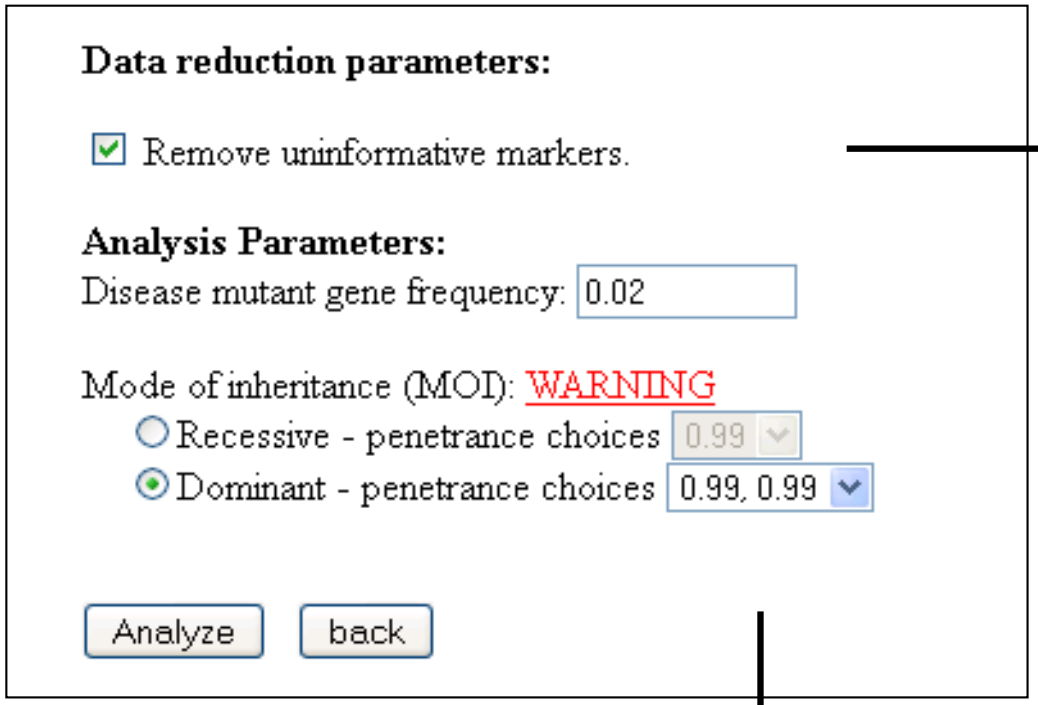
For multiple pedigrees, each pedigree starts with a separate line of the form: pedigree: id

Upload Data

Available actions: Continue | Save and Close | Delete selected

Superlink SNP Online: wizard

Set parameters



e.g. homozygous for whole family

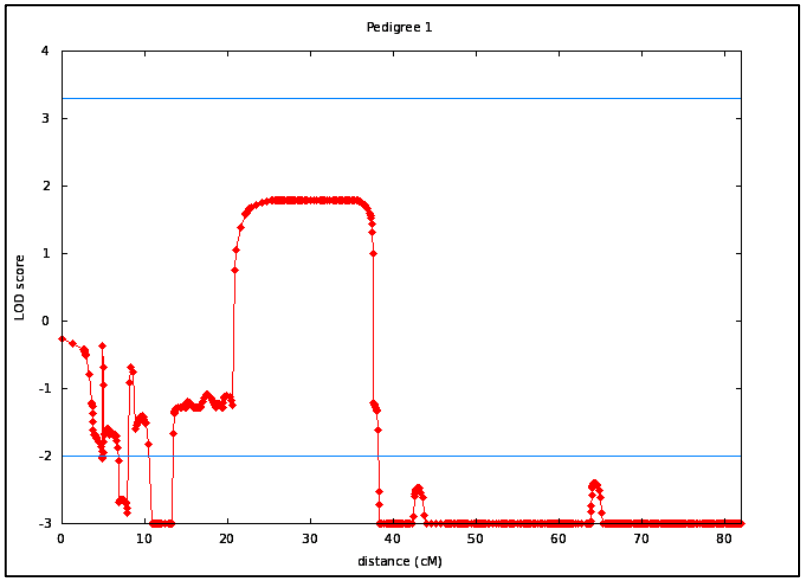
If not fully penetrant power will drop!

Superlink SNP Online: HMM_MultiPoint

Powerful, but computationally expensive

➔ **only possible for small pedigrees**

Marker Name	Marker Position	Affected												Un-Affected				Unknown			
		Typed										Un-Typed				Un-Typed				Un-Typed			
		7 (12)		8 (6)		10 (11)		12 (8)		15 (9)		17 (10)		1 (5)		5 (7)		2 (2)		6 (4)		18 (3)		19 (1)	
rs6599770	0.0000	3	5	3	1	7	2	7	6	8	2	7	2	5	1	6	2	3	4	7	8	5	6	1	2
rs4932679	2.6096	3	5	3	1	7	2	7	6	8	2	7	2	5	1	6	2	3	4	7	8	5	6	1	2
rs748979	2.6931	3	5	3	1	7	2	7	6	8	2	7	2	5	1	6	2	3	4	7	8	5	6	1	2
rs35248560	2.7095	3	5	3	1	7	2	7	6	8	2	7	2	5	1	6	2	3	4	7	8	5	6	1	2
rs7403800	2.7696	3	5	3	1	7	2	7	6	8	2	7	2	5	1	6	2	3	4	7	8	5	6	1	2



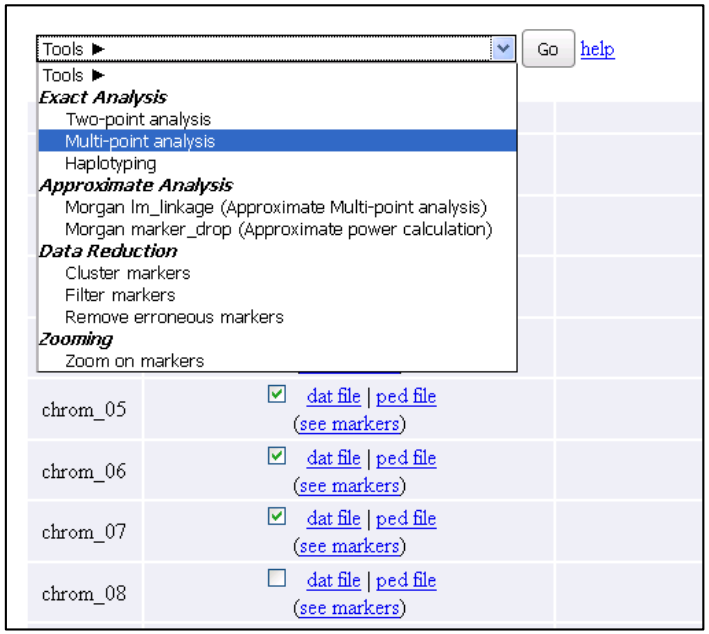
Max LOD = 2 because pedigree is small (6 typed individuals)

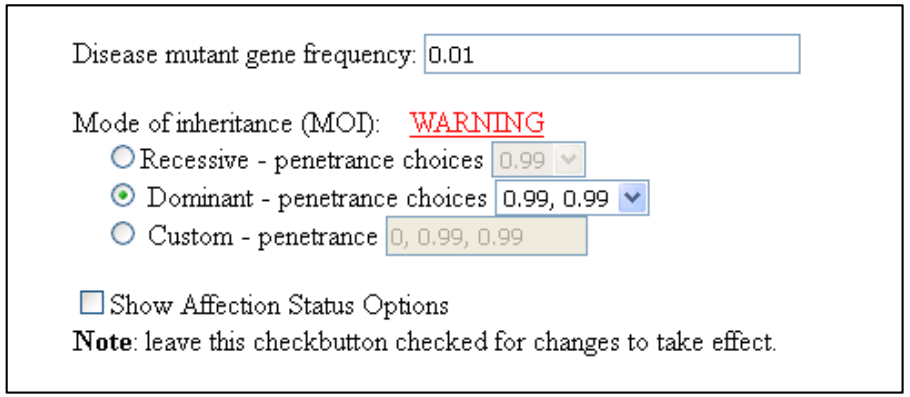
➔ **≈ 10^2 times more likely that the marker is linked with the disease causing mutation than it is not linked with the mutation**

Center for Medical Genetics

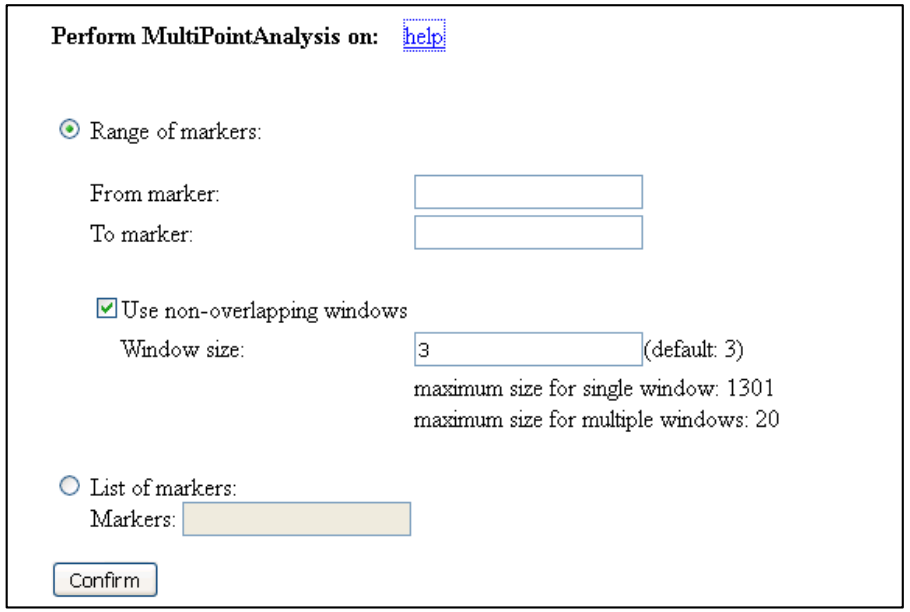
Superlink SNP Online: Exact multi-point analysis

Possible for all pedigrees



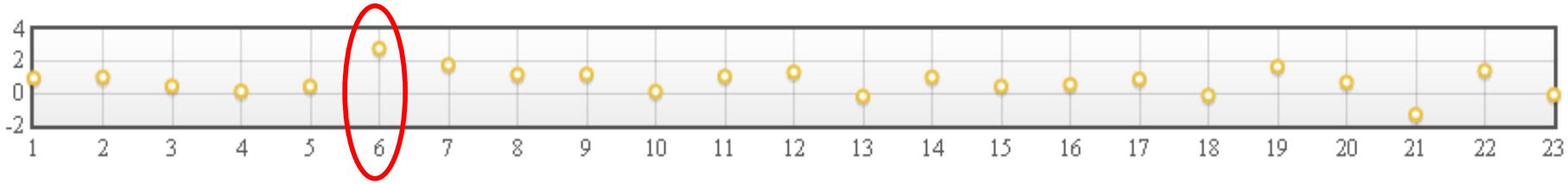


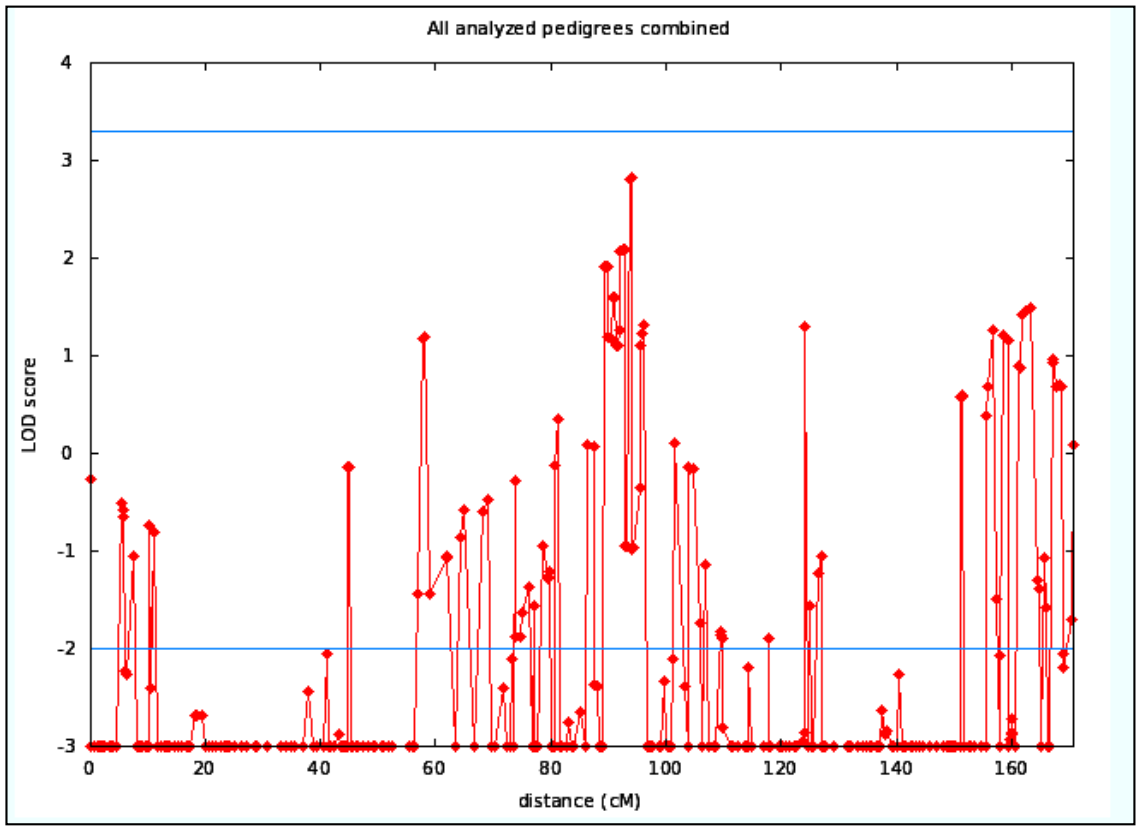
Set parameters



Center for Medical Genetics

Superlink SNP Online: Exact multi-point analysis





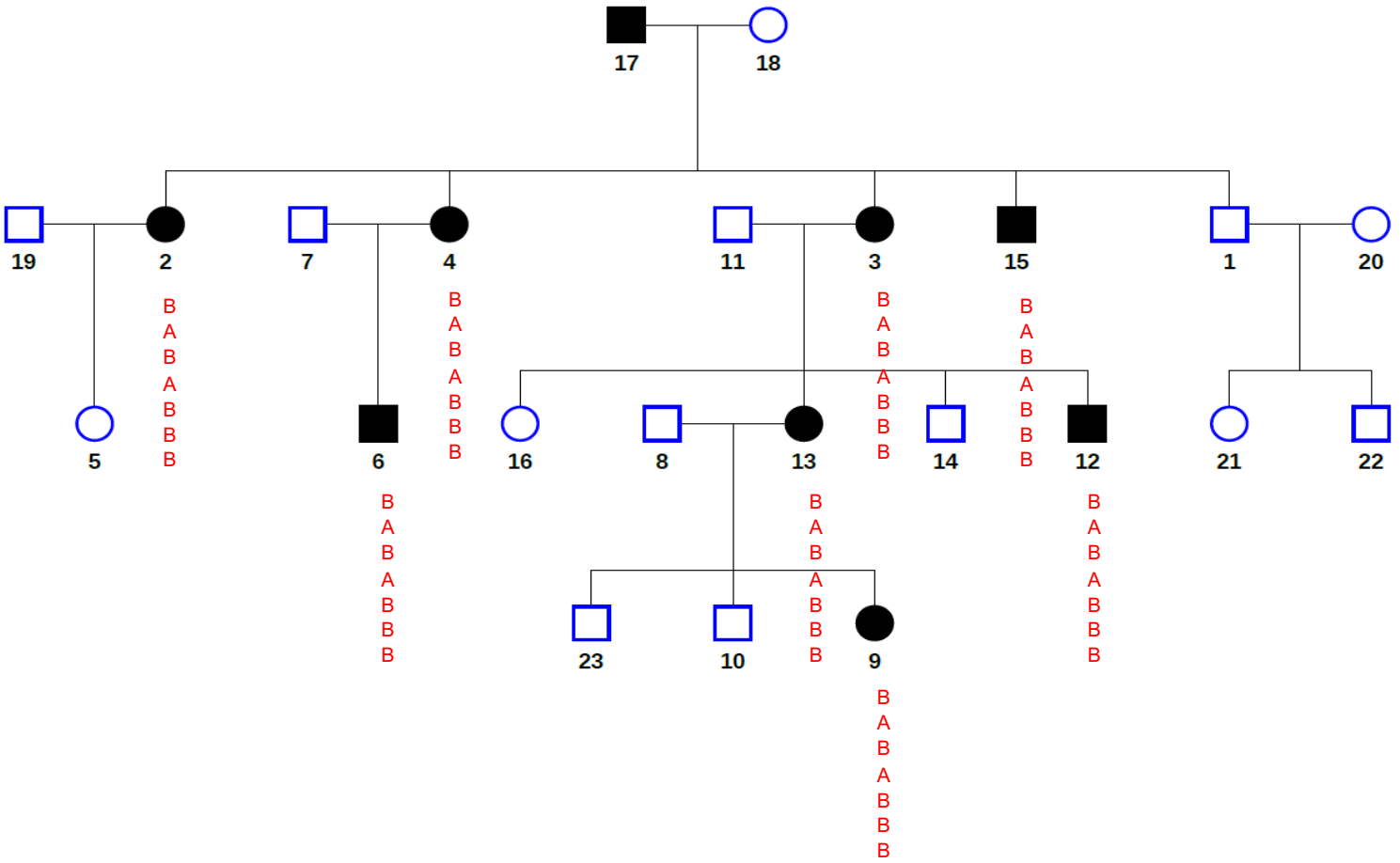
rs632385	-4.2188
rs1884831	-8.4861
rs2246529	1.9108
rs1040676	1.9170
rs205224	1.9169
rs1980986	1.1959
rs7773548	1.1939
rs1179900	1.1810
rs1055403	1.5942
rs285619	1.6051
rs790605	1.6046
rs1145739	1.1064
rs1145773	1.1140
rs1884260	1.2548
rs2144363	2.0748
rs1979797	2.0962
rs1498252	2.0928
rs1373376	-0.9436
rs1863659	-0.9406
rs1427123	-0.9419
rs1040155	2.8121
rs1338248	2.8130
rs1324103	2.8199
rs1548297	-0.9737
rs491112	-0.9578
rs2493964	-0.3563
rs169125	1.0982
rs2890370	1.2314
rs4839952	1.3081
rs7772067	-3.0293
rs2472922	-6.0790

Fetch genes and discover
(e.g. via BioMart)

Max LOD = 2,8

Center for Medical Genetics

Superlink SNP Online: Exact multi-point analysis



All affected individuals should have 1 haplotype in common.
The causal haplotype? Yes, with a certain probability.
BUT…

Center for Medical Genetics

Linkage analysis: difficulties

- Uncertainty about the clinical status of some individuals
  → exclude those individuals (but power will drop)
- Uncertain relationships (software can correct for this)
- Incomplete penetrance
- Phenocopies
- Statistically significant linkage (LOD > 3,2)
  ≠ usefulness of the analysis

↓

Prioritize regions
Exclude regions (f.e. known disease causing genes)
Often followed by exome analysis

↓

Merlin

http://csg.sph.umich.edu/abecasis/Merlin/download/

+ Available for Linux/Win/Mac
+ Powerful
+/- Command line

letter

Merlin—rapid analysis of dense genetic maps using sparse gene flow trees

Gonçalo R. Abecasis[1,2], Stacey S. Cherny[1], William O. Cookson[1] & Lon R. Cardon[1]

Published online: 3 December 2001, DOI: 10.1038/ng786