

LD Pairwise Analysis Matrix Output

Author: Greta Linse Peterson, Golden Helix, Inc.

Overview

LD Analysis is performed on all pairs within a chromosome (if a marker map is applied) or within a haplotype block. This function creates four spreadsheets in matrix form (markers as the row labels and column headers). The spreadsheets contain values for both the EM and CHM methods and both R^2 and D' values.

This script can be run on a spreadsheet that includes both genotype and phenotype data, with or without a marker map. It can also be run on a subset spreadsheet created from an LD plot by selecting **Subset Markers**. You can also run this script on markers in haplotype blocks as specified in a Haplotype Block Spreadsheet.

It is not recommended to run this script on a whole genome dataset as the computations are very time-intensive. Rather it would be best to subset or active only a few hundred markers that are of interest.

Recommended Directory Location

Save the script to the following directory:

***..\Application Data\Golden Helix SVS\UserScripts\Spreadsheet\Analysis**

Note: The **Application Data** folder is a hidden folder on Windows operating systems and its location varies between XP and Vista. The easiest way to locate this directory on your computer is to open SVS and select the **Tools >Open Folder > UserScripts Folder** menu option. If saved to the proper folder, this script will be accessible from the spreadsheet **Analysis** menu.

Using the Script

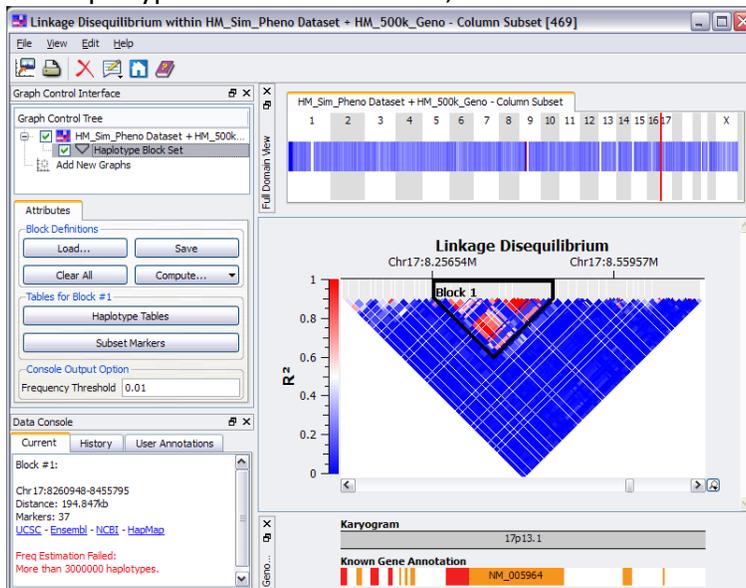
- 1) Method One:
 - a) Open a spreadsheet containing genotype data (possibly including phenotype data and may or may not be marker mapped).

Map	Sample ID	Height (in)	Age	SNP_A-1909444	SNP_A-2237149	SNP_A-4303947	SNP_A-1886933	SNP_A-2236359
Chromosome				1	1	1	1	1
Position				742429	767376	769185	775852	782343
Cytoband				p36.33	p36.33	p36.33	p36.33	p36.33
1	NA12813	60.27476	55.48754	B_B	B_B	A_A	B_B	B_B
2	NA12740	67.06439	54.33223	A_B	B_B	A_A	B_B	B_B
3	NA19193	65.06428	29.11164	A_A	B_B	B_B	A_A	B_B
4	NA18501	66.4639	46.90015	A_B	B_B	A_B	A_B	A_B
5	NA12874	55.6836	34.07425	B_B	B_B	A_A	B_B	B_B
6	NA19154	70.79579	31.26026	A_A	B_B	A_B	A_A	A_B
7	NA19129	71.88845	32.83759	A_A	B_B	A_B	A_A	A_B
8	NA12761	74.85801	50.45818	B_B	B_B	A_A	B_B	B_B
9	NA18537	68.62109	42.67434	A_B	B_B	A_B	A_B	B_B
10	NA19201	60.72689	52.96835	A_A	B_B	B_B	A_A	B_B
11	NA12750	63.3528	23.07136	B_B	B_B	A_A	B_B	B_B
12	NA18998	65.56057	23.33213	B_B	B_B	A_A	B_B	B_B
13	NA19132	71.94062	29.21608	A_A	B_B	B_B	A_A	B_B
14	NA05985	67.98739	27.66916	B_B	T_T	A_A	B_B	B_B
15	NA19132	66.04884	46.07400	A_A	R_R	R_R	A_A	R_R

- b) Select **Analysis > LD Pairwise Analysis Matrix Output**.
- c) Select **No** when asked if using a haplotype block spreadsheet.

2) Method Two:

- a) Open a Plot viewer containing an LD plot. Select a block of markers, and under the Haplotype Block Set attributes, select Subset Markers.



- b) This creates a subset spreadsheet with these markers.

Map	Sample ID	G 1	G 2	G 3	G 4	G 5	G 6	G
Chromosome		17	17	17	17	17	17	17
Position		8260948	8264129	8270690	8273323	8285267	8291756	8316
Cytoband		p13.1	p13.1	p13.1	p13.1	p13.1	p13.1	p13
1	NA12813	B_B	B_B	A_A	A_B	B_B	B_B	
2	NA12740	?_?	B_B	A_A	A_B	B_B	B_B	
3	NA19193	B_B	B_B	B_B	A_A	B_B	B_B	
4	NA18501	A_B	A_B	B_B	A_A	B_B	B_B	
5	NA12874	B_B	B_B	A_A	A_B	B_B	B_B	
6	NA19154	B_B	A_B	B_B	A_A	B_B	B_B	
7	NA19129	B_B	A_B	B_B	A_A	A_B	A_B	
8	NA12761	A_B	A_B	B_B	A_B	B_B	B_B	
9	NA18537	B_B	B_B	A_A	A_B	B_B	B_B	
10	NA19201	A_B	A_B	A_B	A_B	B_B	B_B	
11	NA12760	B_B	B_B	B_B	A_A	B_B	B_B	

- c) Select **Analysis > LD Pairwise Analysis Matrix Output**.
 - d) Select **No** when asked if using a haplotype block spreadsheet.
- 3) Method Three:
- a) Open a spreadsheet containing genotype data (possibly including phenotype data and may or may not be marker mapped).
 - b) Select **Analysis > LD Pairwise Analysis Matrix Output**.
 - c) Select **Yes** when asked if using a haplotype block spreadsheet.
 - d) Select the haplotype block spreadsheet.

Map	Markers	Block #
1	SNP_A-1865924	1
2	SNP_A-2041478	1
3	SNP_A-4274931	2
4	SNP_A-2285039	2
5	SNP_A-1941521	3
6	SNP_A-1860613	3
7	SNP_A-1817285	4
8	SNP_A-1791886	4
9	SNP_A-1938509	5
10	SNP_A-1904645	5
11	SNP_A-1802724	5
12	SNP_A-4174969	5
13	SNP_A-2005048	5
14	SNP_A-2064632	6
15	SNP_A-2241518	6
16	SNP_A-2052894	7
17	SNP_A-1820470	7
18	SNP_A-4281588	7
19	SNP_A-1800897	8
20	SNP_A-2109637	8
21	SNP_A-2245307	9
22	SNP_A-2178502	9
23	SNP_A-1944345	10
24	SNP_A-4196448	10

Haplotype blocks, 27 markers in 10 groups

Four new spreadsheets result with all pair-wise comparisons for both R^2 and D' calculations for both the EM method and the CHM method.

Row	Label	Marker 1	Marker 2	Distance in markers	Distance in kb	EM - R Squared	EM - D Prime	CHM - R Squared	CHM - D Prime
19	SNP_A-2057889	SNP_A-2241518		19	107.925	0.032816660469349	0.313722169258354	0.0393442368417157	0.343500363885569
20	SNP_A-2057889	SNP_A-4249414		20	107.957	0.0110250783864464	0.935280100491622	0.00728897878881073	0.758906993676153
21	SNP_A-2057889	SNP_A-2067666		21	129.544	0.0142806056286833	0.978369860013282	0.0115006143914997	0.877999624332894
22	SNP_A-2057889	SNP_A-1813732		22	129.526	0.0141242111802054	0.97740407860535	0.0112769605075401	0.873355154626512
23	SNP_A-2057889	SNP_A-2026246		23	134.418	0.0141242111802054	0.97740407860535	0.0112769605075401	0.873355154626512
24	SNP_A-2057889	SNP_A-4258149		24	143.653	0.0138524679112468	0.9730744657763629	0.0107786986312603	0.858353470042257
25	SNP_A-2057889	SNP_A-4254680		25	145.107	0.0140978882995759	0.976656735904389	0.0111526572988112	0.868665464200792
26	SNP_A-2057889	SNP_A-1836491		26	147.266	0.0136379366703397	0.98236258051233	0.0137460144001237	0.986247405717891
27	SNP_A-2057889	SNP_A-4223873		27	154.134	0.0141242111802054	0.97740407860535	0.0112769605075401	0.873355154626512
28	SNP_A-2057889	SNP_A-2149200		28	154.503	0.0141242111802054	0.97740407860535	0.0112769605075401	0.873355154626512
29	SNP_A-2057889	SNP_A-1859515		29	156.52	0.0514980734105764	0.47456305552540	0.0597440397369558	0.51153967923242
30	SNP_A-2057889	SNP_A-1836510		30	163.752	0.01492010320102058	0.97941085915853	0.013377805828589	0.927401427646584
31	SNP_A-2057889	SNP_A-2056328		31	164.351	0.0142554543292355	0.977647314330333	0.0113750979578992	0.873311630712922
32	SNP_A-2057889	SNP_A-2052894		32	164.46	0.0489264524934604	0.590906549065138	0.036320336859323	0.50918491377441
33	SNP_A-2057889	SNP_A-1820470		33	168.575	0.0514579067102755	0.614482602750051	0.0351954899290527	0.508190939910701
34	SNP_A-2057889	SNP_A-4281588		34	176.835	0.0519240867565488	0.617530525084385	0.0352082847855593	0.508506408799012
35	SNP_A-2057889	SNP_A-2005049		35	183.595	0.009950306630302	0.9518673892162	0.00759370240452801	0.831543997116625
36	SNP_A-2057889	SNP_A-1800897		36	184.847	0.0001430549173746	0.8518512829430728	0.0015418703694644	0.021510012403124
37	SNP_A-2032943	SNP_A-1804870		1	6.561	0.17952787842027	0.999473967316139	0.17352321794388	0.98264314060071
38	SNP_A-2032943	SNP_A-2159364		2	9.194	0.0089570853880188	0.220727609566008	0.010292389601539	0.248856640702853
39	SNP_A-2032943	SNP_A-1788959		3	21.138	0.063562383378042	0.9990294242232	0.077466007744058	1.10488976167192
40	SNP_A-2032943	SNP_A-2260393		4	27.627	0.0801888187306131	0.999619962339513	0.0988207228939977	1.10989853808265
41	SNP_A-2032943	SNP_A-2272371		5	51.913	0.0638888189404851	0.99901904940492	0.0779136056481325	1.10323662446392
42	SNP_A-2032943	SNP_A-1810767		6	57.026	0.0017149525109944	0.22025440759884	0.0013536742766707	0.195684104818262
43	SNP_A-2032943	SNP_A-4252161		7	57.212	0.14445801156575	0.99678887545118	0.039462598221795	0.94260147911788
44	SNP_A-2032943	SNP_A-1938059		8	61.471	0.109448204284284	0.99930623973817	0.15745853799745	1.1989487120924
45	SNP_A-2032943	SNP_A-2090882		9	66.16	0.0453576400501332	0.99748787888474	0.042843996584366	0.957800037635631
46	SNP_A-2032943	SNP_A-1904645		10	68.749	0.0698833943623289	0.66066242172157	0.093803066685992	0.765422089792705
47	SNP_A-2032943	SNP_A-1802724		11	71.72	0.0930110890476407	0.44702234835432	0.10962660252642	0.485310679685882