

LD Pairwise Analysis

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. The resulting spreadsheet 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

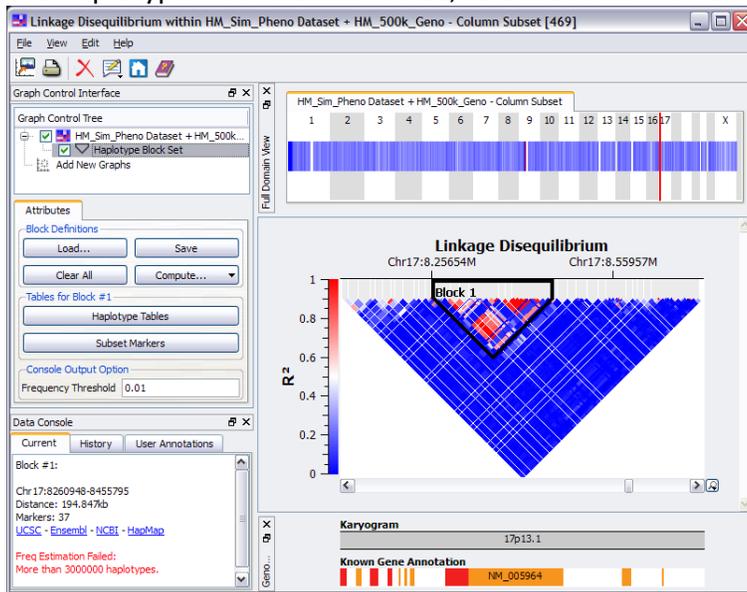
- 1) Method One:
 - a) Open a spreadsheet containing genotype data (possibly including phenotype data and may or may not be marker mapped).

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	NA12760	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	NA06985	67.98739	27.66916	B_B	T_T	A_A	B_B	B_B
15	NA19049	65.19484	46.17499	A_A	R_R	R_R	A_A	R_R

- b) Activate at most 75 markers from each chromosome, including unmapped markers. Phenotype data does not have to be inactivated.
- c) Select **Analysis > LD Pairwise Analysis**.
- d) 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	SNP_A-2057889	SNP_A-2032943	SNP_A-2048970	SNP_A-2159364	SNP_A-1788959	SNP_A-2260393	SNP_A-2...
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	B_B
2	NA12740	?_?	B_B	A_A	A_B	B_B	B_B	B_B
3	NA19193	B_B	B_B	B_B	A_A	B_B	B_B	B_B
4	NA18501	A_B	A_B	B_B	A_A	B_B	B_B	B_B
5	NA12874	B_B	B_B	A_A	A_B	B_B	B_B	B_B
6	NA19154	B_B	A_B	B_B	A_A	B_B	B_B	B_B
7	NA19129	B_B	A_B	B_B	A_A	A_B	A_B	A_B
8	NA12761	A_B	A_B	B_B	A_B	B_B	B_B	B_B
9	NA18537	B_B	B_B	A_A	A_B	B_B	B_B	B_B
10	NA19201	A_B	A_B	A_B	A_B	B_B	B_B	B_B
11	NA12760	B_B	B_B	B_B	A_A	B_B	B_B	B_B

- c) Select **Analysis > LD Pairwise Analysis**.
- 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**.
- 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

The result is a new spreadsheet with the first two columns listing the markers that were compared (all pair-wise comparisons, only listing one occurrence of the pair), the distance in markers, the distance in kb (if the genotype spreadsheet was marker mapped), both R^2 and D' values for the EM method and the CHM method.

Marker 1	Marker 2	Distance in markers	Distance in kb	EM - R Squared	EM - D Prime	CHM - R Squared	CHM - D Prime
SNP_A-2057889	SNP_A-2241518	19	107.925	0.0128166604869349	0.313722169258354	0.0393442360417157	0.343509363885569
SNP_A-2057889	SNP_A-4240414	20	107.957	0.011025078386464	0.935280100491622	0.00725897678881073	0.758906993676153
SNP_A-2057889	SNP_A-2067066	21	129.544	0.014280656286833	0.978368988003282	0.0115006143924997	0.8779906922435284
SNP_A-2057889	SNP_A-1913732	22	129.926	0.0141242111802054	0.977410407860535	0.0112769605075401	0.873355154626512
SNP_A-2057889	SNP_A-2020546	23	134.418	0.0141242111802054	0.977410407860535	0.0112769605075401	0.873355154626512
SNP_A-2057889	SNP_A-4159149	24	143.653	0.0138154679112468	0.97027461786269	0.0107769605075401	0.868352470042257
SNP_A-2057889	SNP_A-4254680	25	145.107	0.014097882995759	0.97665675904389	0.011152657298112	0.868654642017952
SNP_A-2057889	SNP_A-1836491	26	147.266	0.0136379366703397	0.98236258051233	0.0137760144001237	0.986247805717891
SNP_A-2057889	SNP_A-4223873	27	154.134	0.0141242111802054	0.977410407860535	0.0112769605075401	0.873355154626512
SNP_A-2057889	SNP_A-2149290	28	154.303	0.0141242111802054	0.977410407860535	0.0112769605075401	0.873355154626512
SNP_A-2057889	SNP_A-1858915	29	156.52	0.051496873109784	0.474563855562549	0.059744093796958	0.511153967523242
SNP_A-2057889	SNP_A-1836510	30	163.752	0.0149201052010058	0.97941059158833	0.0133775805828589	0.927401427646584
SNP_A-2057889	SNP_A-2056128	31	164.351	0.014255450292355	0.97764731330333	0.011375097078892	0.87311830729222
SNP_A-2057889	SNP_A-2052894	32	164.46	0.0480264524934604	0.50005639065138	0.036329336859323	0.5091891377941
SNP_A-2057889	SNP_A-1820470	33	168.575	0.0514579067102755	0.614482602750051	0.035195469290527	0.508190939910701
SNP_A-2057889	SNP_A-4281588	34	176.835	0.0519240867565488	0.617530525084385	0.035208284785593	0.508506408799012
SNP_A-2057889	SNP_A-2005049	35	183.595	0.009950305663003	0.9518677892162	0.00759370240452801	0.8315397116625
SNP_A-2057889	SNP_A-1800897	36	194.847	0.000114316447173746	0.0185212839436728	0.00015418703696464	0.0215100124530124
SNP_A-2032943	SNP_A-2048970	1	6.361	0.179527678745207	0.999473967316339	0.173022027974388	0.982614314000071
SNP_A-2032943	SNP_A-2157084	2	9.104	0.00809789932080180	0.2207276036008	0.0102923396601559	0.24855664702853
SNP_A-2032943	SNP_A-1788809	3	21.138	0.063562381378042	0.99902954242232	0.0777466007746058	1.10488976387392
SNP_A-2032943	SNP_A-2260793	4	27.627	0.0801588187306131	0.999619962339513	0.098820728935977	1.10988953808265
SNP_A-2032943	SNP_A-2272371	5	51.913	0.0638886189404851	0.999019049404492	0.0779136056481325	1.1032366246392
SNP_A-2032943	SNP_A-1910767	6	57.026	0.00171495225109944	0.220254410798844	0.00135367342466707	0.195684104818262
SNP_A-2032943	SNP_A-4252261	7	57.212	0.0444358021565175	0.996768887245318	0.0397452339822175	0.94269347561708
SNP_A-2032943	SNP_A-1938509	8	61.471	0.10948202842384	0.99930623973837	0.157454833799745	1.19839457120924
SNP_A-2032943	SNP_A-2208082	9	66.16	0.0435578480511332	0.9971078888474	0.04181596589496	0.95766037635631
SNP_A-2032943	SNP_A-1904645	10	68.740	0.068883394252289	0.66062421712157	0.09380306668592	0.76542289762705
SNP_A-2032943	SNP_A-1802724	11	71.72	0.0930110890476407	0.447022348354232	0.10962602525642	0.485310679685882