

PGA Seattle SNPs Interactive Tutorial II – Linkage Disequilibrium and TagSNPs
March 20-21, 2006
Dana Crawford, PhD

Goal: This tutorial introduces several websites and tools useful for determining linkage disequilibrium for your gene or region of interest and tagSNP selection. In this section, you will cover the following topics.

- SeattleSNPs website tools
 - Visual Genotype (VG2)
 - Visual Haplotype (VH1)
- TagSNP selection tools
 - LDSelect
 - Haploview

Part 1. Visual Genotype (VG2) using Genome Variation Server

1. Go to the SeattleSNPs home page at <http://pga.gs.washington.edu> (or directly to the Genome Variation Server at <http://gvs.gs.washington.edu/GVS/>). Check that your browser is set to NOT block pop-up windows. GVS uses pop-up windows to display the images and tables.
2. Click the Gene Name search button
3. Enter 'USF1' for gene name, '2500' in upstream field, and '1500' in downstream field.
4. Click on "Search."
5. A table displaying number of variations, number of genotypes, populations, and data submitters will appear. By default, GVS has selected the European-descent (CEPH) population re-sequenced by SeattleSNPs (Submitter is PGA-UW-FHCRC). For this gene, we know that Perlegen produced data for the sample samples re-sequenced by SeattleSNPs. Check the box for "AFD EUR PANEL" with the Perlegen submitter.
6. Scroll down the page, and click on "display genotypes." Two pop-up windows appear: one displaying a visual representation of all the SNPs available for your search criteria and one displaying a table containing the genome position of the SNPs available for your search criteria. For the visual genotype table, the numbers at the top of the image represent the SNPs (numbered along a reference sequence used in re-sequencing the gene). The numbers on the left side of the image represent the sample ID. Each square represents an individual sample's genotype: homozygous for the common allele (blue), heterozygous (red), and homozygous for the rare allele (yellow). The SNPs are color coded to describe their location within the gene.
7. Close the pop-up windows. To find the rs numbers for these SNPs, go to the pull-down menu for "Output SNPs by" and select "RS_ID."
8. Click on "display genotypes."

9. To save the image to your computer, right-click on the image and choose “save as.”

GVS: Genome Variation Server BETA
Sponsored by [SeattleSNPs](#)

An NHLBI Program for Genomic Applications

Gene Name: **USF1** [expand search](#)

Gene ID: 7391

Chromosome 1: 157822114 - 157828830 (-)

Total chromosome span: 157819614 - 157830330

Genes in this region: F11R USF1 ARHGAP30

Select Population(s)

Check at least one population. You may select up to 4 populations if all those selected have individuals in common.

Select	Number Variations	Number Genotypes	Population	Submitter	Check to Select Only Unrelated Individuals
<input checked="" type="checkbox"/>	46	1058	PGA-EUROPEAN-PANEL	PGA-UW-FHCRC	
<input type="checkbox"/>	46	1104	PGA-AFRICAN-PANEL	PGA-UW-FHCRC	
<input type="checkbox"/>	13	1170	HapMap-YRI	CSHL-HAPMAP	<input type="checkbox"/>
<input type="checkbox"/>	13	1170	HapMap-CEU	CSHL-HAPMAP	<input type="checkbox"/>
<input type="checkbox"/>	12	540	HapMap-JPT	CSHL-HAPMAP	
<input type="checkbox"/>	12	540	HapMap-HCB	CSHL-HAPMAP	
<input type="checkbox"/>	10	240	AFD_CHN_PANEL	PERLEGEN	
<input type="checkbox"/>	10	230	AFD_AFR_PANEL	PERLEGEN	
<input checked="" type="checkbox"/>	10	240	AFD_EUR_PANEL	PERLEGEN	

Set up parameters for display and analysis

Data Output and Display

Output SNPs By	RS_ID	Display SNPs By	Table/Image
Allele Frequency Cutoff (%)	0	No Monomorphic Sites	<input type="checkbox"/>
Clustering in Graphic Display			
Cluster SNPs	<input type="checkbox"/>	Cluster Samples	<input type="checkbox"/>
Selecting Tag SNPs			
r^2 Threshold (0.0-1.0)	0.80		
Color-Coding For LD Plot			
LD Minimum (0.0-1.0)	0.1	LD Maximum (0.0-1.0)	1.0

Display Results

display genotypes →
display tag snps →
display linkage disequilibrium →
display snp summary →

reset ↻

National Heart, Lung, and Blood Institute

Programs for Genomic Applications

Part 2. Linkage Disequilibrium (LD) using Genome Variation Server

1. With steps 1 through 5 completed from above.
2. Scroll down. Under “Allele frequency cutoff (%)”, type “5.” This will exclude SNPs with a minor allele frequency less than 5%.
3. Check the box for “Cluster SNPs.”
4. Now Choose the LD statistic (r^2) using the pull-down menu “Linkage Disequilibrium Plot.”
5. Click on “Display linkage disequilibrium”
6. You should have an image of the genotypes and an LD plot appearing in a pop-up window. You should also have a table displaying the pair-wise LD values (r^2) for each SNP pair.
7. To save the image to your computer, right-click on the image and choose “save as.”

<input type="checkbox"/>	12	540	HapMap-JPT	CSHL-HAPMAP
<input type="checkbox"/>	12	540	HapMap-HCB	CSHL-HAPMAP
<input type="checkbox"/>	10	240	AFD_CHN_PANEL	PERLEGEN
<input type="checkbox"/>	10	230	AFD_AFR_PANEL	PERLEGEN
<input checked="" type="checkbox"/>	10	240	AFD_EUR_PANEL	PERLEGEN

Set up parameters for display and analysis

Data Output and Display	
Output SNPs By	<input type="text" value="SNP_Position"/>
Display SNPs By	<input type="text" value="Table/Image"/>
Filtering SNPs	
Allele Frequency Cutoff (%)	<input type="text" value="5"/>
No Monomorphic Sites	<input type="checkbox"/>
Clustering in Graphic Display	
Cluster SNPs	<input checked="" type="checkbox"/>
Cluster Samples	<input type="checkbox"/>
Selecting Tag SNPs	
r^2 Threshold (0.0-1.0)	<input type="text" value="0.80"/>
Color-Coding For LD Plot	
LD Minimum (0.0-1.0)	<input type="text" value="0.1"/>
LD Maximum (0.0-1.0)	<input type="text" value="1.0"/>

Display Results

→
 →
 →
 →


National Heart, Lung, and Blood Institute

Programs for Genomic Applications

Part 3. TagSNP Selection (LDSelect) using Genome Variation Server

1. With steps 1 through 3 completed from above.
2. Click on “display tagSNPs.” Two pop-up windows should appear: the visual genotype display and a table of the tagSNPs. TagSNPs are chosen using the algorithm LDSelect at the default $r^2 > 0.80$. The visual genotype output clusters SNPs together that are in the same “bin” (denoted by a line), and, within each bin, SNPs that can be considered tagSNPs are marked as “*.” Only *one* tagSNP per

Part 4. Using Visual Haplotype (VH1) for Haplotype tagSNP Selection in SeattleSNPs

1. Go to <http://pga.gs.washington.edu>
2. Find VH1 under “Software” (left-hand grey-colored panel of website). This software has a interface similar to VG2 but will display haplotypes. Haplotypes represent the alleles of each SNP assigned to an individuals chromosomes. Each individual has two chromosomes representing the maternal and paternal chromosomes inherited from his or her parents. The visual haplotype will be twice as long as the visual genotype because now each individual is represented by two rows of data (haplotypes) instead of just one row of data (genotypes).
NOTE: Be aware that a proportion of the genes re-sequenced by SeattleSNPs are X-linked. In this situation, males have one X chromosome and females have two X chromosomes.
3. Choose a population (European-descent) using the pull-down menu for “PGA Finished Gene PopulationFilter.”
4. Choose the gene *CRP* re-sequenced by SeattleSNPs using the pull-down menu for “PGA Finished Gene Prettybase Input File.”
5. To pick tagSNPs to represent common genetic variation, we suggest you filter by minor allele frequency for common SNPs. Enter 10 in “Rare Allele Percentage (integer, 0 to 50).” Note that the list of SNPs at a specific MAF will depend on the population you choose in step #3.
6. Under haplotype sorting, choose “haplotype by frequency.”
7. To identify SNPs in haplotypes that are correlated (or contained in a “block”), sort by site. At “Cluster By:” choose “SITE.”
8. Click on “Run VH1 on the Web!”
9. You should have an image of the haplotypes in a pop-up window. The numbers at the top of the image represent the SNPs (numbered along a reference sequence used in re-sequencing the gene). The SNPs here are sorted according to site relatedness. The numbers on the side of the image represent the sample ID. Each square represents an individual sample’s allele: common (blue) and rare (yellow) allele. Each row represents the individual sample’s haplotype, and each individual will have two rows representing the two chromosomes. You can identify “blocks” manually using VH1.

Questions: How many haplotypes do you have? How many tagSNPs would you genotype?

SeattleSNPs
Variation Discovery Resource

Search Site

Home
Gene Resources
 Genes Sequenced for SNPs
 Genes in Progress
 Proposed Gene Targets
 Summary Statistics
 Summary Data
 Data Download
 Gene Nomination
Software
 Polyphred
 VG2
 VH1
 LDSelect
 PCR Overlap
 GeneHunter
Education
 2005 Workshop
 Previous Workshops
 PGA Symposium Download

**Displaying Estimated Haplotype Data:
Visual Haplotypes**

Local Phasebase Input File: Browse...

PGA Finished Gene Population Filter: Display Only European Descent

Haplotype Sorting: Haplotype by Frequency

PGA Finished Gene Phasebase Input File: CRP

Rare Allele Percentage (integer, 0 to 50): 10 Cluster By: SITE

Generate Representative Phasebase:

Run VH1 on the Web!

Part 5. Where to Find Haplotypes in SeattleSNPs

1. In addition to VH1, we offer PHASEv2.0 output for each of our genes re-sequenced on the SeattleSNPs website. On the home page, click on “Genes Sequenced for SNPs” (left side).
2. Choose *MCP*.
3. PHASE output is found in the “Haplotyping Data” section of the gene’s web page. We also offer a static image of the haplotypes in this section. To manipulate this image, use VH1 under software and create a new haplotype image for your gene of interest.

Part 6. Downloading Genotype Data from HapMap

1. Go to www.hapmap.org
2. Click on “Browse Project Data” on left side of website.
3. In “Landmark or Region” field, type “*ABCE1*.” Select the latest phase II data (from January 2006) from the pull-down menu.
4. Click on “Search.” You should see the gene *ABCE1* with 25 genotyped SNPs (denoted by little pie charts symbolizing the allele frequency for each population sample genotyped). If you do not see the pie charts, scroll down and check “Genotyped SNPs” under “Variation” and then click on the “Update Image” button. If you still don’t see the pie charts, click on the “Set Track Options” button and check track 11. Click the “Accept Changes and Return” button.
5. At the “Reports & Analysis” pull-down menu, choose “Download SNP genotype data.”
6. Click on “Configure.” Choose a population (CEU is CEPH or European-descent). The click on “Save to Disk.” Alternatively, you can click on “Open directly in HaploView” if you have Haploview loaded on your computer. Click “Go.”

Part 7. Using HapMap Data in Haploview

1. Download and install Haploview 3.2 from <http://www.broad.mit.edu/mpg/haploview/index.php>

2. Open Haploview. Click on “Load HapMap data.” Load the file you saved in the previous section (Downloading Genotype Data from HapMap). If you are connected to the internet, click on “Download and show HapMap info track?” “Click “OK.” Alternatively, if you did not save the file from the previous section, you can download the file “ABCE1_dumped_region” from http://pga.gs.washington.edu/workshop_2006/materials.html. Load this file onto Haploview and click “OK.”
3. The first view of the data is the “check markers” window. This provides a nice summary of the marker data, including name of the markers, genomic position of the markers, observed heterozygosity, predicted heterozygosity, Hardy Weinberg, % samples successfully genotyped, the number of fully genotyped family trios for each marker, the number of Mendelian inheritance errors, minor allele frequency, and pass/fail quality control for each marker. Seventeen markers failed (denoted in red) in the *ABCE1* dataset because they are monomorphic in the samples genotyped (no heterozygotes or homozygotes for the rare allele).
4. Haploview offers a visual of the LD statistic. Click on the “LD” tab. You can change haplotype block definitions by going to “Analysis” and select the block definition. The default is the block definition by Gabriel et al in *Science* (2002). To change the LD statistic, click on “Display” and select the statistic of your choice. For this example, choose “four-gamete rule”.

Questions: How many blocks are in *ABCE1* for the European-descent population using the default block definition in Haploview?

5. By default, if the LD statistic is 1.0 for a particular marker pair, the number 1.0 is not shown in the figure. Any LD statistic less than 1.0 will be shown in the figure. Right-click on the most upper left red square in the block. This pop-up will give you statistics related to the pair of markers used in calculating this LD statistic.

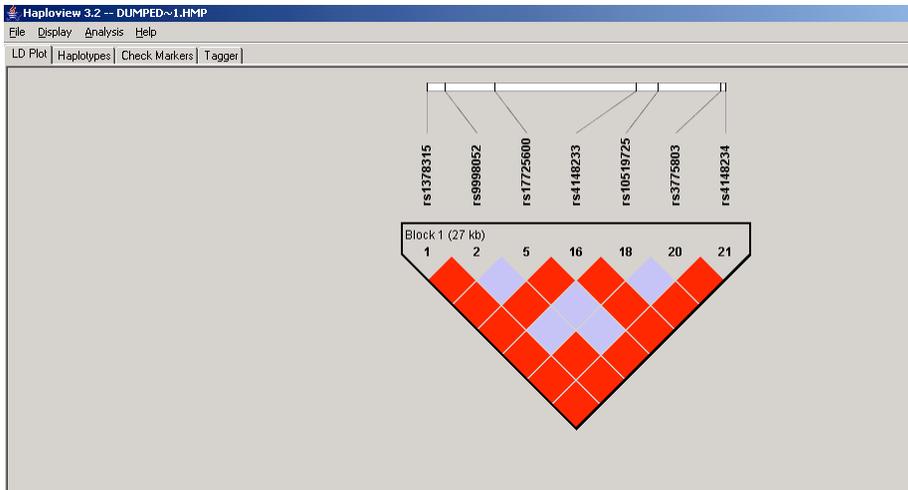
Questions: How far apart are these two markers physically?

6. For haplotypes, click on the “Haplotypes” tab. Haplotype frequencies are displayed on the right side of each haplotype. The triangles above the haplotypes denote the haplotype tagging SNPs. In cases of complex haplotypes (not shown here for the HapMap data for *ABCE1*), there will be lines connecting haplotypes, denoting their relationship to one another. Also, there will be a multiallelic D' value.

Questions: How many haplotypes were identified in this dataset? How many haplotype tagging SNPs were identified?

7. For the minimal set of tagSNPs, go to the “Tagger” tab. You can choose the algorithm used to define tagSNPs. For this example, choose “pairwise tagging only”. Then click “Run Tagger.” The results are displayed so that the tagSNPs are on the left of the screen. The right side of the screen shows which SNPs are being tagged by other SNPs.

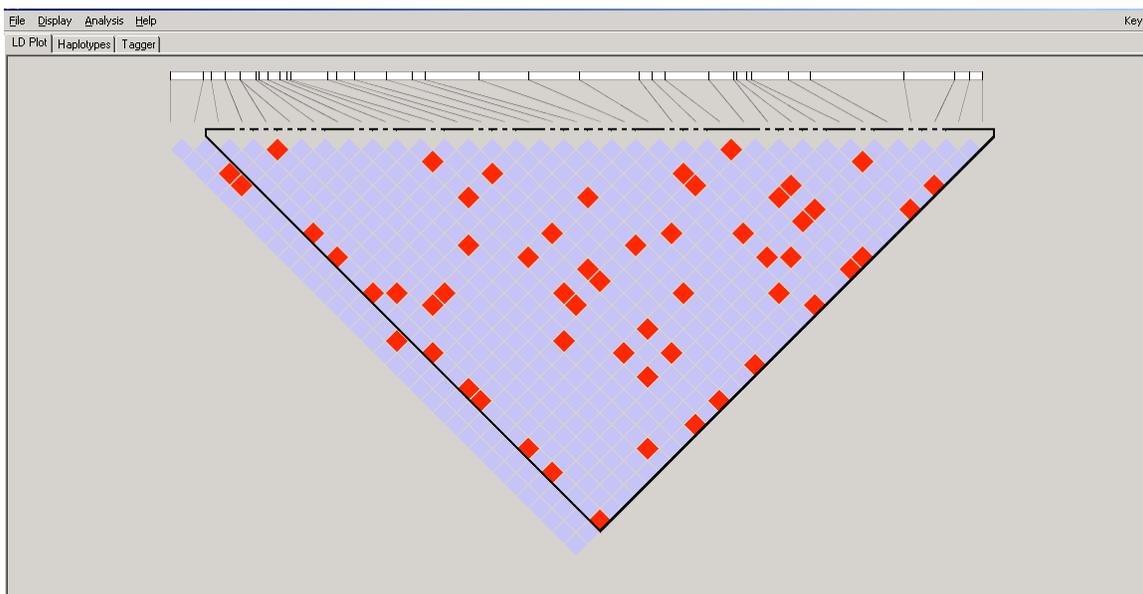
Questions: Using “Tagger” and “pair-wise tagging only,” how many Haploview tagSNPs are in *ABCE1* for the European-descent HapMap data?



Part 8. Using SeattleSNPs Data in Haploview

1. Open Haploview. If Haploview is already opened (for the previous exercise), under “File,” choose “Open genotype data.”
2. When loading the SeattleSNPs data for *ABCE1* for this exercise, click on “load phased haplotypes.”
3. Download “abce1.ED_gene.hapmap.inp” and “abce1_locus_info.txt” from http://pga.gs.washington.edu/workshop_2006/materials.html. The input file here is *ABCE1* haplotype data (using PHASEv2.1) for European-Americans using all SNPs (no minor allele frequency cut-off). Load this file onto Haploview and click “OK.” Repeat steps 4, 5, 6, and 7 from “Using HapMap Data in Haploview.” Note the difference between complete variation data and common (HapMap) variation data.

Questions: How many tagSNPs are identified using pair-wise tagging only in “Tagger” using SeattleSNPs data for *ABCE1*?



Answers to Questions:

How many bins are in *USF1* for the European-descent population? **7 bins: four bins with >1 SNP; three “bins” with only one SNP.**

How many tagSNPs must be genotyped directly because they are not contained within a bin with another SNP? **3**

Are there any exonic SNPs that are considered tagSNPs for *USF1*? **Yes. Two SNPs in the untranslated region.**

Which population requires more tagSNPs for *USF1*? **African-descent**

Which population sample requires more tagSNPs to represent *ADH1C*: African-descent or Asian-descent? **African-descent (23 tagSNPs). Asians-descent requires 7 tagSNPs.**

What happens when you decrease the r^2 threshold? **The number of tagSNPs decreases** (remember that you will have to increase your sample size to have the same power to detect an association with an r^2 threshold of 0.80).

How many haplotypes do you have? **4**

How many tagSNPs would you genotype? **3- 4**

How many blocks are in *ABCE1* for the European-descent population using the default block definition in Haploview? **One**

How far apart are these two markers physically? **1.6kb**

How many haplotypes were identified in this dataset? **5**

How many haplotype tagging SNPs were identified? **4**

Using “Tagger” and “pair-wise tagging only,” how many Haploview tagSNPs are in *ABCE1* for the European-descent HapMap data? **4**

How many tagSNPs are identified using pair-wise tagging only in “Tagger” using SeattleSNPs data for *ABCE1*? **19** (but note that there was no minor allele frequency cut-off, so this would tag all SNPs, both common and rare, in *ABCE1*).