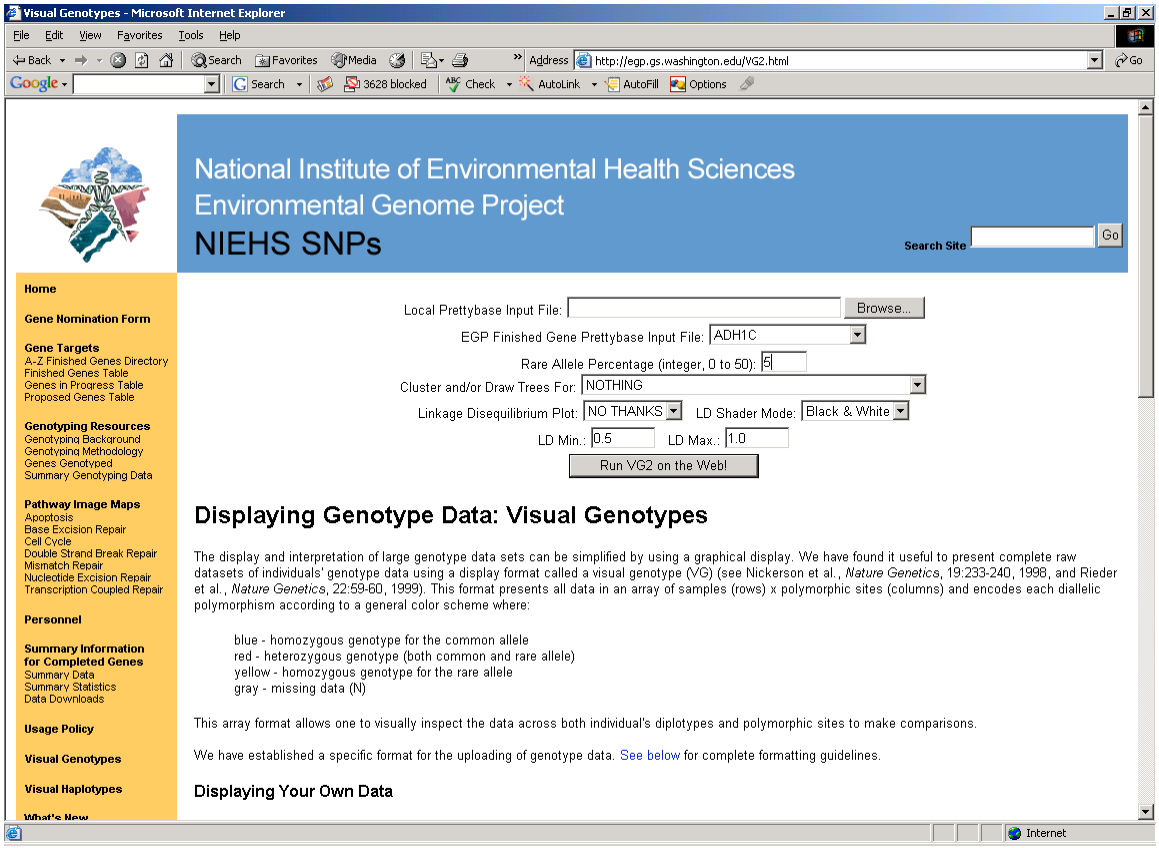
NIEHS SNPs Interactive Tutorial II – Linkage Disequilibrium and TagSNPs
January 31, 2006
Dana Crawford, PhD

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.

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

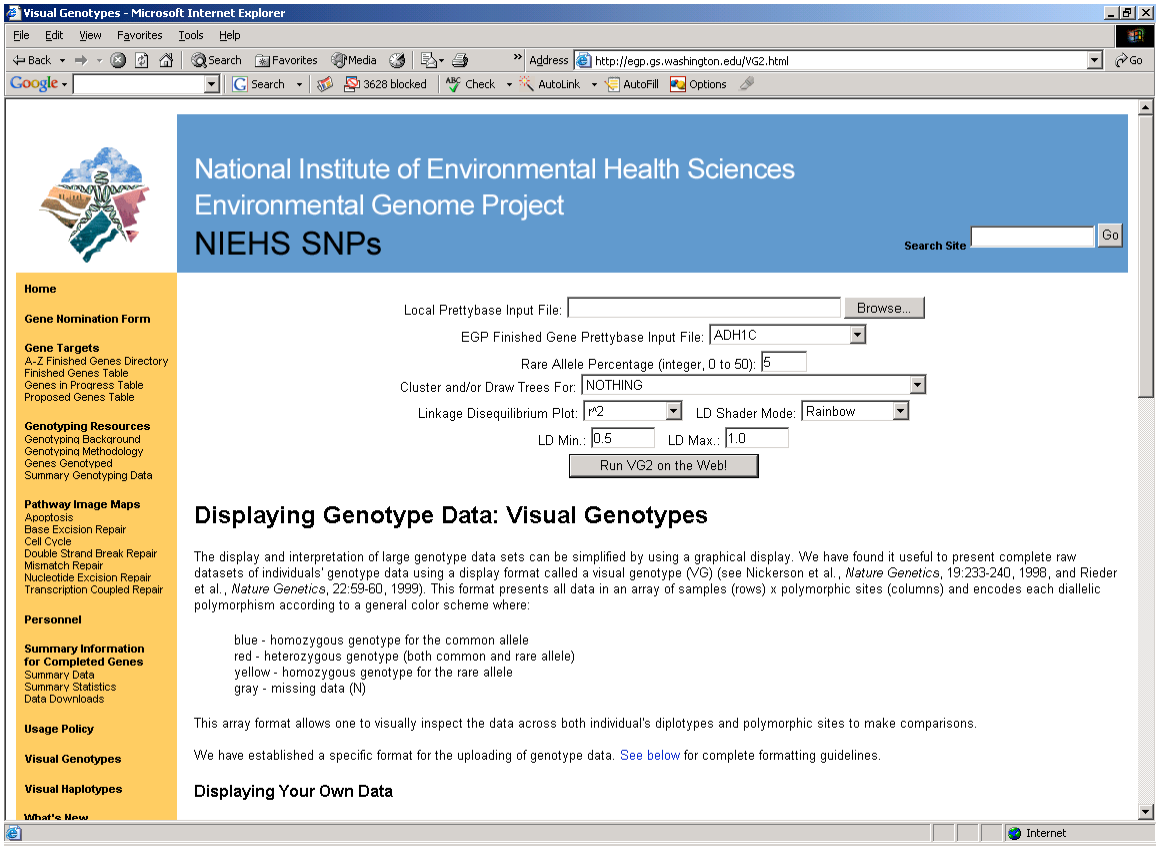
Part 1. Using Visual Genotype (VG2) in NIEHS SNPs

1. Go to http://egp.gs.washington.edu
2. Find VG2 by clicking on "Visual Genotypes" (within left-hand yellow-colored panel of website).
3. Choose the gene *ADH1C* re-sequenced by NIEHS SNPs using the pull-down menu for "EGP Finished Gene Prettybase Input File."
4. Enter "5" in "Rare Allele Percentage (integer, 0 to 50)." This filter allows you to display only common SNPs (>5% minor allele frequency) for *ADH1C*.
5. Click on "Run VG2 on the Web!."
6. This will return an image of the genotypes for *ADH1C* 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 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).
7. To save the image to your computer, right-click on the image and choose "save as."



Part 2. Linkage Disequilibrium (LD) Using VG2 in NIEHS SNPs

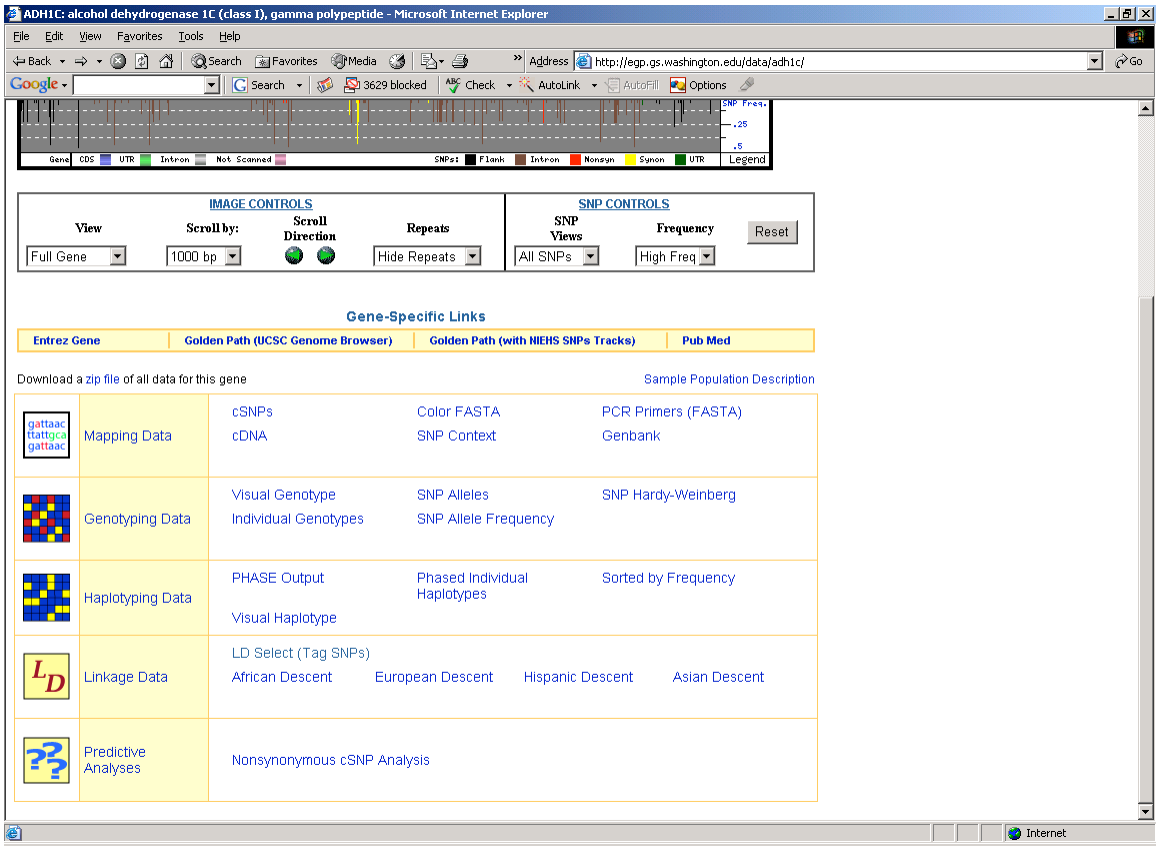
1. With steps 1 through 5 completed from above.
2. Now Choose the LD statistic (r^2) using the pull-down menu "Linkage Disequilibrium Plot."
3. Choose the color of the LD plot (rainbow) using the pull-down menu "LD Shader Mode".
4. Click on "Run VG2 on the Web!."
5. You should have an image of the genotypes and an LD plot appearing in a pop-up window.
6. To save the image to your computer, right-click on the image and choose "save as."
7. The defaults for LD min and max are 0.5 to 1 but you can change this parameter to 0 to 1. Try this option and then again run the default option.



Part 3. TagSNP Selection (LDSelect) in NIEHS SNPs

1. Under "Gene Targets" (within left-hand yellow-colored panel of website), click on "A-Z Finished Genes Directory."
2. Choose "A" and then "ADH1C" to access the gene page data for *ADH1C*.
3. To find the tagSNPs for ADH1C, scroll down the page to the "LD Linkage Data" section.
4. Click on a population for tagSNPs specifically chosen for that population. TagSNPs were chosen for each population from all SNPs regardless of minor allele frequency using the algorithm LDSelect at the default $r^2 > 0.64$. The output lists each bin, and, within each bin, lists SNPs that can be considered tagSNPs to represent that bin in a genetic association study. Only *one* tagSNP per bin is required to represent the genetic diversity of that bin (NOTE: This is different than other tagSNP algorithms based on haplotypes). If a tagSNP is not in a bin (e.g., there is only one SNP in the "bin"), it must be genotyped directly because no other SNP will serve as a sufficient proxy.

Questions: How many bins are in *ADH1C* for the European-descent population? How many tagSNPs must be genotyped directly because they are not contained within a bin with another SNP? Which population sample requires more tagSNPs to represent *ADH1C*: African-descent or Asian-descent?

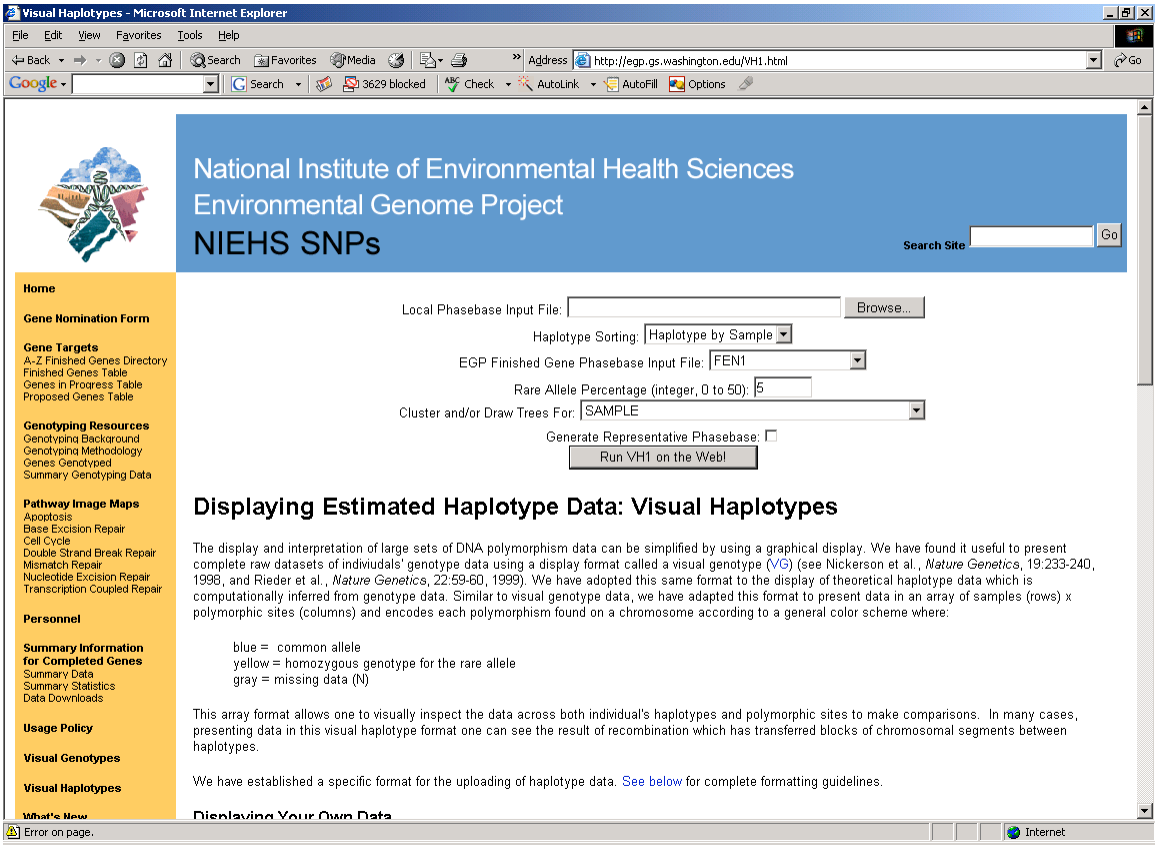


Part 4. Using Visual Haplotype (VH1) for Haplotype tagSNP Selection in NIEHS SNPs

1. Go to http://egp.gs.washington.edu
2. Click on "Visual Haplotypes" (within left-hand yellow-colored panel of website). This software has an interface similar to "Visual Genotypes" but it will display haplotypes. Haplotypes represent the alleles of each SNP assigned to an individual's 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 NIEHS SNPs are X-linked. In this situation, males have one X chromosome and females have two X chromosomes.
3. Choose the gene *FEN1* re-sequenced by NIEHS SNPs using the pull-down menu for "EGP Finished Gene Prettybase Input File."
4. To pick tagSNPs to represent common genetic variation, we suggest you filter by minor allele frequency for common SNPs. Enter 5 in "Rare Allele Percentage (integer, 0 to 50)."

5. To identify the number of haplotypes in your population sample, sort by sample. At '"Cluster By:" choose "SAMPLE."
6. Click on "Run VH1 on the Web!"
7. 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 samples with the same haplotype. 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 the number of common haplotypes manually using VH1.

Questions: How many haplotypes do you have? How many haplotype tagSNPs would you genotype to resolve all common haplotypes?



Part 5. Where to Find Haplotypes in NIEHS SNPs

1. In addition to VH1, we offer PHASEv2.0 output for each of the genes re-sequenced on the NIEHS SNPs website. On the home page, Under "Gene Targets" (within left-hand yellow-colored panel of website), click on "A-Z Finished Genes Directory."
2. Choose 'F' and then *FEN1*.

3. PHASE output is found in the "Haplotyping Data" section of the gene's web page.

Part 6. Downloading Genotype Data from HapMap

1. Go to http://www.hapmap.org
2. Click on "Browse Project Data" on left side of website.
3. In "Landmark or Region" field, type "*VNN2*."
4. Click on "Details." You should see the gene *VNN2* with 6 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 "vnn2x_hapmap_dumped_region" from http://egp.gs.washington.edu/niehs/data_files/datafiles.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. Three markers failed (denoted in red) in the *VNN2* 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 *VNN2* 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 is shown in the figure. Right-click on the 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.

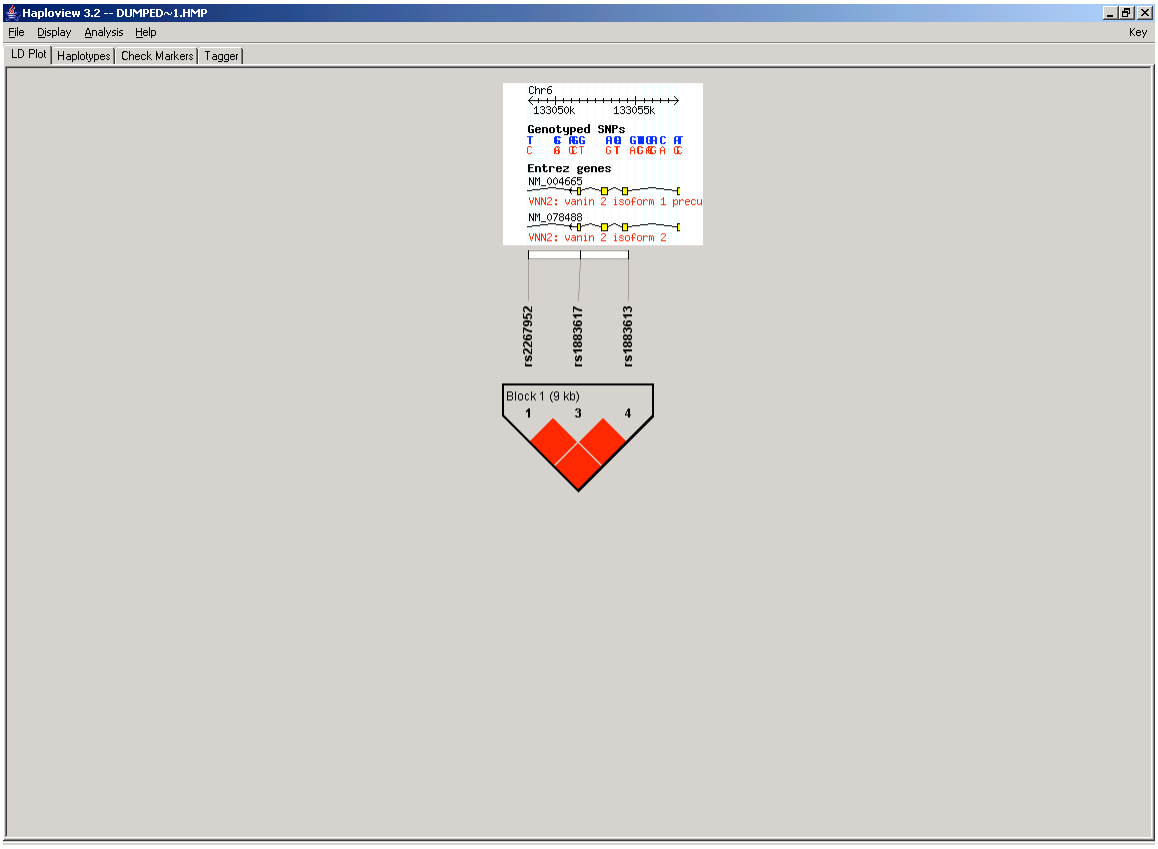
Questions: How far apart are these two markers physically?

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 *VNN2*), 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.

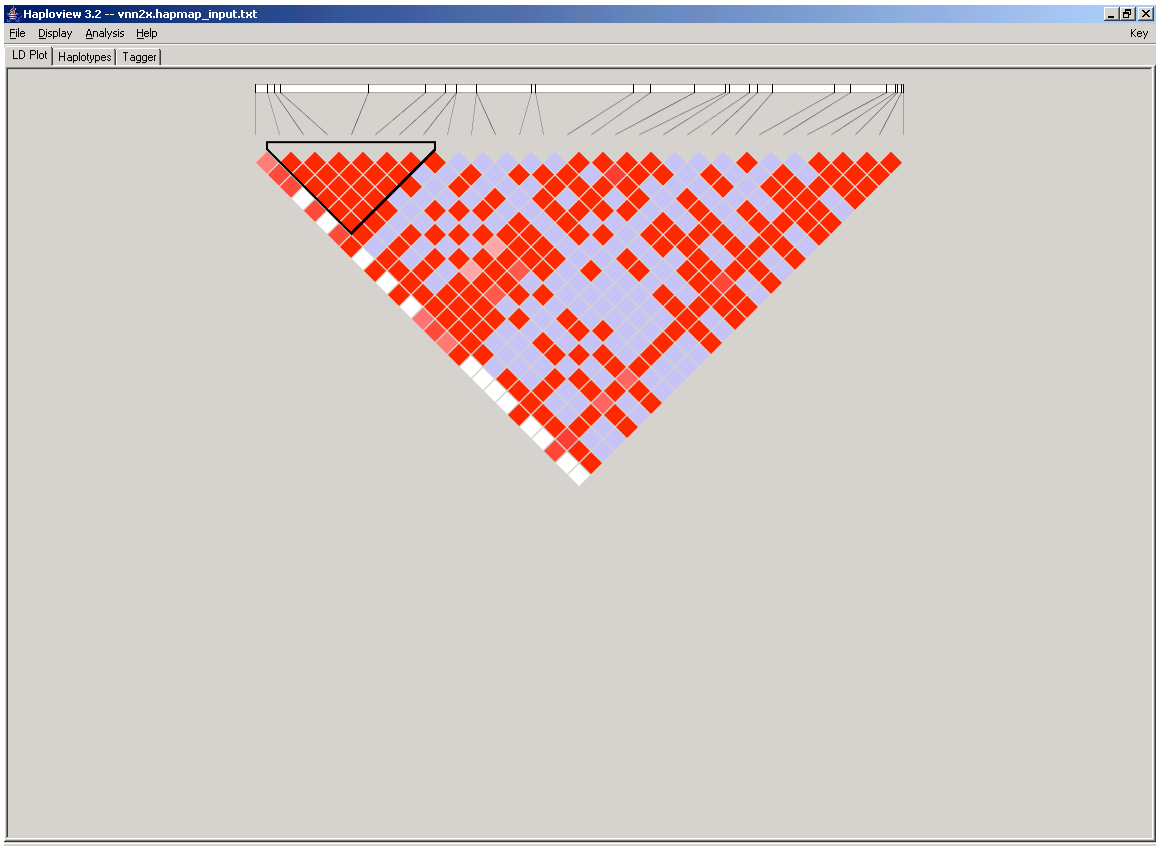
***Questions*:** Using "Tagger" and "pair-wise tagging only," how many Haploview tagSNPs are in *VNN2* for the European-descent HapMap data?



Part 8. Using NIEHS SNPs 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 NIEHS SNPs data for *VNN2* for this exercise, click on "load phased haplotypes."
3. Download "vnn2x.hapmap_input" and "vnn2x_locus_info_file" from http://egp.gs.washington.edu/niehs/data_files/datafiles.html. The input file here is *VNN2* haplotype data (using PHASEv2.1) for European-Americans with a minor allele frequency >10%. 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 sampled variation data.

Questions: How many tagSNPs are identified using pair-wise tagging only in "Tagger" using NIEHS SNPs data for *VNN2*?



Answers to Questions:

How many bins are in *ADH1C* for the European-descent population? ***15 bins: eight bins with >1 SNP; seven "bins" with only one SNP.***

How many tagSNPs must be genotyped directly because they are not contained within a bin with another SNP? *7*

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

How many haplotypes do you have? ***3***

How many haplotype tagSNPs would you genotype to resolve all common haplotypes? ***Probably 2: 1175 and either 995 or 5213.***

How many blocks are in *VNN2* for the European-descent population using the default block definition in Haploview? ***One***

How far apart are these two markers physically? ***4.8kb***

How many haplotypes were identified in this dataset? ***3***

How many haplotype tagging SNPs were identified? ***2***

Using "Tagger" and "pair-wise tagging only," how many Haploview tagSNPs are in *MCP* for the European-descent HapMap data? ***2***

How many tagSNPs are identified using pair-wise tagging only in "Tagger" using NIEHS SNPs data for VNN2? ***9***