

## Picking SNPs Application to Association Studies

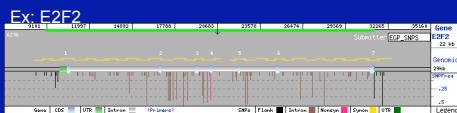
Dana Crawford, PhD

NIEHS  
Environmental Genome Project  
January 30, 2006

## Outline of Tutorial

- Concepts of tagSNPs
- LD and haplotype definitions
- Haplotype blocks and definitions
- Tools to identify tagSNPs

## Why Do We Need tagSNPs?



## Too Many SNPs to Genotype!

Whole Genome:

- 15,000,000 SNPs
- 6,000,000 SNPs  $\geq$  5% MAF

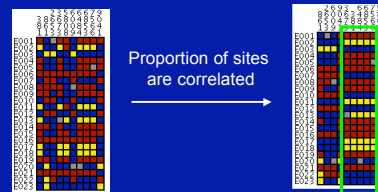
Average Gene:

- 26.5 kb
- 130 SNPs
- 44 SNPs  $\geq$  5% MAF

## SNPs Are Correlated (aka linkage disequilibrium)

"the nonindependence of alleles at different sites." Pritchard and Przeworski 2001

Genotype at one site can predict genotype at another site



## Measuring Pair-wise SNP Correlations

- SNP correlation described by linkage disequilibrium (LD)
- Pair-wise measures of LD:  $D'$  and  $r^2$

$$D = p_{AB} - p_A p_B; D' = D/D_{max} \quad \text{Recombination}$$

$$r^2 = \frac{D^2}{f(A_1)f(A_2)f(B_1)f(B_2)} \quad \text{Power}$$

## LD Statistics: Practical Uses

- $r^2$  is inversely related to power

$$\frac{1}{r^2}$$

1,000 cases	} $r^2=1.0$	1,250 cases	} $r^2 = 0.80$
1,000 controls		1,250 controls	

- $D'$  is related to recombination history

$D' = 1$       no recombination  
 $D' < 1$       historical recombination

## Where to Find Population LD Statistics

For your gene or region of interest, search

- HapMap [www.hapmap.org](http://www.hapmap.org)
- Perlegen [genome.perlegen.com](http://genome.perlegen.com)
- SeattleSNPs PGA [pga.gs.washington.edu](http://pga.gs.washington.edu)
- NIEHS SNPs [egp.gs.washington.edu](http://egp.gs.washington.edu)

## Where to Find Population LD Statistics

For your gene or region of interest, search

- HapMap [www.hapmap.org](http://www.hapmap.org)
- Perlegen [genome.perlegen.com](http://genome.perlegen.com)
- SeattleSNPs PGA [pga.gs.washington.edu](http://pga.gs.washington.edu)
- NIEHS SNPs [egp.gs.washington.edu](http://egp.gs.washington.edu)

## Visualizing Pair-wise LD

National Institute of Environmental Health Sciences  
Environmental Genome Project  
NIEHS SNPs

Welcome to the NIEHS SNPs Program

**Introduction**

The NIEHS Environmental Genome Project is a multi-disciplinary, collaborative effort focused on examining the relationship between environmental exposures, individual sequence variation in human genes, and disease risk in U.S. populations. The NIEHS SNPs Program at the University of Washington is targeted on the systematic identification and genotyping of single nucleotide polymorphisms (SNPs) in environmental response genes. The first phase of the effort focused on finding common sequence variation (SNPs) in human genes involved in DNA repair and cell cycle pathways (see links under Gene Targets in the navigation menu on the left). Ultimately, the project will provide dense genetic maps of human genes that can be applied in evaluating human disease risk with environmental exposures.

**GeneSNPs Database**

NIEHS SNPs are available in the GeneSNPs database as well as the national database resource, dbSNP. GeneSNPs provides a gene-centric map of the genome structure, coding sequences, and identified allelic variation in genes being targeted by a study in disease susceptibility by the NIEHS. This database provides a graphical view of all available SNP data including allele frequencies and genotypes in select populations. This information is key in selecting the polymorphic sites needed to examine disease risk in human population studies.

**Polymorphism Analysis**

Automated DNA sequencing is being used to identify and genotype SNPs in human candidate genes (see dbSNP). Candidate genes are being sequenced to identify common sequence variation for structural analysis and population-based studies. Candidate genes were formerly sequenced across a panel of 90 individuals representative of the U.S. population (see Sample Population Descriptions Panel 1). Candidate genes are now being sequenced across a panel of 95 individuals of known ethnicities (see Sample Population Descriptions Panel 2). All SNPs have been identified using only high-quality sequence data (Q >

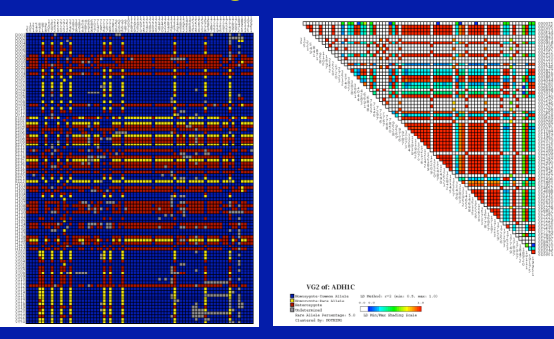
## Visualizing Pair-wise LD

National Institute of Environmental Health Sciences  
Environmental Genome Project  
NIEHS SNPs

Displaying Genotype Data: Visual Genotypes

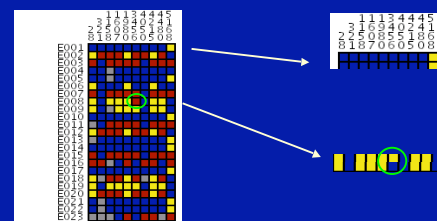
The display and interpretation of large genotype data sets can be simplified by using a visual display. We have found it useful to present complete raw datasets of individual genotype data using a display format called a visual genotype (VG) (see Neuvonen et al., Nature Genetics, 19:233-240, 1998, and Flecker et al., Nature Genetics, 22:86-90, 1999). This format presents all data in an array of squares (rows = a polymorphic sites (columns) and encodes each diallelic polymorphism according to a general color scheme where:

## Visualizing Pair-wise LD



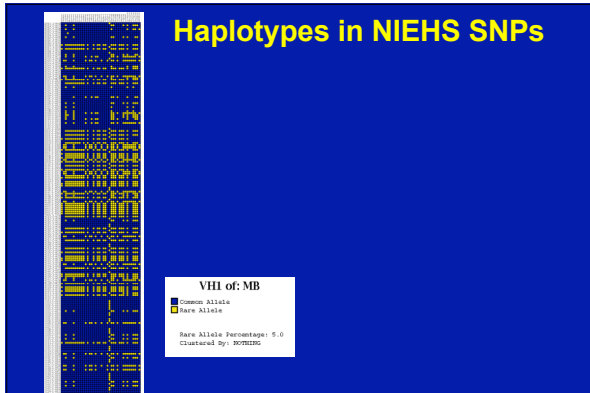
## Multi-SNP Correlations (aka Haplotypes)

"...a unique combination of genetic markers present in a chromosome." pg 57 in Hartl & Clark, 1997









## Haplotypes in NIEHS SNPs

National Institute of Environmental Health Sciences  
Environmental Genome Project  
NIEHS SNPs

Local Phasebase Input File:

Haplotype Cloning: [Haplotype by Sample]

EGP Finished Gene Phasebase Input File: [MB]

Rare Allele Percentage (Integer): [MAXIMUM]

Cluster and/or Draw Trees For: [SAMPLE]

Gene: [NIEHS]

Site: [NIEHS]

Site with Tree:

Site and Sample with Site Tree:

Site and Sample with Sample Tree:

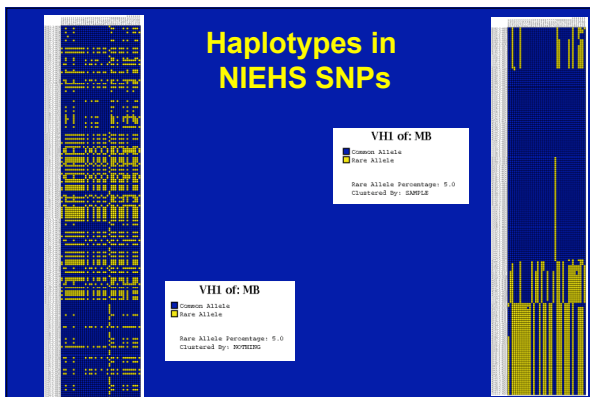
**Displaying Estimated Haplotype**

The display and interpretation of large sets of DNA polymorphism **SITE AND SAMPLE WITH SITE AND SAMPLE TREES**... found it useful to present complete raw datasets of individual genotypes in a display format called a visual genotype (chromosome) as in: Nature Genetics, 19:233-240, 1998, and Foster et al., Nature Genetics, 22:504D, 1999. We have adopted this same format in the display of theoretical haplotype data which is computationally derived from genotype data. Similar to visual genotype data, we have adapted this format to present data in an array of samples (rows) a polymorphic sites (columns) and encodes each polymorphism found on a chromosome according to a general color scheme where:

- blue = common allele
- yellow = chromogenic genotype for the rare allele
- gray = missing data [N]

This array format allows one to visually inspect the data across both individual's haplotypes and polymorphic sites to make comparisons. In many cases, presenting data in this visual haplotype format one can see the result of recombination which has transferred blocks of chromosomal segments between haplotypes.

We have established a specific format for the uploading of haplotype data. See below for complete formatting guidelines.



## Haplotypes in GeneSNPs

Return to NIEHS SNPs Home | Return to EGP Finished Genes Directory

National Institute of Environmental Health Sciences  
Environmental Genome Project  
NIEHS SNPs

**MB:myoglobin**  
Chromosomal Location: 22q13.1

GeneSNPs Image

View: Full Gene | Search by: 1000 bp | Display: Hide Repeats | Repeat: Hide Repeats | SNP Controls: All SNPs | View: High Freq | Reset

Gene-Specific Links

Selected Gene: [Golden Path \(BCSC Genome Browser\)](#) | [Golden Path with NIEHS SNPs Tracks](#) | [Pub Med](#)

Download a zip file of all data for this gene | Sample Population Description

Mapping Data	cdNA	Color FASTA	PCR Primers (FASTA)
Genotyping Data	Visual Genotype	SNP Alleles	SNP Hardy-Weinberg
	Individual Genotypes	SNP Allele Frequency	

## Haplotypes in GeneSNPs

GeneSNPs | My GeneSNPs | Gene Lists | EGP Lists | Tools | Open Query

**GeneSNPs**

**About**

Sponsored by The National Institute of Environmental Health Sciences  
Developed by The National Center for Human Genome Research

The Environmental Genome Project web resource integrates gene, sequence and polymorphism data into a richly annotated gene profile. The browser genes included are related to DNA repair, cell cycle control, cell signaling, cell division, locomotion and metabolism, and are thought to play a role in susceptibility to environmental exposure.

**What's New**

Last Updated: 04-2006

- 04-2006
  - Workflows: Report Options Added to SNPTable
  - SNPs: Locus dBSNP Added
- 12-2005
  - NIEHS: New EGP Data for 71801
  - NIEHS: New EGP Data for 71802
  - NIEHS: New EGP Data for 71803
  - NIEHS: New EGP Data for 71804
  - NIEHS: New EGP Data for 71805
  - NIEHS: New EGP Data for 71806
  - NIEHS: New EGP Data for 71807
  - NIEHS: New EGP Data for 71808
  - NIEHS: New EGP Data for 71809
  - NIEHS: New EGP Data for 71810
  - NIEHS: New EGP Data for 71811
  - NIEHS: New EGP Data for 71812
  - NIEHS: New EGP Data for 71813
  - NIEHS: New EGP Data for 71814
  - NIEHS: New EGP Data for 71815
  - NIEHS: New EGP Data for 71816
  - NIEHS: New EGP Data for 71817
  - NIEHS: New EGP Data for 71818
  - NIEHS: New EGP Data for 71819
  - NIEHS: New EGP Data for 71820
  - NIEHS: New EGP Data for 71821
  - NIEHS: New EGP Data for 71822
  - NIEHS: New EGP Data for 71823
  - NIEHS: New EGP Data for 71824
  - NIEHS: New EGP Data for 71825
  - NIEHS: New EGP Data for 71826
  - NIEHS: New EGP Data for 71827
  - NIEHS: New EGP Data for 71828
  - NIEHS: New EGP Data for 71829
  - NIEHS: New EGP Data for 71830
  - NIEHS: New EGP Data for 71831
  - NIEHS: New EGP Data for 71832
  - NIEHS: New EGP Data for 71833
  - NIEHS: New EGP Data for 71834
  - NIEHS: New EGP Data for 71835
  - NIEHS: New EGP Data for 71836
  - NIEHS: New EGP Data for 71837
  - NIEHS: New EGP Data for 71838
  - NIEHS: New EGP Data for 71839
  - NIEHS: New EGP Data for 71840
  - NIEHS: New EGP Data for 71841
  - NIEHS: New EGP Data for 71842
  - NIEHS: New EGP Data for 71843
  - NIEHS: New EGP Data for 71844
  - NIEHS: New EGP Data for 71845
  - NIEHS: New EGP Data for 71846
  - NIEHS: New EGP Data for 71847
  - NIEHS: New EGP Data for 71848
  - NIEHS: New EGP Data for 71849
  - NIEHS: New EGP Data for 71850
  - NIEHS: New EGP Data for 71851
  - NIEHS: New EGP Data for 71852
  - NIEHS: New EGP Data for 71853
  - NIEHS: New EGP Data for 71854
  - NIEHS: New EGP Data for 71855
  - NIEHS: New EGP Data for 71856
  - NIEHS: New EGP Data for 71857
  - NIEHS: New EGP Data for 71858
  - NIEHS: New EGP Data for 71859
  - NIEHS: New EGP Data for 71860
  - NIEHS: New EGP Data for 71861
  - NIEHS: New EGP Data for 71862
  - NIEHS: New EGP Data for 71863
  - NIEHS: New EGP Data for 71864
  - NIEHS: New EGP Data for 71865
  - NIEHS: New EGP Data for 71866
  - NIEHS: New EGP Data for 71867
  - NIEHS: New EGP Data for 71868
  - NIEHS: New EGP Data for 71869
  - NIEHS: New EGP Data for 71870
  - NIEHS: New EGP Data for 71871
  - NIEHS: New EGP Data for 71872
  - NIEHS: New EGP Data for 71873
  - NIEHS: New EGP Data for 71874
  - NIEHS: New EGP Data for 71875
  - NIEHS: New EGP Data for 71876
  - NIEHS: New EGP Data for 71877
  - NIEHS: New EGP Data for 71878
  - NIEHS: New EGP Data for 71879
  - NIEHS: New EGP Data for 71880
  - NIEHS: New EGP Data for 71881
  - NIEHS: New EGP Data for 71882
  - NIEHS: New EGP Data for 71883
  - NIEHS: New EGP Data for 71884
  - NIEHS: New EGP Data for 71885
  - NIEHS: New EGP Data for 71886
  - NIEHS: New EGP Data for 71887
  - NIEHS: New EGP Data for 71888
  - NIEHS: New EGP Data for 71889
  - NIEHS: New EGP Data for 71890
  - NIEHS: New EGP Data for 71891
  - NIEHS: New EGP Data for 71892
  - NIEHS: New EGP Data for 71893
  - NIEHS: New EGP Data for 71894
  - NIEHS: New EGP Data for 71895
  - NIEHS: New EGP Data for 71896
  - NIEHS: New EGP Data for 71897
  - NIEHS: New EGP Data for 71898
  - NIEHS: New EGP Data for 71899
  - NIEHS: New EGP Data for 71900
  - NIEHS: New EGP Data for 71901
  - NIEHS: New EGP Data for 71902
  - NIEHS: New EGP Data for 71903
  - NIEHS: New EGP Data for 71904
  - NIEHS: New EGP Data for 71905
  - NIEHS: New EGP Data for 71906
  - NIEHS: New EGP Data for 71907
  - NIEHS: New EGP Data for 71908
  - NIEHS: New EGP Data for 71909
  - NIEHS: New EGP Data for 71910
  - NIEHS: New EGP Data for 71911
  - NIEHS: New EGP Data for 71912
  - NIEHS: New EGP Data for 71913
  - NIEHS: New EGP Data for 71914
  - NIEHS: New EGP Data for 71915
  - NIEHS: New EGP Data for 71916
  - NIEHS: New EGP Data for 71917
  - NIEHS: New EGP Data for 71918
  - NIEHS: New EGP Data for 71919
  - NIEHS: New EGP Data for 71920
  - NIEHS: New EGP Data for 71921
  - NIEHS: New EGP Data for 71922
  - NIEHS: New EGP Data for 71923
  - NIEHS: New EGP Data for 71924
  - NIEHS: New EGP Data for 71925
  - NIEHS: New EGP Data for 71926
  - NIEHS: New EGP Data for 71927
  - NIEHS: New EGP Data for 71928
  - NIEHS: New EGP Data for 71929
  - NIEHS: New EGP Data for 71930
  - NIEHS: New EGP Data for 71931
  - NIEHS: New EGP Data for 71932
  - NIEHS: New EGP Data for 71933
  - NIEHS: New EGP Data for 71934
  - NIEHS: New EGP Data for 71935
  - NIEHS: New EGP Data for 71936
  - NIEHS: New EGP Data for 71937
  - NIEHS: New EGP Data for 71938
  - NIEHS: New EGP Data for 71939
  - NIEHS: New EGP Data for 71940
  - NIEHS: New EGP Data for 71941
  - NIEHS: New EGP Data for 71942
  - NIEHS: New EGP Data for 71943
  - NIEHS: New EGP Data for 71944
  - NIEHS: New EGP Data for 71945
  - NIEHS: New EGP Data for 71946
  - NIEHS: New EGP Data for 71947
  - NIEHS: New EGP Data for 71948
  - NIEHS: New EGP Data for 71949
  - NIEHS: New EGP Data for 71950
  - NIEHS: New EGP Data for 71951
  - NIEHS: New EGP Data for 71952
  - NIEHS: New EGP Data for 71953
  - NIEHS: New EGP Data for 71954
  - NIEHS: New EGP Data for 71955
  - NIEHS: New EGP Data for 71956
  - NIEHS: New EGP Data for 71957
  - NIEHS: New EGP Data for 71958
  - NIEHS: New EGP Data for 71959
  - NIEHS: New EGP Data for 71960
  - NIEHS: New EGP Data for 71961
  - NIEHS: New EGP Data for 71962
  - NIEHS: New EGP Data for 71963
  - NIEHS: New EGP Data for 71964
  - NIEHS: New EGP Data for 71965
  - NIEHS: New EGP Data for 71966
  - NIEHS: New EGP Data for 71967
  - NIEHS: New EGP Data for 71968
  - NIEHS: New EGP Data for 71969
  - NIEHS: New EGP Data for 71970
  - NIEHS: New EGP Data for 71971
  - NIEHS: New EGP Data for 71972
  - NIEHS: New EGP Data for 71973
  - NIEHS: New EGP Data for 71974
  - NIEHS: New EGP Data for 71975
  - NIEHS: New EGP Data for 71976
  - NIEHS: New EGP Data for 71977
  - NIEHS: New EGP Data for 71978
  - NIEHS: New EGP Data for 71979
  - NIEHS: New EGP Data for 71980
  - NIEHS: New EGP Data for 71981
  - NIEHS: New EGP Data for 71982
  - NIEHS: New EGP Data for 71983
  - NIEHS: New EGP Data for 71984
  - NIEHS: New EGP Data for 71985
  - NIEHS: New EGP Data for 71986
  - NIEHS: New EGP Data for 71987
  - NIEHS: New EGP Data for 71988
  - NIEHS: New EGP Data for 71989
  - NIEHS: New EGP Data for 71990
  - NIEHS: New EGP Data for 71991
  - NIEHS: New EGP Data for 71992
  - NIEHS: New EGP Data for 71993
  - NIEHS: New EGP Data for 71994
  - NIEHS: New EGP Data for 71995
  - NIEHS: New EGP Data for 71996
  - NIEHS: New EGP Data for 71997
  - NIEHS: New EGP Data for 71998
  - NIEHS: New EGP Data for 71999
  - NIEHS: New EGP Data for 72000
- 12-2005
  - NIEHS: New EGP Data for 92848A

## Haplotypes in GeneSNPs

GeneSNPs | My GeneSNPs | Gene Lists | EGP Lists | Tools | Open Query

**GeneSNPs**

**About**

Sponsored by The National Institute of Environmental Health Sciences  
Developed by The National Center for Human Genome Research

The Environmental Genome Project web resource integrates gene, sequence and polymorphism data into a richly annotated gene profile. The browser genes included are related to DNA repair, cell cycle control, cell signaling, cell division, locomotion and metabolism, and are thought to play a role in susceptibility to environmental exposure.

**What's New**

Last Updated: 04-2006

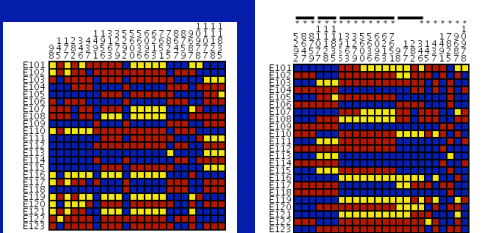
- 04-2006
  - Workflows: Report Options Added to SNPTable
  - SNPs: Locus dBSNP Added
- 12-2005
  - NIEHS: New EGP Data for 71801
  - NIEHS: New EGP Data for 71802
  - NIEHS: New EGP Data for 71803
  - NIEHS: New EGP Data for 71804
  - NIEHS: New EGP Data for 71805
  - NIEHS: New EGP Data for 71806
  - NIEHS: New EGP Data for 71807
  - NIEHS: New EGP Data for 71808
  - NIEHS: New EGP Data for 71809
  - NIEHS: New EGP Data for 71810
  - NIEHS: New EGP Data for 71811
  - NIEHS: New EGP Data for 71812
  - NIEHS: New EGP Data for 71813
  - NIEHS: New EGP Data for 71814
  - NIEHS: New EGP Data for 71815
  - NIEHS: New EGP Data for 71816
  - NIEHS: New EGP Data for 71817
  - NIEHS: New EGP Data for 71818
  - NIEHS: New EGP Data for 71819
  - NIEHS: New EGP Data for 71820
  - NIEHS: New EGP Data for 71821
  - NIEHS: New EGP Data for 71822
  - NIEHS: New EGP Data for 71823
  - NIEHS: New EGP Data for 71824
  - NIEHS: New EGP Data for 71825
  - NIEHS: New EGP Data for 71826
  - NIEHS: New EGP Data for 71827
  - NIEHS: New EGP Data for 71828
  - NIEHS: New EGP Data for 71829
  - NIEHS: New EGP Data for 71830
  - NIEHS: New EGP Data for 71831
  - NIEHS: New EGP Data for 71832
  - NIEHS: New EGP Data for 71833
  - NIEHS: New EGP Data for 71834
  - NIEHS: New EGP Data for 71835
  - NIEHS: New EGP Data for 71836
  - NIEHS: New EGP Data for 71837
  - NIEHS: New EGP Data for 71838
  - NIEHS: New EGP Data for 71839
  - NIEHS: New EGP Data for 71840
  - NIEHS: New EGP Data for 71841
  - NIEHS: New EGP Data for 71842
  - NIEHS: New EGP Data for 71843
  - NIEHS: New EGP Data for 71844
  - NIEHS: New EGP Data for 71845
  - NIEHS: New EGP Data for 71846
  - NIEHS: New EGP Data for 71847
  - NIEHS: New EGP Data for 71848
  - NIEHS: New EGP Data for 71849
  - NIEHS: New EGP Data for 71850
  - NIEHS: New EGP Data for 71851
  - NIEHS: New EGP Data for 71852
  - NIEHS: New EGP Data for 71853
  - NIEHS: New EGP Data for 71854
  - NIEHS: New EGP Data for 71855
  - NIEHS: New EGP Data for 71856
  - NIEHS: New EGP Data for 71857
  - NIEHS: New EGP Data for 71858
  - NIEHS: New EGP Data for 71859
  - NIEHS: New EGP Data for 71860
  - NIEHS: New EGP Data for 71861
  - NIEHS: New EGP Data for 71862
  - NIEHS: New EGP Data for 71863
  - NIEHS: New EGP Data for 71864
  - NIEHS: New EGP Data for 71865
  - NIEHS: New EGP Data for 71866
  - NIEHS: New EGP Data for 71867
  - NIEHS: New EGP Data for 71868
  - NIEHS: New EGP Data for 71869
  - NIEHS: New EGP Data for 71870
  - NIEHS: New EGP Data for 71871
  - NIEHS: New EGP Data for 71872
  - NIEHS: New EGP Data for 71873
  - NIEHS: New EGP Data for 71874
  - NIEHS: New EGP Data for 71875
  - NIEHS: New EGP Data for 71876
  - NIEHS: New EGP Data for 71877
  - NIEHS: New EGP Data for 71878
  - NIEHS: New EGP Data for 71879
  - NIEHS: New EGP Data for 71880
  - NIEHS: New EGP Data for 71881
  - NIEHS: New EGP Data for 71882
  - NIEHS: New EGP Data for 71883
  - NIEHS: New EGP Data for 71884
  - NIEHS: New EGP Data for 71885
  - NIEHS: New EGP Data for 71886
  - NIEHS: New EGP Data for 71887
  - NIEHS: New EGP Data for 71888
  - NIEHS: New EGP Data for 71889
  - NIEHS: New EGP Data for 71890
  - NIEHS: New EGP Data for 71891
  - NIEHS: New EGP Data for 71892
  - NIEHS: New EGP Data for 71893
  - NIEHS: New EGP Data for 71894
  - NIEHS: New EGP Data for 71895
  - NIEHS: New EGP Data for 71896
  - NIEHS: New EGP Data for 71897
  - NIEHS: New EGP Data for 71898
  - NIEHS: New EGP Data for 71899
  - NIEHS: New EGP Data for 71900
  - NIEHS: New EGP Data for 71901
  - NIEHS: New EGP Data for 71902
  - NIEHS: New EGP Data for 71903
  - NIEHS: New EGP Data for 71904
  - NIEHS: New EGP Data for 71905
  - NIEHS: New EGP Data for 71906
  - NIEHS: New EGP Data for 71907
  - NIEHS: New EGP Data for 71908
  - NIEHS: New EGP Data for 71909
  - NIEHS: New EGP Data for 71910
  - NIEHS: New EGP Data for 71911
  - NIEHS: New EGP Data for 71912
  - NIEHS: New EGP Data for 71913
  - NIEHS: New EGP Data for 71914
  - NIEHS: New EGP Data for 71915
  - NIEHS: New EGP Data for 71916
  - NIEHS: New EGP Data for 71917
  - NIEHS: New EGP Data for 71918
  - NIEHS: New EGP Data for 71919
  - NIEHS: New EGP Data for 71920
  - NIEHS: New EGP Data for 71921
  - NIEHS: New EGP Data for 71922
  - NIEHS: New EGP Data for 71923
  - NIEHS: New EGP Data for 71924
  - NIEHS: New EGP Data for 71925
  - NIEHS: New EGP Data for 71926
  - NIEHS: New EGP Data for 71927
  - NIEHS: New EGP Data for 71928
  - NIEHS: New EGP Data for 71929
  - NIEHS: New EGP Data for 71930
  - NIEHS: New EGP Data for 71931
  - NIEHS: New EGP Data for 71932
  - NIEHS: New EGP Data for 71933
  - NIEHS: New EGP Data for 71934
  - NIEHS: New EGP Data for 71935
  - NIEHS: New EGP Data for 71936
  - NIEHS: New EGP Data for 71937
  - NIEHS: New EGP Data for 71938
  - NIEHS: New EGP Data for 71939
  - NIEHS: New EGP Data for 71940
  - NIEHS: New EGP Data for 71941
  - NIEHS: New EGP Data for 71942
  - NIEHS: New EGP Data for 71943
  - NIEHS: New EGP Data for 71944
  - NIEHS: New EGP Data for 71945
  - NIEHS: New EGP Data for 71946
  - NIEHS: New EGP Data for 71947
  - NIEHS: New EGP Data for 71948
  - NIEHS: New EGP Data for 71949
  - NIEHS: New EGP Data for 71950
  - NIEHS: New EGP Data for 71951
  - NIEHS: New EGP Data for 71952
  - NIEHS: New EGP Data for 71953
  - NIEHS: New EGP Data for 71954
  - NIEHS: New EGP Data for 71955
  - NIEHS: New EGP Data for 71956
  - NIEHS: New EGP Data for 71957
  - NIEHS: New EGP Data for 71958
  - NIEHS: New EGP Data for 71959
  - NIEHS: New EGP Data for 71960
  - NIEHS: New EGP Data for 71961
  - NIEHS: New EGP Data for 71962
  - NIEHS: New EGP Data for 71963
  - NIEHS: New EGP Data for 71964
  - NIEHS: New EGP Data for 71965
  - NIEHS: New EGP Data for 71966
  - NIEHS: New EGP Data for 71967
  - NIEHS: New EGP Data for 71968
  - NIEHS: New EGP Data for 71969
  - NIEHS: New EGP Data for 71970
  - NIEHS: New EGP Data for 71971
  - NIEHS: New EGP Data for 71972
  - NIEHS: New EGP Data for 71973
  - NIEHS: New EGP Data for 71974
  - NIEHS: New EGP Data for 71975
  - NIEHS: New EGP Data for 71976
  - NIEHS: New EGP Data for 71977
  - NIEHS: New EGP Data for 71978
  - NIEHS: New EGP Data for 71979
  - NIEHS: New EGP Data for 71980
  - NIEHS: New EGP Data for 71981
  - NIEHS: New EGP Data for 71982
  - NIEHS: New EGP Data for 71983
  - NIEHS: New EGP Data for 71984
  - NIEHS: New EGP Data for 71985
  - NIEHS: New EGP Data for 71986
  - NIEHS: New EGP Data for 71987
  - NIEHS: New EGP Data for 71988
  - NIEHS: New EGP Data for 71989
  - NIEHS: New EGP Data for 71990
  - NIEHS: New EGP Data for 71991
  - NIEHS: New EGP Data for 71992
  - NIEHS: New EGP Data for 71993
  - NIEHS: New EGP Data for 71994
  - NIEHS: New EGP Data for 71995
  - NIEHS: New EGP Data for 71996
  - NIEHS: New EGP Data for 71997
  - NIEHS: New EGP Data for 71998
  - NIEHS: New EGP Data for 71999
  - NIEHS: New EGP Data for 72000
- 12-2005
  - NIEHS: New EGP Data for 92848A



## LDSelect: Using LD to Pick tagSNPs

### LDSelect

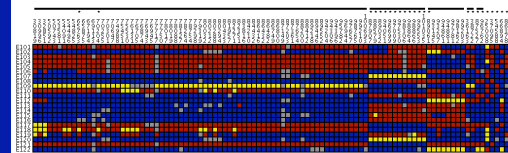
- Uses SNP discovery data (not haplotypes)
- Finds all correlated SNPs to minimize the total number
- Maintains genetic diversity of locus



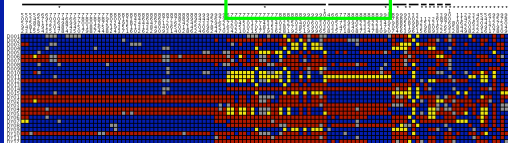
Carlson et al. AJHG (2004)

## TagSNPs Are Population Specific

### European-descent (BLM)



### African-descent (BLM)



## SNP Selection: tagSNP Data

**BLM**

View  
Full Gene

Enter Gene

Download a zip file of:

- Mapping
- Genotype
- Haplotype
- Linkage

Site 1	rs1041103	11	494446	AGTAA	1000
Site 2	rs1041103	11	494446	AGTAA	1000
Site 3	rs1041103	11	494446	AGTAA	1000
Site 4	rs1041103	11	494446	AGTAA	1000
Site 5	rs1041103	11	494446	AGTAA	1000
Site 6	rs1041103	11	494446	AGTAA	1000
Site 7	rs1041103	11	494446	AGTAA	1000
Site 8	rs1041103	11	494446	AGTAA	1000
Site 9	rs1041103	11	494446	AGTAA	1000
Site 10	rs1041103	11	494446	AGTAA	1000
Site 11	rs1041103	11	494446	AGTAA	1000
Site 12	rs1041103	11	494446	AGTAA	1000

## Side Note: Categorizing tagSNPs

- SNP context  
Nonrepetitive > repetitive
- Location of SNP  
Coding > noncoding
- Function  
Nonsynonymous > synonymous

## Categorizing tagSNPs

**LPO**

View  
Full Gene

Enter Gene

Download a zip file of all data for this gene

Mapping Data

Hugo	AllelePos	a.a.Pos	Residue	Variant	Freq	PPHpredict	SIFTpredict
lpo	010050	105	T	I	0.03	possibly damaging	TOLERATED
lpo	013794	244	A	T	0.01	benign	TOLERATED
lpo	018169	414	R	Q	0.01	probably damaging	INTOLERANT
lpo	018189	421	V	M	0.11	benign	TOLERATED
lpo	023597	514	R	Q	0.01	probably damaging	INTOLERANT
lpo	030719	614	I	T	0.01	possibly damaging	INTOLERANT
lpo	031176	700	D	N	0.01	benign	INTOLERANT

## Haplotypes in Genetic Association Studies

Two main approaches with haplotypes:

Haplotypes → Pick tagSNPs → Genotype samples

Pick tagSNPs → Infer haplotypes → Test for association

## Haplotypes in Genetic Association Studies

Two main approaches with haplotypes:

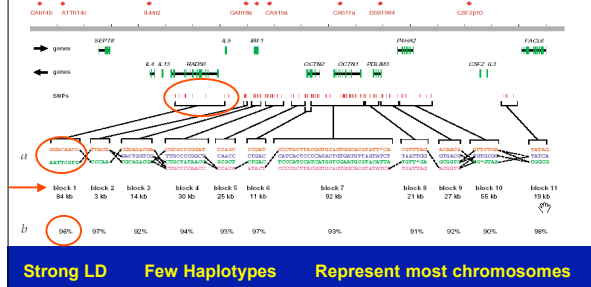
Haplotypes → Pick tagSNPs → Genotype samples

Recombination  
Natural selection  
Population history  
Population demography

Pick tagSNPs → Infer haplotypes → Test for association

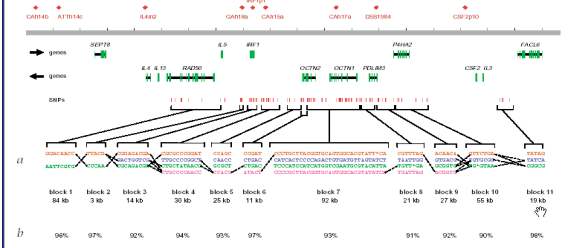
## Haplotype "Blocks"

Daly et al Nat. Genet. (2001)



## Block Definitions

Daly et al Nat. Genet. (2001)



$D'$  [Gabriel et al Science (2002)]

## Block Definitions

Four-gamete test:



<4 haplotypes,  $D' = 1$  → block

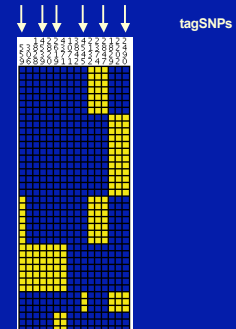
4 haplotypes,  $D' < 1$  → boundary

## Haplotype Blocks and tagSNPs

Identifying blocks and tagSNPs:

- Manually  
Visual haplotype
- Algorithms  
HapMap and Haploview

## Haplotype Blocks and tagSNPs



LTA:  
16 SNPs (MAF > 10%)  
6 "common" haplotypes



# Haplotype Blocks and tagSNPs

## Identifying blocks and tagSNPs:

- Manually Visual Haplotype
- Algorithms HapMap and HaploView

# HapMap Data and Haploview

**International HapMap Project**  
Home | About the Project | Data | Publications | Tutorial

The International HapMap Project is a partnership of scientists and funding agencies from Canada, China, Japan, Nigeria, the United Kingdom and the United States to develop a public resource that will help researchers find genes associated with human disease and response to pharmaceuticals. See "About the International HapMap Project" for more information.

**Project Information**  
About the Project  
HapMap Publications  
HapMap Tutorial  
HapMap Mapping List  
HapMap Project Participants  
HapMap Mirror Site in Japan

**News**

- 2005-10-24: **HapMap Public Release #19**  
Genotypes, frequencies, and assays for phase I and phase II of the HapMap project are now available for bulk download. The files contain all phase I and II data combined.
- 2005-09-28: **Recombination rates and hotspots for HapMap Phase I dataset released**  
Recombination rates and hotspot data is now available for bulk download.
- 2005-09-28: **HapMap Public Release #18**  
Genotypes, frequency, and assays for the chromosomes 2, 8, 11, 14, 15 and 21 is now available for bulk download only. Please note that this represents complete phase II data on these chromosomes for CEU and YRI samples. This is a dump with all phase I and II data combined. Genotype summary for SNPs in this release is as follows:

Populations	CEU	CHB	JPT	YRI
Total SNPs	3,294,260	3,262,524	3,262,222	3,099,520
Total Genotype Data	5,094,684	5,019,390	5,012,360	4,387,448

**Helpful Links**  
HapMap Project Press Release  
NHGRI HapMap Page  
NCBI Variation Database (dbSNP)  
Japanese SNP Database (JSNP)

**www.hapmap.org**

**International HapMap Project**  
Home | About the Project | Data | Publications | Tutorial

**Instructions:** Search using a sequence name, gene name, locus, or other landmark. The wildcard character \* is allowed. To center on a location, click the ruler. Use the ScrollZoom buttons to change magnification and position.

Examples: Chr20, Chr9:660,000,760,000, SNP:r4677660, NM\_153254, BRCA2, Srs11, ENH010

**Search**  
[Locus] [tag SNPs] [Phased Haplotypes] [Genotype data] [Frequency data] [Symbols and colours used]

**Landmark or Region**  
[Search] [Reset] [Reports & Analysis] [Annotate LD Plot] [Configure] [Go]

**Data Source**  
[HapMap Data Phase I] [Phase I] [Phase II] [dbSNP] [dbSNP] [dbSNP]

**Population descriptors:** YRI Yoruba in Ibadan, Nigeria, JPT Japanese in Tokyo, Japan, CHB Han Chinese in Beijing, China, CEU CEPH Gish residents with ancestry from northern and western Europe

**Overview** [Update Image]

For performing in depth LD and Haplotype analysis of genotype data install Haploview in your local machine. Haploview ver 3.32 is now available for download.

**Tracks** [Display Settings] [Add your own tracks]

**International HapMap Project**  
Home | About the Project | Data | Publications | Tutorial

Showing 19.59 kbp from chr6, positions 133,045,579 to 133,065,167

**Instructions:** Search using a sequence name, gene name, locus, or other landmark. The wildcard character \* is allowed. To center on a location, click the ruler. Use the ScrollZoom buttons to change magnification and position.

Examples: Chr20, Chr9:660,000,760,000, SNP:r4677660, NM\_153254, BRCA2, Srs11, ENH010

**Search**  
[Locus] [tag SNPs] [Phased Haplotypes] [Genotype data] [Frequency data] [Symbols and colours used]

**Landmark or Region**  
[Search] [Reset] [Reports & Analysis] [Annotate LD Plot] [Configure] [Go]

**Data Source**  
[HapMap Data Phase I] [Phase I] [Phase II] [dbSNP] [dbSNP] [dbSNP]

**Population descriptors:** YRI Yoruba in Ibadan, Nigeria, JPT Japanese in Tokyo, Japan, CHB Han Chinese in Beijing, China, CEU CEPH Gish residents with ancestry from northern and western Europe

**Overview**  
[Update Image]

For performing in depth LD and Haplotype analysis of genotype data install Haploview in your local machine. Haploview ver 3.32 is now available for download.

**Tracks** [Display Settings] [Add your own tracks]

**International HapMap Project**  
Home | About the Project | Data | Publications | Tutorial

Showing 19.59 kbp from chr6, positions 133,045,579 to 133,065,167

**Instructions:** Search using a sequence name, gene name, locus, or other landmark. The wildcard character \* is allowed. To center on a location, click the ruler. Use the ScrollZoom buttons to change magnification and position.

Examples: Chr20, Chr9:660,000,760,000, SNP:r4677660, NM\_153254, BRCA2, Srs11, ENH010

**Search**  
[Locus] [tag SNPs] [Phased Haplotypes] [Genotype data] [Frequency data] [Symbols and colours used]

**Landmark or Region**  
[Search] [Reset] [Reports & Analysis] [Annotate LD Plot] [Configure] [Go]

**Data Source**  
[HapMap Data Phase I] [Phase I] [Phase II] [dbSNP] [dbSNP] [dbSNP]

**Population descriptors:** YRI Yoruba in Ibadan, Nigeria, JPT Japanese in Tokyo, Japan, CHB Han Chinese in Beijing, China, CEU CEPH Gish residents with ancestry from northern and western Europe

**Overview**  
[Update Image]

For performing in depth LD and Haplotype analysis of genotype data install Haploview in your local machine. Haploview ver 3.32 is now available for download.

**Tracks** [Display Settings] [Add your own tracks]

**International HapMap Project**  
Home | About the Project | Data | Publications | Tutorial

Showing 19.59 kbp from chr6, positions 133,045,579 to 133,065,167

**Instructions:** Search using a sequence name, gene name, locus, or other landmark. The wildcard character \* is allowed. To center on a location, click the ruler. Use the ScrollZoom buttons to change magnification and position.

Examples: Chr20, Chr9:660,000,760,000, SNP:r4677660, NM\_153254, BRCA2, Srs11, ENH010

**Search**  
[Locus] [tag SNPs] [Phased Haplotypes] [Genotype data] [Frequency data] [Symbols and colours used]

**Landmark or Region**  
[Search] [Reset] [Reports & Analysis] [Annotate LD Plot] [Configure] [Go]

**Data Source**  
[HapMap Data Phase I] [Phase I] [Phase II] [dbSNP] [dbSNP] [dbSNP]

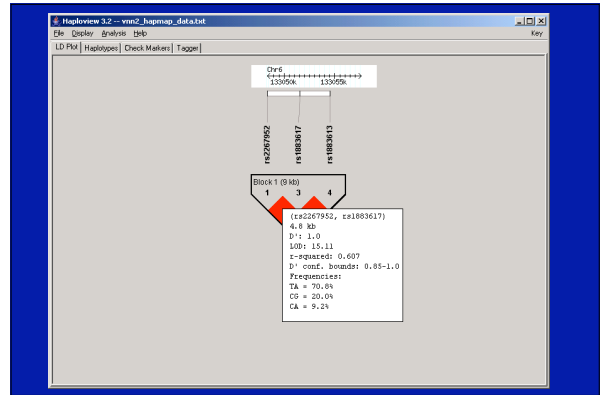
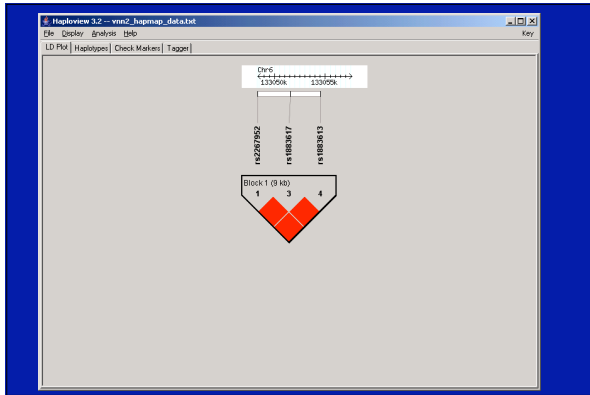
**Population descriptors:** YRI Yoruba in Ibadan, Nigeria, JPT Japanese in Tokyo, Japan, CHB Han Chinese in Beijing, China, CEU CEPH Gish residents with ancestry from northern and western Europe

**Overview**  
[Update Image]

For performing in depth LD and Haplotype analysis of genotype data install Haploview in your local machine. Haploview ver 3.32 is now available for download.

**Tracks** [Display Settings] [Add your own tracks]





Export data

Tab to Export  
 LD  
 Haplotypes  
 Data Checks

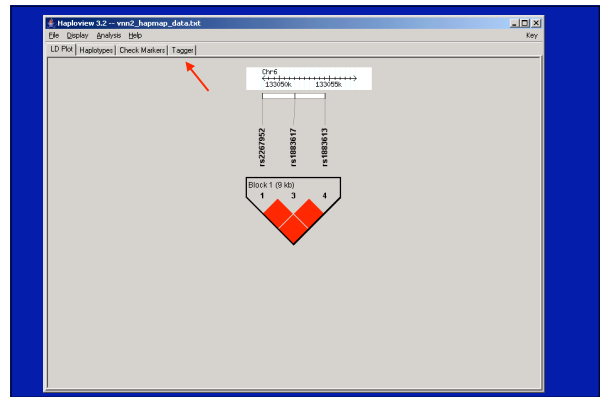
Output Format  
 Text  PNG Image  Compress image (smaller file)

Range  
 All  Marker  No  Adjacent markers only

OK Cancel

BLOCK 1, MARKERS: 1! 3! 4  
414 (0.706)  
232 (0.200)  
214 (0.092)

May not be minimal set



Configuration Results

#	Name	Position	Force Include	Force Exclude	Capture this Allele?
1	rs2267952	133048252	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3	rs1883617	133053220	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
4	rs1883613	133057933	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Minimal set of tagSNPs based on  $r^2$

pairwise tagging only  $r^2$  threshold [0.8]  
 aggressive tagging: use 2 marker haplotypes  
 aggressive tagging: use 2- and 3-marker haplotypes LOD threshold for multi-marker tests [3.0]

Run Tagger Reset Table

Tests

Allele	Test	r	p
rs2267952	rs2267952	1.0	
rs1883617	rs1883617	1.0	
rs1883613	rs1883617	1.0	

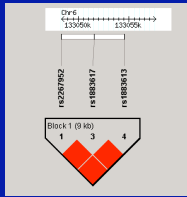
Alleles captured by Current Selection  
rs1883613  
rs1883613

Captured 3 alleles with mean  $r^2$  of 1.0  
Captured 100 percent of alleles with  $r^2 > 0.8$   
Using 2 SNPs in 2 tests.

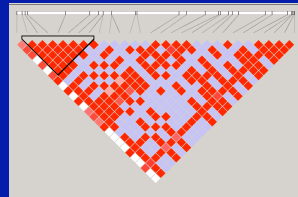
Dump Tests File

Note: HapMap is not complete variation data

## Variation data, LD, and tagSNPs for VNN2 in European-Americans



HapMap  
3 tagSNPs



NIEHS SNPs  
9 tagSNPs

## Where to Find Tagging Software

HaploBlockFinder	<a href="http://cgi.uc.edu/cgi-bin/kzhang/haploBlockFinder.cgi">http://cgi.uc.edu/cgi-bin/kzhang/haploBlockFinder.cgi</a>
Haploview	<a href="http://www.broad.mit.edu/mpg/haploview/">http://www.broad.mit.edu/mpg/haploview/</a>
LDSelect	<a href="http://droog.gs.washington.edu/ldSelect.html">http://droog.gs.washington.edu/ldSelect.html</a>
SNPtagger	<a href="http://www.well.ox.ac.uk/~xiayi/haplotype/index.html">http://www.well.ox.ac.uk/~xiayi/haplotype/index.html</a>
TagIT	<a href="http://popgen.biol.ucl.ac.uk/software.html">http://popgen.biol.ucl.ac.uk/software.html</a>
tagSNPs	<a href="http://www.rcf.usc.edu/~stram/tagSNPs.html">http://www.rcf.usc.edu/~stram/tagSNPs.html</a>

## Haplotypes, TagSNPs, and Caveats

- Haplotypes are inferred
- Block-like structure assumed for some software
- Different block definitions
- Block boundaries sensitive to marker density
- Genotype savings may not be great (recombination)

## Common Errors in Association Studies

Bell and Cardon (2001)

- Small sample size
- Subgroup analysis and multiple testing
- Random error
- Poorly matched control group
- Failure to attempt study replication e.g., Second case/control study  
Gene expression studies
- ✓ Failure to detect LD with adjacent loci
- Over interpreting results and positive publication bias
- Unwarranted 'candidate gene' declaration after identifying association in arbitrary genetic region

## Picking SNPs Application to Association Studies Summary

- Resources available for pair-wise LD and haplotypes
- Software for tagSNP selection available
- Be aware the limitations of the approach you choose
- Replication required by several journals