

The HapMap Project Tutorial



Presented by

Derek Aguiar

email

rap@cs.brown.edu

Goals

- After this tutorial you will hopefully know...
 - What HapMap is
 - What HapMap catalogs
 - How to use basic HapMap features
 - Find SNPs near a gene
 - Visualize allele frequencies
 - Download a set of SNPs in the area
 - Visualize LD
 - Visualize Haplotypes

What the HapMap Project?

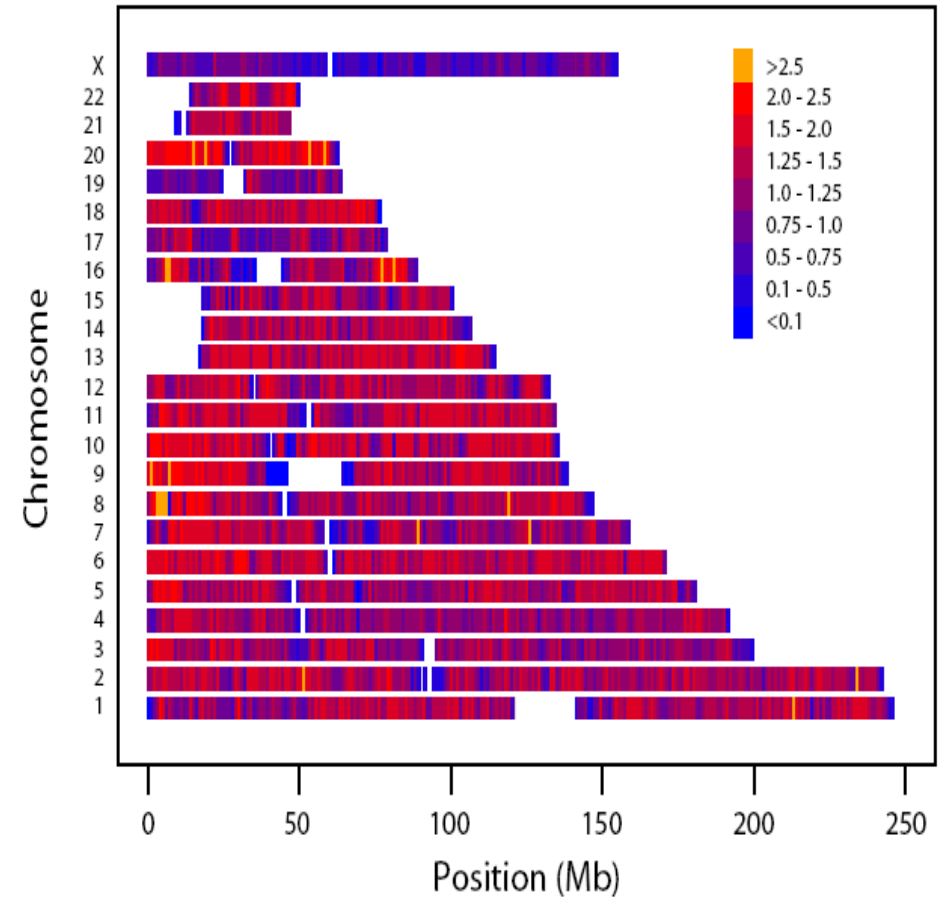
- International effort to identify and record genetic variation in *H. sapiens* only.
- Describes:
 - what they are – Type of SNP? alleles? etc...
 - where they are – genomic position
 - how they are distributed (intra and inter-population).

What the HapMap Project?

- The HapMap Project does *not* attempt to correlate genotype and phenotype.
- HapMap also includes some datasets on CNV
 - Copy Number Variation are larger chromosomal events in which large contiguous sections of DNA are deleted, inserted, reversed, or translocated
 - CNV accounts for most of the genetic variation between two individuals (estimated at about 0.4%)

Genetic Variation

- SNPs (pronounced “snips”) are by far the most common form of genetic variation



International HapMap Consortium
(2007). Nature 449:851-861

Genetic Variation Estimations

- There is about 1 SNP every 500-700bp (on average)
- 10 million SNPs in the Human genome
- 300k-600k tag SNPs that can contain the information of the 10 million (these SNPs are the most important for HapMap to catalog)

Let's Get To It!

- Open www.hapmap.org
- Select Phase III Data
 - Phase III is the most recent phase containing many more individuals from more diverse populations than Phase II and I. However, Phase III does not contain some of the analysis we will need for later on in the tutorial.

HLA-DRA

- Search for HLA-DRA
 - You can search for a gene by name, a chromosomal band, a phrase like “cell proliferation” (will return many results!)
 - HLA-DRA is the most strongly associated locus to MS according to many GWAS
- Re-center
- View SNP tables

SNP Details

- Left click on a SNP to view the SNP report
- Reports and Analysis – Download SNP data
 - You will get a report for the SNPs present in your window
 - Can choose “Configure” to use more options (including output data formats)

GWAS Hits

- Turn on GWAS Study annotations in region
- Re-center around GWAS study
- Mouse over for a summary
- Click for more info

Linkage Disequilibrium and Phased Haplotypes

- Go to the Phase I and II data on the main page
- Search again for HLA-DRA
- Turn on Phased Haplotype Display and LD Plot

Linkage Disequilibrium and Phased Haplotypes

- By default, triangle plot shows LD values using LOD scores.
 - Can be customized to use r^2 or D' which we will discuss soon
 - Can be customized to show more than one population
- For phased haplotypes, alleles colored yellow and blue
- Phasing was done in two steps:
 - Phase data using family info where possible
 - IMPUTE++ was used (HMM/MCMC)

HapMap

- Play with different settings and tools
- dbSNP and dbGaP are another two good resources
 - <http://www.ncbi.nlm.nih.gov/sites/gquery>
 - You might find references to these two sites in HapMap
- Homework assignment using HapMap is coming soon!