

## Haplovew [www.broad.mit.edu/mpg/haplovew/](http://www.broad.mit.edu/mpg/haplovew/)

### Software Demo

Summer Institute in Statistical Genetics 2014  
Module 10  
Topic 11

## Useful for...

- Visualizing LD and “haplotype blocks”
- Haplotype population frequency estimation
- Other possible uses:
  - Simple association tests
  - Tag SNP selection
  - Visualizing GWAS results from PLINK

## “Linkage” format data

- Pedigree Name: A unique alphanumeric identifier for this individual's family. Unrelated individuals should not share a pedigree name.
- Individual ID: An alphanumeric identifier for this individual. Should be unique within his family (see above).
- Father's ID: Identifier corresponding to father's individual ID or "0" if unknown father. Note that if a father ID is specified, the father must also appear in the file.
- Mother's ID: Identifier corresponding to mother's individual ID or "0" if unknown mother Note that if a mother ID is specified, the mother must also appear in the file.
- Sex: Individual's gender (1=MALE, 2=FEMALE).
- Affection status: Affection status to be used for association tests (0=UNKNOWN, 1=UNAFFECTED,2=AFFECTED).
- Marker genotypes

3

3	12	8	9	1	2	1	2	3	3	0	0	4	2
a	b	c	d	e	f							g	

- Individual 12 from pedigree 3
- Father is individual 8 and mother is individual 9
- Male
- Affected
- Genotype A C for first marker
- Genotype G G for second marker
- Missing genotype for third marker

4

## HaploView and LocusZoom

Rows: individuals of Ashkenazi Jewish ancestry  
Columns: after “intro” columns, alleles for some SNPs in a stretch of chromosome12 (in this file, 1=reference allele and 2 = alternative allele)

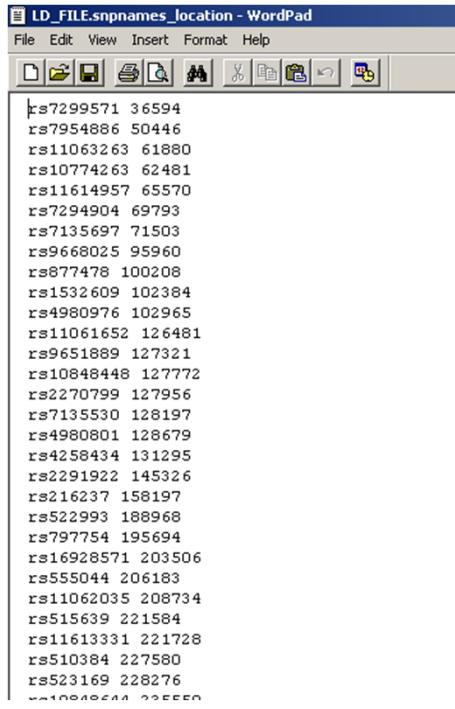
5

Each marker is represented by two columns (one for each allele, separated by a space) . Standard coding is either ACGT or 1-4 where: 1=A, 2=C, 3=G, T=4.

A “0” in any of the marker genotype position indicates missing data.

# HaploView and LocusZoom

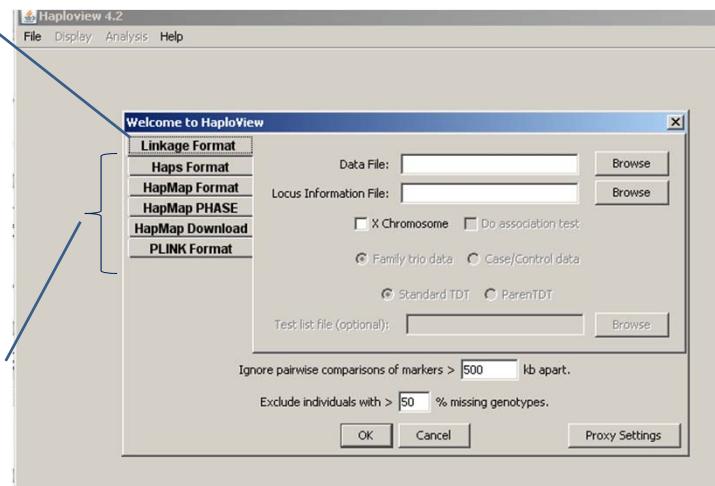
Names and locations  
of the SNPs  
represented in the  
data file



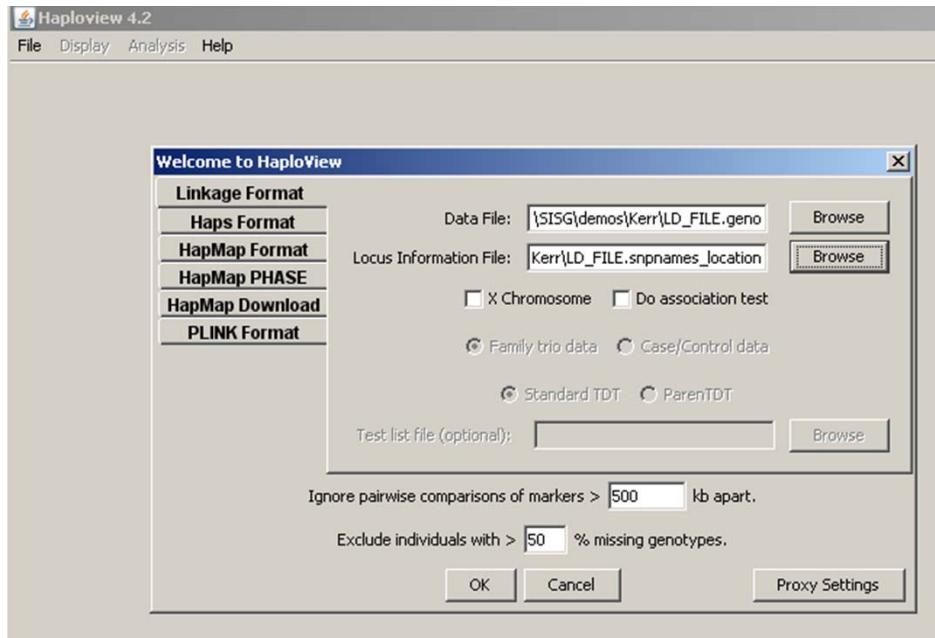
```
LD_FILE.snpnames_location - WordPad
File Edit View Insert Format Help
rs7299571 36594
rs7954886 50446
rs11063263 61880
rs10774263 62481
rs11614957 65570
rs7294904 69793
rs7135697 71503
rs9668025 95960
rs877478 100208
rs1532609 102384
rs4980976 102965
rs11061652 126481
rs9651889 127321
rs10848448 127772
rs2270799 127956
rs7135530 128197
rs4980801 128679
rs4258434 131295
rs2291922 145326
rs216237 158197
rs522993 188968
rs797754 195694
rs16928571 203506
rs555044 206183
rs11062035 208734
rs515639 221584
rs11613331 221728
rs510384 227580
rs523169 228276
--100000000 225550
```

Our data  
are in  
“Linkage  
Format”

Other  
formats can  
be used



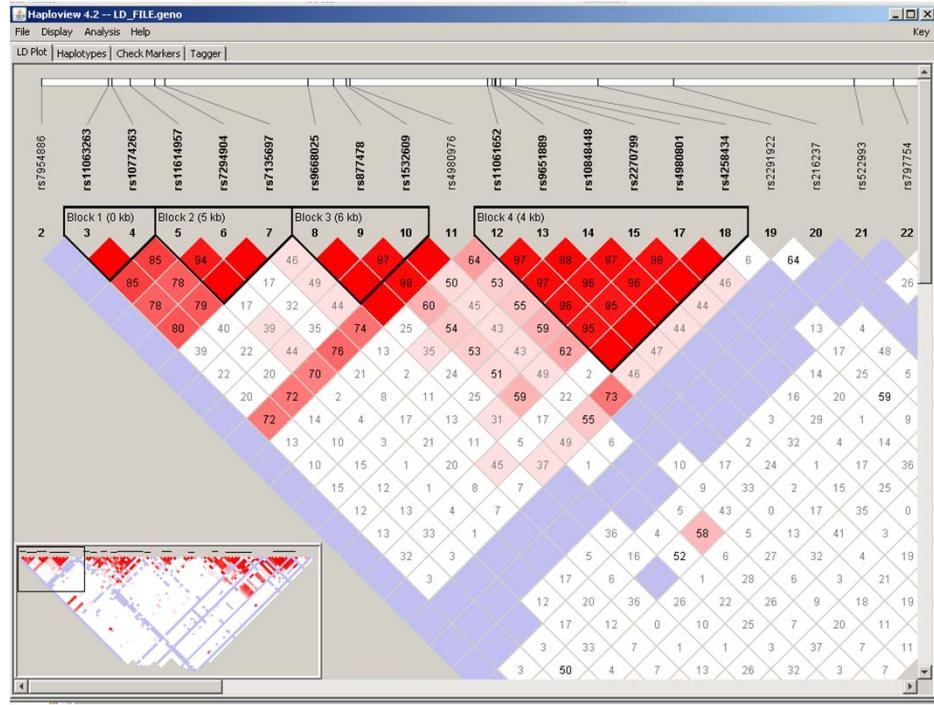
# HaploView and LocusZoom



9

#	Name	Position	ObsHET	PredHET	HWpval	%Geno	FamTrio	MendErr	MAF	Alleles	Rating
1	rs7299571	36594	0.915	0.496	3.394E-39	100.0	0	0	0.457	A:C	
2	rs7954886	50446	0.0050	0.0050	1.0	98.0	0	0	0.0030	C:A	✓
3	rs11063263	61880	0.533	0.498	0.4238	98.0	0	0	0.472	A:C	✓
4	rs10774263	62481	0.538	0.499	0.3549	100.0	0	0	0.475	C:A	✓
5	rs11614957	65570	0.487	0.489	1.0	100.0	0	0	0.425	C:A	✓
6	rs7294904	69793	0.482	0.477	1.0	100.0	0	0	0.392	A:C	✓
7	rs7135697	71503	0.503	0.499	1.0	100.0	0	0	0.477	C:A	✓
8	rs9668025	95960	0.332	0.339	0.6842	100.0	0	0	0.216	C:A	✓
9	rs877478	100208	0.446	0.446	1.0	98.0	0	0	0.336	C:A	✓
10	rs1532609	102384	0.472	0.453	0.6843	100.0	0	0	0.347	C:A	✓
11	rs4980976	102965	0.392	0.344	0.0772	100.0	0	0	0.221	C:A	✓
12	rs11061652	126481	0.387	0.466	0.0233	100.0	0	0	0.369	C:A	✓
13	rs9651889	127321	0.421	0.486	0.0778	98.0	0	0	0.415	C:A	✓
14	rs10849448	127772	0.421	0.473	0.1487	98.0	0	0	0.385	C:A	✓
15	rs2270799	127956	0.421	0.46	0.2773	98.0	0	0	0.359	C:A	✓
16	rs7135530	128197	0.0	0.0	1.0	98.0	0	0	0.0	C:C	
17	rs4980801	128679	0.395	0.464	0.0488	98.0	0	0	0.367	C:A	✓
18	rs4258434	131295	0.291	0.327	0.178	100.0	0	0	0.206	C:A	✓
19	rs2291922	145326	0.451	0.422	0.4475	98.0	0	0	0.303	C:A	✓
20	rs216237	158197	0.03	0.03	1.0	100.0	0	0	0.015	C:A	✓
21	rs522993	188968	0.01	0.01	1.0	100.0	0	0	0.0050	C:A	✓
22	rs797754	195694	0.01	0.01	1.0	98.0	0	0	0.0050	A:C	✓
23	rs16928571	203506	0.0	0.0	1.0	100.0	0	0	0.0	C:C	
24	rs555044	206183	0.39	0.396	0.9338	98.0	0	0	0.272	C:A	✓
25	rs11062035	208734	0.106	0.118	0.3374	100.0	0	0	0.063	A:C	✓
26	rs515639	221584	0.322	0.381	0.0429	100.0	0	0	0.256	C:A	✓
27	rs11613331	221728	0.492	0.477	0.7782	100.0	0	0	0.392	A:C	✓
28	rs510384	227580	0.205	0.216	0.6493	98.0	0	0	0.123	A:C	✓
29	rs523169	228276	0.342	0.386	0.1435	100.0	0	0	0.261	C:A	✓
30	rs11613334	228276	0.474	0.474	0.4747	100.0	0	0	0.346	A:C	✓

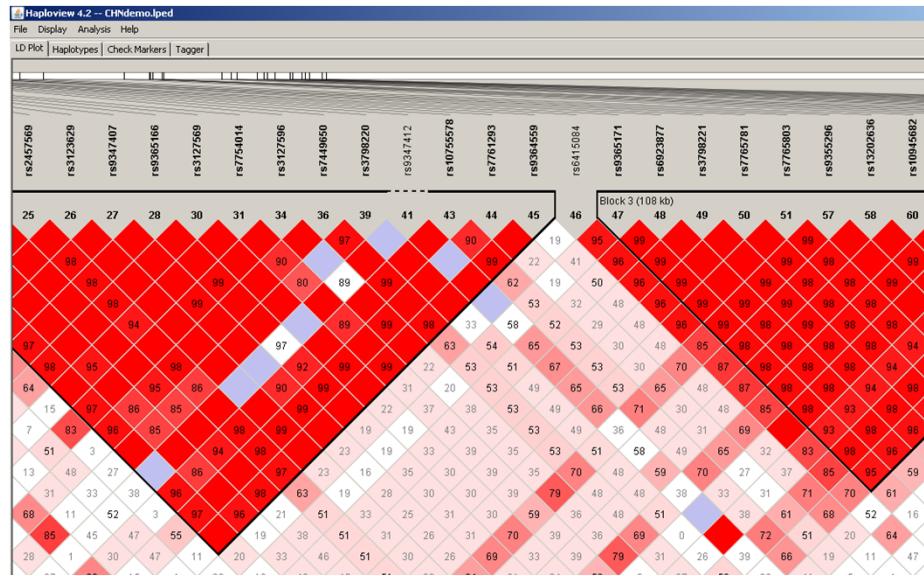
HW p-value cutoff: 0.0010  
 Min genotype %: 75  
 Max # mendel errors: 1  
 Minimum minor allele freq.: 0.0010



## Another Example

- Data on Chinese-Americans
- CHNdemo.lped
- CHNmarkerinfo.txt
- Lots of LD

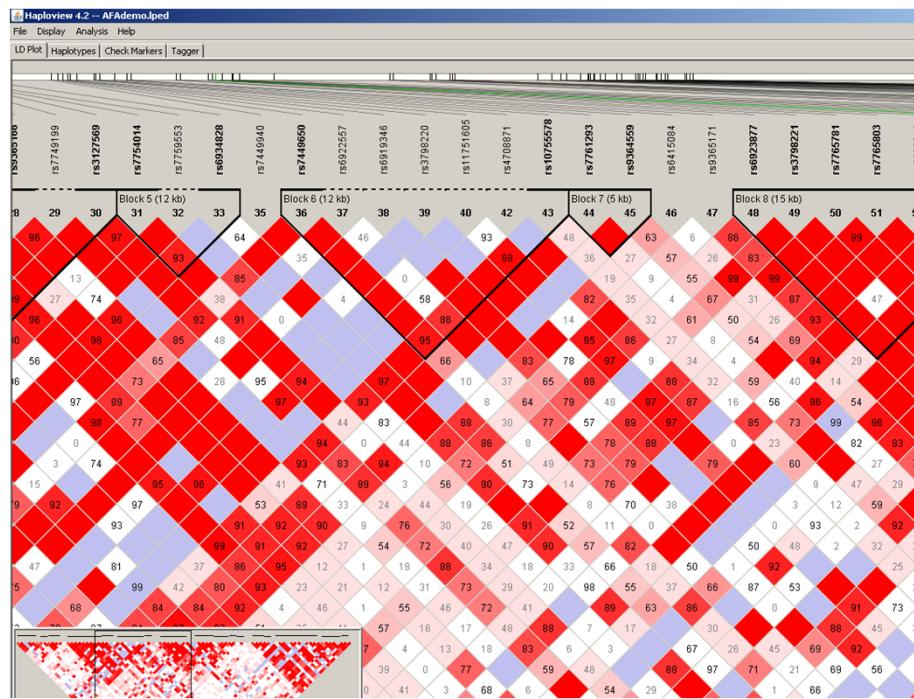
# HaploView and LocusZoom



13

- Contrast with the same region in a sample of African-American
  - AFAdemo.lped
  - AFAmarkerinfo.txt
- Lots less LD

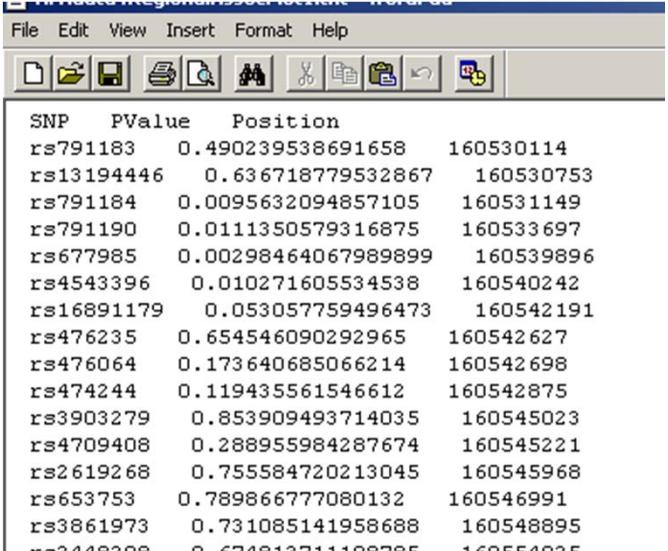
14



**Regional Association Plots**  
**Web based tools available at**  
[www.broadinstitute.org/mpg/snap/ldplot.php](http://www.broadinstitute.org/mpg/snap/ldplot.php)

Software Demo

## Input Data



SNP	PValue	Position
rs791183	0.490239538691658	160530114
rs13194446	0.636718779532867	160530753
rs791184	0.0095632094857105	160531149
rs791190	0.0111350579316875	160533697
rs677985	0.00298464067989899	160539896
rs4543396	0.010271605534538	160540242
rs16891179	0.053057759496473	160542191
rs476235	0.654546090292965	160542627
rs476064	0.173640685066214	160542698
rs474244	0.119435561546612	160542875
rs3903279	0.853909493714035	160545023
rs4709408	0.288955984287674	160545221
rs2619268	0.755584720213045	160545968
rs653753	0.789866777080132	160546991
rs3861973	0.731085141958688	160548895
rs2448298	0.674813711198785	160554025

17

## Required Columns

- SNP – must follow a standard nomenclature unless POSITION also specified
- PValue

18

## Optional Columns

- POSITION – the coordinate of the SNP on the chromosome
  - If not supplied, then Hapmap position will be used. If not in Hapmap, then SNP won't be plotted
  - ALL SNPs must be on the same chromosome within 1Mb

19

## Optional Columns

- RSquared – measure of LD between this SNP and target SNP
  - If not supplied, it is estimated from Hapmap data. If one or both SNPs in not in Hapmap, then LD assumed to be 0
- SnpType – “typed” or “imputed”

20

## Other Inputs

- The “target” SNP is the focus of the plot. The  $r^2$  values are computed from this SNP to all other SNPs. If not specified, then software defaults to use the SNP with the smallest pvalue.
- Population panel. Used, for example, to estimate  $r^2$  values if not input by user.

21

Input file example: AFADATA4RegionalAssocPlot1.txt

The screenshot shows the SNAP web application interface. At the top, there is a header with the title "SNAP" and "SNP Annotation and Proxy Search". To the right of the title, it says "Version 2.2" and has links for "FAQ | Documentation | Home". The "BROAD INSTITUTE" logo is also present. Below the header, there are four buttons: "Proxy Search", "Pairwise LD", "Plots", and "Map SNP IDs".

The main content area has a sub-header "Generate regional association plots or linkage disequilibrium plots." followed by a "Plot Type" section with a dropdown menu set to "Regional Association Plot".

In the "Association Data" section, there is a label "Select a file that contains snps and associated p-values ([file format](#)):". Below this are two input fields: "Association data:" with a "Browse..." button and an "Example" link, and "Target SNP:" with a note "(optional) Default is lowest p-value SNP".

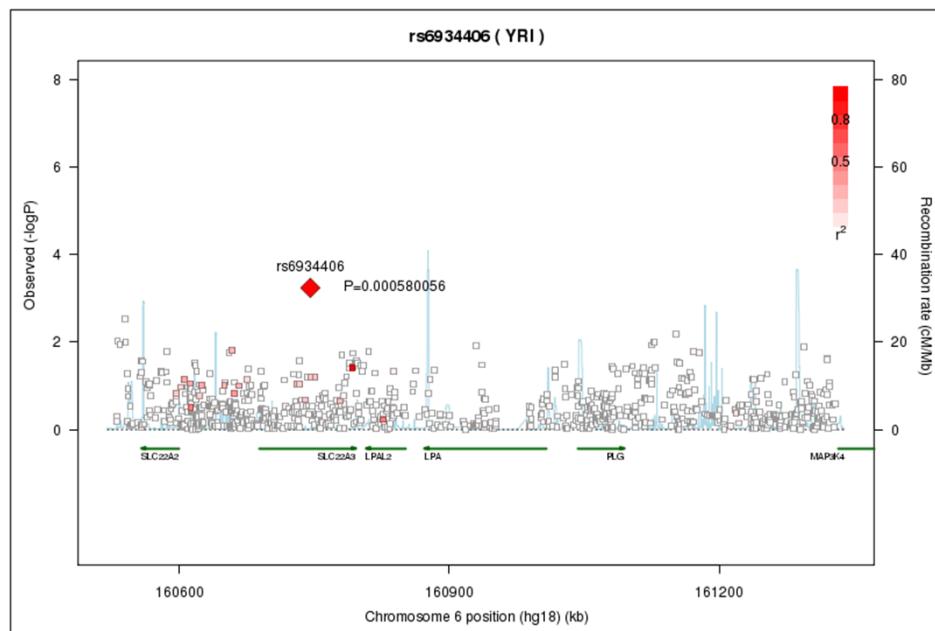
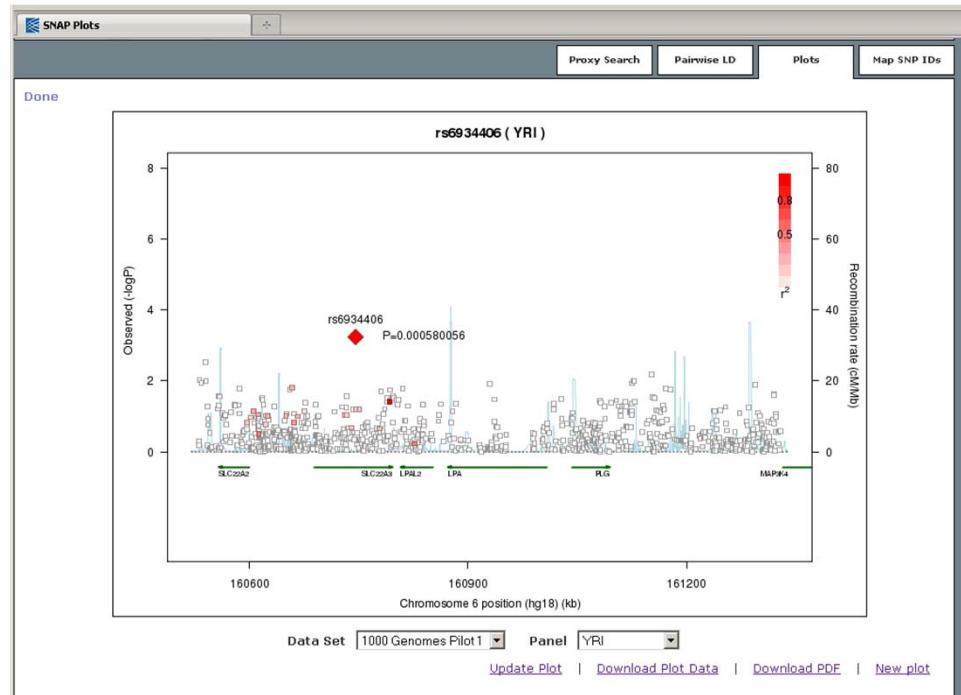
The "Plot Options" section contains the following settings:

- "SNP data set:" dropdown set to "1000 Genomes Pilot1"
- " $r^2$  threshold:" dropdown set to "0.8"
- "Population panel:" dropdown set to "CEU"
- "Distance limit:" dropdown set to "500"

At the bottom of the form, there are two buttons: "plot" and "reset".

At the very bottom of the page, there is a footer with the text "© 2008 Broad Institute".

# HaploView and LocusZoom



# Regional Association Plots

## LocusZoom

<http://csg.sph.umich.edu/locuszoom/>

### Software Demo

J. SENG M. WELCH J.

Search site

Site map

**UNIVERSITY OF MICHIGAN**  
CENTER FOR STATISTICAL GENETICS

**LocusZoom**

LocusZoom is a tool to plot regional association results from genome-wide association scans or candidate gene studies. This is Version 1.1

Report problems to [cristen@umich.edu](mailto:cristen@umich.edu)

We are pleased to announce that our paper on "LocusZoom" has been published. [ [ABSTRACT](#) | [PDF](#) ]

REFERENCE:  
Purcell SJ\*, Welch RP\*, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. (2010) LocusZoom: Regional visualization of genome-wide association scan results. Bioinformatics 2010 September 15; 26(18): 2336-2337.

Count of Successful Plots

Period	Count
This week	1025
Year 2012	32570
Jan	5205
Feb	5580
Mar	6494
Apr	6816
May	5959
Jun	2516
Year 2011	50775
Year 2010	22184

**Links**

- [Plot Using Your Data](#)
- [Plots Using Your Data and Your HitSpec File](#) Batch mode, results returned via Email
- [Plot Using Published GWAS Results](#)

Lipid (Total Cholesterol, HDL-C, LDL-C, Triglycerides)

- Teslovich et al. Nature 2010 (<http://www.sph.umich.edu/csg/abecasis/public/lipids2010>)

GIANT ([http://www.broadinstitute.org/collaboration/giant/index.php/Main\\_Page](http://www.broadinstitute.org/collaboration/giant/index.php/Main_Page))

- BMI: Speliotes et al. Nature Genetics 2010
- Height: Langston et al. Nature 2010

\* Whole genome association scan results from the Wellcome Trust Case Control Consortium 2010

# HaploView and LocusZoom

T  
AIR  
IED  
r

**Links**

- [Plot Using Your Data](#)
- [Plots Using Your Data and Your Hitspec File](#) Batch mode, results returned via Email
- [Plot Using Published GWAS Results](#)

Lipid (Total Cholesterol, HDL-C, LDL-C, Triglycerides)

- Teslovich et al. Nature 2010 (<http://www.sph.umich.edu/csg/abecasis/public/lipids2010/>)

GIANT ([http://www.broadinstitute.org/collaboration/giant/index.php/Main\\_Page](http://www.broadinstitute.org/collaboration/giant/index.php/Main_Page))

- BMI: Speliotes et al. Nature Genetics 2010
- Height: Lango Allen et al. Nature 2010
- Waist-hip ratio adjusted for BMI: Heid et al. Nature Genetics 2010

MAGIC (<http://www.magicinvestigators.org/>)

- 2hr Glucose Adjusted For BMI: Saxena R et al. Nature Genetics 2010
- Fasting Glucose, Fasting Insulin, HOMA-B, HOMA-IR: Dupuis J et al. Nature Genetics 2010

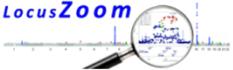
ICBP-GWAS ([http://www.georgehretlab.org/icbp\\_088023401234-9812599.html](http://www.georgehretlab.org/icbp_088023401234-9812599.html))

- ICBP-GWAS: Georg B. Ehret et al. Nature 2011

**Documentation**

- [FAQ](#) Frequently Asked Questions
- [Stand Alone Software](#) Run LocusZoom on your own machines
- [Change Log](#) What's changed

Input file example: AFADATA4RegionalAssocPlot1.csv

**LocusZoom**  **LocusZoom - Plot with Your Data**

---

**Plot Your Data**

Depending on the size of your data, runs can require 30-60 seconds to generate a plot

<b>Provide Details for Your Data</b>  <small>Required: Fill in Only ONE of These Three</small>	<b>Path to Your File</b> <input type="text" value="N:\Talks\2011\SiSG\demos\Kerr\AFADATA4F"/> <a href="#">Browse...</a> <small>File will be sent to server and used for plotting (Maximum 200MB) [Help]</small>	<small>Set for PLINK data or WigGWA data</small>
	<b>P-Value Column Name</b> <input type="text"/> <small>Default is P_value</small>	
	<b>Marker Column Name</b> <input type="text"/> <small>Default is MarkerName</small>	
	<b>Column Delimiter</b> <input type="text" value="Comma"/> <small>Default is tab</small>	
<b>Specify Region to Display</b>	<b>SNP</b> <input type="text" value="rs6934406"/> <small>SNP Reference Name</small> <input type="text" value="+/-"/> <input type="text" value="400"/> <small>Kb Flanking Size</small>	
	<b>Gene</b> <input type="text"/> <small>+/-</small> <input type="text" value="200"/> <small>Kb Flanking Size</small> <input type="text"/> <small>Optional Index SNP Default=lowest p-value</small>	
	<b>Region</b> <small>Chr:</small> <input type="text" value="None"/> <small>Mb</small> <b>through</b> <input type="text"/> <small>Mb</small> <input type="text"/> <small>Optional Index SNP Default=lowest p-value</small>	
<b>Custom Annotation</b>	<b>Column Name</b> <input type="text"/> <small>Name of annotation column</small>	

# HaploView and LocusZoom

Optional: This overrides Show Annotation below		Category Order	Order of annotation categories
<b>Optional Controls</b>			
Title On Plot	SISG Example Plot <b>Default is name of plot data</b>		
Genome Build/ LD Population	hg19/1000 Genomes Nov 2010 AFR		
Legend Location	auto		
Show RUG	HapMap <input type="checkbox"/> Use SNPs in Data <input checked="" type="checkbox"/> <input type="text" value="Plotted SNPs"/> Label for RUG of your SNPs		
Maximum Rows of Gene Names	<input type="text" value="10"/>		
Point Size Proportional to Sample Size	<input type="text" value="Weight"/> Specify NOWEIGHT as the column name to disable this		
Highlight Region of Interest	<input type="text"/> Mb through <input type="text"/> Mb Leave blank if no highlighting is needed		
Theme	Best for Viewing		
Show Annotation	<input type="checkbox"/> Framestop <input type="checkbox"/> Splice <input type="checkbox"/> NonSynonymous <input type="checkbox"/> Synonymous <input type="checkbox"/> UTR <input type="checkbox"/> * TFBScs <input checked="" type="checkbox"/> MCS44 Placental <input type="checkbox"/> Nothing		
Other Options	<input checked="" type="checkbox"/> Recombination Rate Overlay <input checked="" type="checkbox"/> Display Warning for Omitted Genes		

