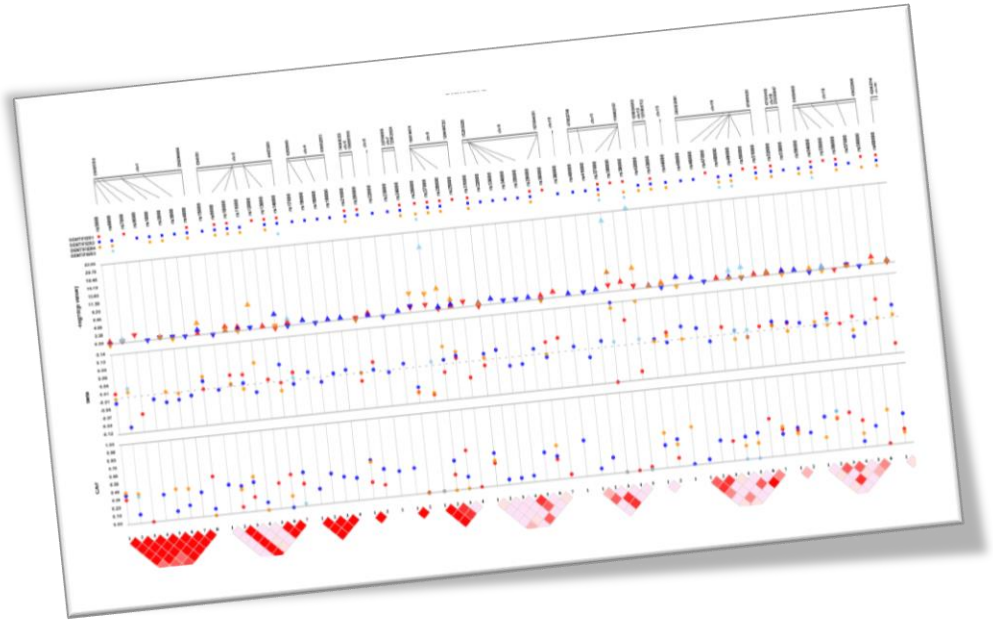
Synthesis-View



Visualization and interpretation of SNP association results for multi-cohort, multi-phenotype data and meta-analysis

Table of Contents

Introduction

Synthesis-View is a data visualization application. Through the use of stacked data-tracks, information on SNP genomic locations, presence of a SNP in a specific study or analysis, as well as related information such as genetic effect size and summary phenotype information, can be plotted according to user preference. Synthesis-View can be used either at the command line or using our web interface at http://visualization.ritchielab.psu.edu/synthesis_views/plot. Figure 1 shows a screen capture of the web interface.

Installation for Command Line Version

For Windows

1. Download and install Ruby. Use the one-click installer as the simplest option.

http://dl.bintray.com/oneclick/rubyinstaller/rubyinstaller-1.9.3-p545.exe?direct

Check 'Add Ruby executables to your PATH'
Check 'Associate .rb and .rbw files with this Ruby installation' if you want to be able to double click on a script to start it or just type it at the shell prompt.

2. Download and install Ruby development kit. Download it, run it to extract it somewhere (permanent). Then cd to it, run 'ruby dk.rb init' and 'ruby dk.rb install' (without quotes) to bind it to ruby installations in your path.

https://github.com/downloads/oneclick/rubyinstaller/DevKit-tdm-32-4.5.2-20111229-1559-sfx.exe

3. Download and Install ImageMagick. The latest version is not compatible with RMagick on Windows but version 6.8.7-8 is and can be downloaded from the site below. Check 'Install development headers and libraries for C and C++' during installation and install it into a directory without spaces (not in 'Program files').

http://ftp.sunet.se/pub/multimedia/graphics/ImageMagick/binaries/ImageMagick-6.8.7-8-Q16-x86-dll.exe

4. Open a command prompt and install RMagick. Replace the <Installation directory> in the command below with the location of the ImageMagick installation.

gem install rmagick --platform=ruby -- --with-opt-lib=C:\<Installation directory>\lib --with-opt-include=C:\<Installation directory>\include

For example, if it is installed in C:\ImageMagick-6.8.7-Q16 the command would be:

gem install rmagick --platform=ruby -- --with-opt-lib=C:\ImageMagick-6.8.7-Q16\lib --with-opt-include=C:\ImageMagick-6.8.7-Q16\include

5. Download pheno_gram.rb from

http://ritchielab.psu.edu/software/phenogram-downloads

For Linux/Mac

1. Download and install Ruby if not already installed. See the following site for instructions on Linux and OS X.

http://www.ruby-lang.org/en/downloads/

2. Download and install ImageMagick using the method appropriate for your operating system.

http://www.imagemagick.org/script/binary-releases.php

3. Install the RMagick gem.

    gem install rmagick

3. Download pheno_gram.rb from

http://ritchielab.psu.edu/software/phenogram-downloads

Synthesis-View Command Line Options

For those using the command line version of Synthesis-View rather than the web interface, here are the command-line arguments for controlling the behavior of the Synthesis-View script. Only the –e option is required, specifying the main input for the script to generate the standard Synthesis-View plot.

-e [synth_file]	Synthesis-View format file for input
-G [log_file_name]	Options log file
-s [phenotypic_summary_file]	Group summary file for input
-g [gene_file]	Gene summary file
-l [ldfile]	Linkage disequilibrium file
-b [abbrev_file]	Abbreviation definition file
-i [SNP info file]	Tags each SNP with information (such as source)
-t [title_str]	Main title for plot (enclose in quotes)
-M, --more-locs	Label SNP locations at additional spots along chromosome
-k, --group-key-end	Place group color key at bottom of plot
-T	Larger plot font
-S, --cleaner-axis	Axes on plots "cleaner". Default uses min and max values as axis boundaries
-J, --jitter	Offsets results from different groups horizontally
-c, --no-triangle	P values will be drawn as circles instead of triangles that show direction of effect
-U [highlighted]	Group passed as argument will be drawn as diamonds
-A, --pval	Plot p value track
-B, --beta	Plot effect size track
-n [eff_name]	Effect size name (beta or es)
-x, --ancestry-track	Plot ancestry track
-Z, --rank-track	Plot rank track
-u [columns]	Columns to plot that are not standard
-p [pthresh]	p value threshold
-P [pheno_title]	Phenotype title for phenotype summary plot (enclose in quotes)
-m, --pmin [pmin_val]	Maximum y-axis setting for p value track
-N, --plot-study-total	Plot number of studies in meta analyis track
-d	Draws d-prime result plot
-r	Draws r-squared result plot

-a, --rotate	Rotate final plot
-F, --forest	Draw Forest plot
-D, --forest-legend	Draw legend on Forest plot
-I, --interaction	Draw interaction plot
-W, --power-plot	Draw power track
-L, --large-or	Significant odds ratios larger in size on plot
-R, --or-zero	Set minimum value on Forest plot x-axis to zero
-C, --case-control	Plot case and control totals on separate plots
-O, --circle-case-con	Plots cases and controls as closed and open circles on same track
-H, --caf-case-control	Plots case and control caf on separate plots
-K, --caf-circle-case-control	Plots case and control caf as closed and open circles on same track
-z	Draws high resolution figure (300dpi)
-Y	Create grayscale image
-w	Produce html file with links to dbSNP
-f [image_type]	Image format for output (png default)
-o [output_name]	Optional output name for the image
-h, --help	Show this usage statement
-v, --version	Show version

Creating Sample Synthesis-View Plot at the Command Line

Type at the command prompt:

ruby synthesis_view.rb –e sample_input.txt

To specify a title for the plot:

ruby synthesis_view.rb –e sample_input.txt –t "My sample plot" –o sample

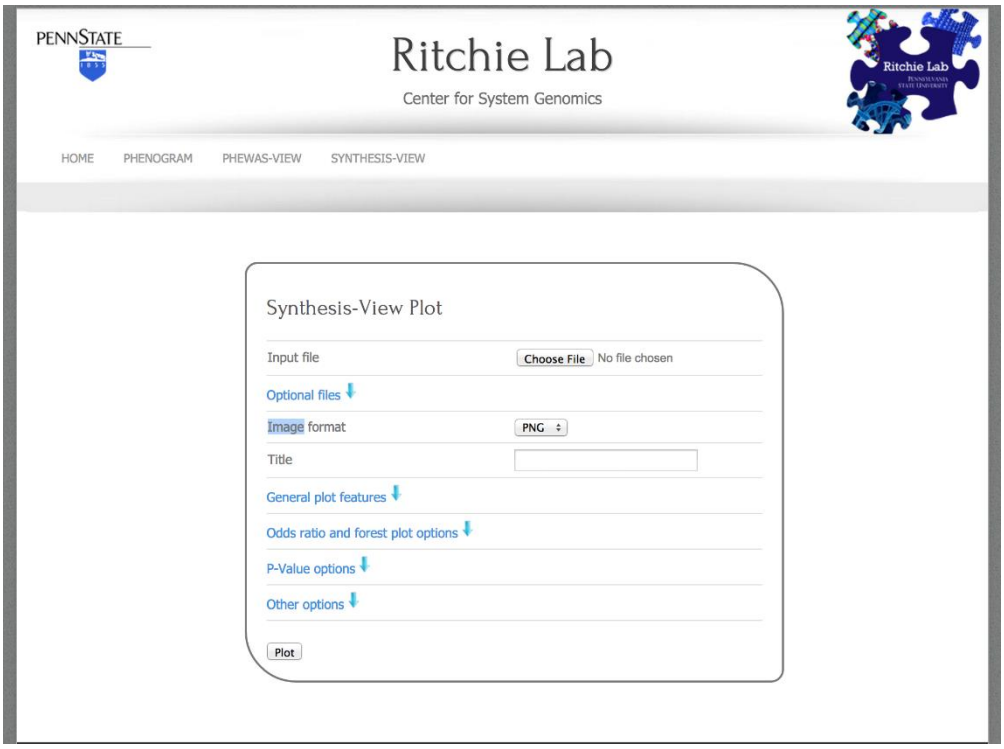


Figure 1: Web interface available at http://visualization.ritchielab.psu.edu/synthesis_views/plot

Synthesis-View Input File Format

Standard Synthesis View Plot

All Synthesis-View input files are tab-delimited text files, and the first row must be a header column. To create a standard Synthesis-View plot (Figure 2) the input file columns take the format *group:type* where *group* is a set of results (such as multiple phenotypes, multiple studies, or multiple ancestry groups) and *type* specifies the type of result. Accepted types are *pval* (p value), *es* (effect size), *N* (sample size), *caf* (coded allele frequency), *cases* (total cases), *controls* (total controls), *cafcases* (coded allele frequency for cases), *cafcontrols* (coded allele frequency for controls), *study* (number of studies included) and rank(ranking of the groups for that SNP). A simple example is shown in the table below. The SNP column with SNP rs IDs is required in the file unless an interaction plot is specified. For an interaction plot, the input file must have SNP1 and SNP2 columns. The chromosome number and base pair location must also be supplied by the user. An example Synthesis-View file is available at http://visualization.ritchielab.psu.edu/synthesis_views/document. A similar file was used to generate Figure 2. Figure 3 shows an example plot from Pendergrass et al. 2010 [1], where multiple phenotypes are plotted as the "groups". Note, in Figure 3, p-values are plotted as triangles (the plotting of p-values as triangles is optional). The base of the triangle corresponds to the p-value, and the direction of the triangle corresponds to the direction of the effect.

Figure 4 shows an example of screen output from Synthesis-View, in this case a forest plot (see the manual section on forest plots in Synthesis-View). When the plot appears in the web interface after plotting data within Synthesis-View, SNP rs IDs within this plot can be selected with the mouse, and entrez SNP information for the SNP appears in a new window. *Note*: this feature only works in cases when the SNP rs ID numbers are matched with current publically available SNP rs ID information.

SNP	Chromosome	Location	Group1:pval	Group1:es	Group2:pval	Group2:es
rs1	1	484938	0.1053	-0.0293	0.252	0.2293
rs2	1	582839	0.3501	0.01451	0.822	-0.0939
rs3	1	786111	0.421	0.0198	0.121	0.3210
rs4	2	144183	0.888	0.0129	0.402	-0.4554
rs5	2	203394	0.082	-0.9951	0.782	0.00832

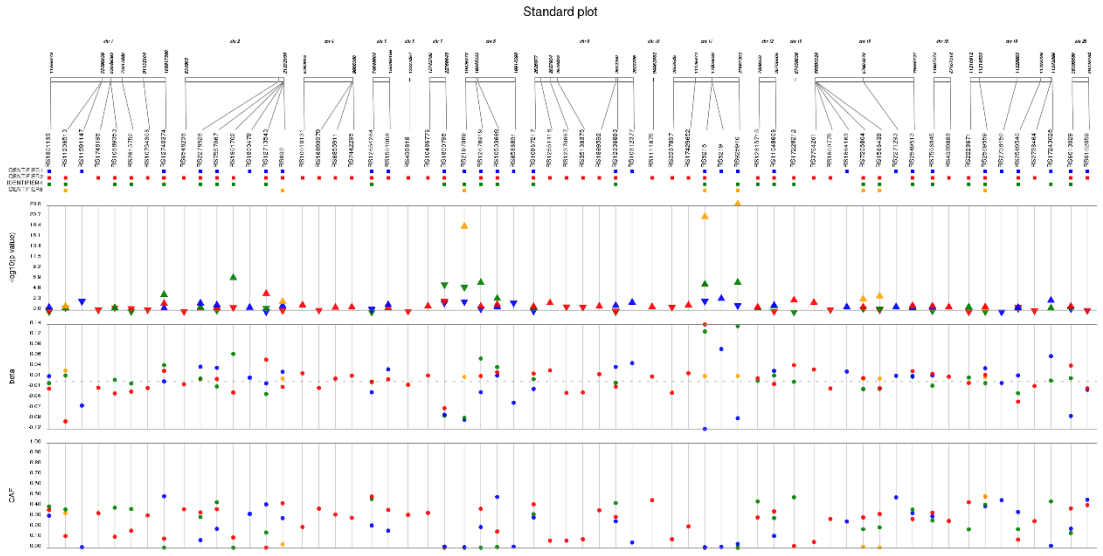


Figure 2: Standard Synthesis–View plot using the example file available at http://visualization.ritchielab.psu.edu/synthesis_views/document

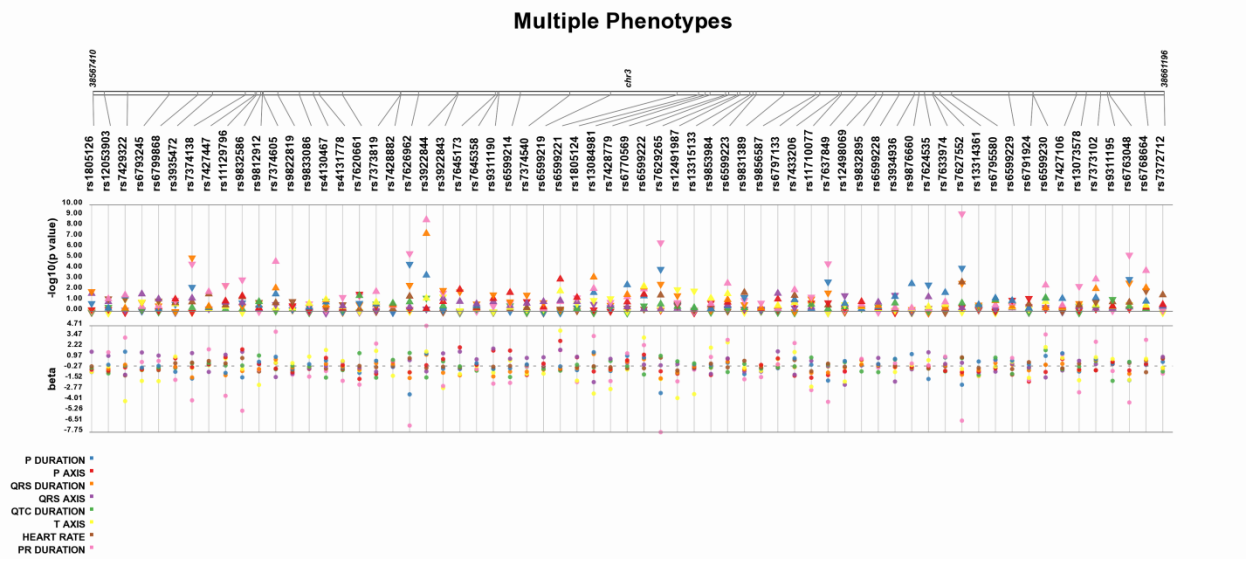


Figure 3: Synthesis–View plot showing multiple phenotypes (Pendergrass et al. 2010 [1])

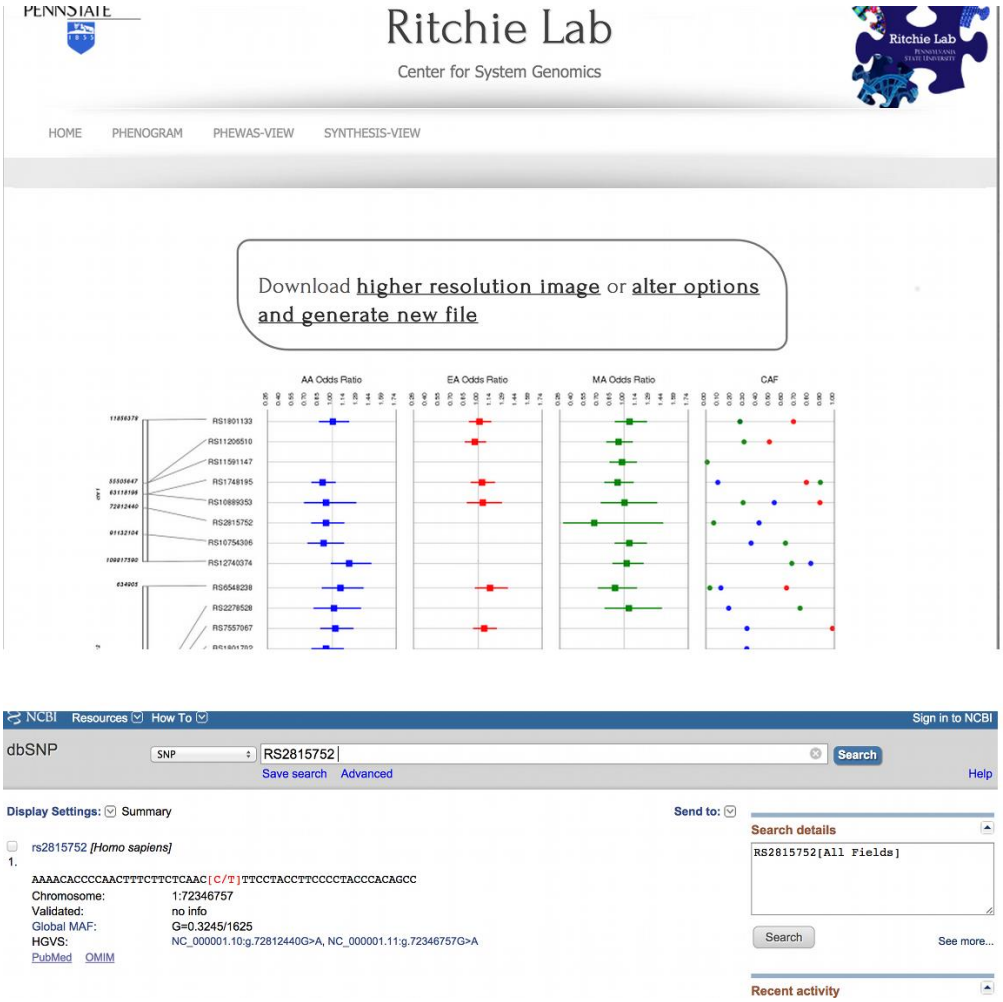


Figure 4: Synthesis–View web interface interactive output plot

Subgroups and Main Groups

Subgroups can be specified for the main groups. In this case the format is *group:subgroup:type* and each subgroup will have its own plot drawn on a separate track (Figure 5) so that differences/similarities between subgroups can be visualized:

SNP	Chromosome	Location	Group1:sub1:pval	Group1:sub1:es	Group1:sub2:pval	Group1:sub2:es
rs1	1	484938	0.1053	-0.0293	0.252	0.2293
rs2	1	582839	0.3501	0.01451	0.822	-0.0939
rs3	1	786111	0.421	0.0198	0.121	0.3210
rs4	2	144183	0.888	0.0129	0.402	-0.4554
rs5	2	203394	0.082	-0.9951	0.782	0.00832

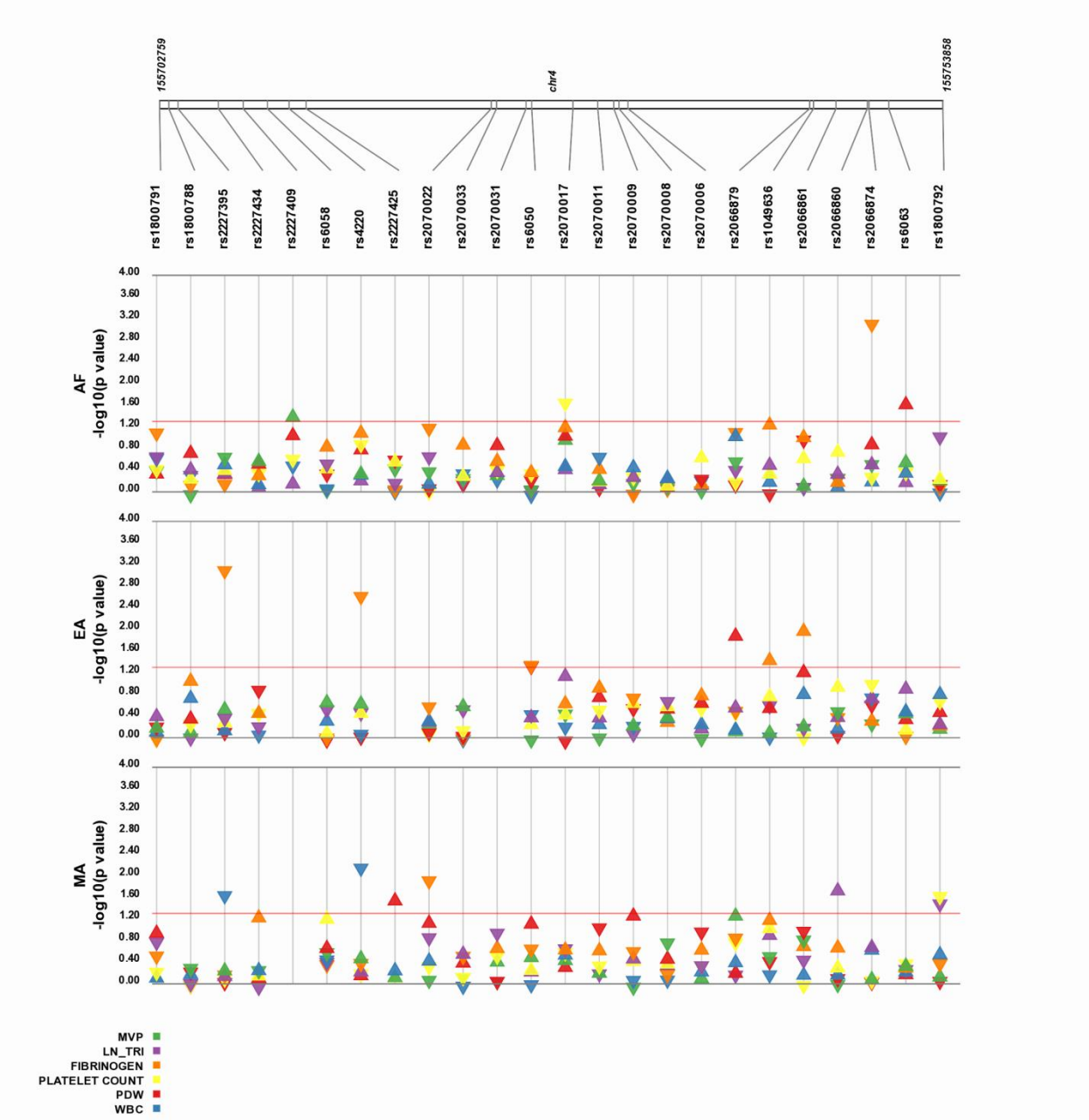


Figure 5: Synthesis–View plot using multiple groups (phenotypes) and multiple subgroups (ancestry) from Pendergrass et al. 2010 [1].

Optional Phenotype or Sample Size Summary, Color Selection File

Synthesis-View accepts an optional file listing phenotypic and/or sample size summary information for the groups in the Synthesis-View input file. This is also the way to optionally specify colors for the various groups. The required column is *Group* (information on what group the sample size and/or phenotype measurements correspond to). For plotting sample size for each group, include the optional column *Samples.* For plotting phenotypic summary information, include the columns *Avg Pheno* (the mean of the phenotype for that respective group), and *Std Dev* (standard deviation of the respective phenotype). Alternately, to create a box plot of the summary information, include the following columns: *Median* (median value for each group), *25%* (25th percentile), *75%* (75th percentile), *Min* (minimum value in the group), and *Max* (maximum value in the group). To specify colors for the groups use the column header *Color*. The colors can be

specified in one of several formats recognized by RMagick (http://www.imagemagick.org/RMagick/doc/imusage.html#color_names).

Average phenotype column example input:

Group	Samples	Avg Pheno	Std Dev
Group1	500	4.3	0.5
Group2	370	3.8	0.28
Group3	580	4.7	0.44

Color specification example input:

Group	Color
Group1	Blue
Group2	Black
Group3	Red

Box plot example input:

Group	Median	25%	75%	Min	Max
Group1	8.8	6.0	10.1	4.4.	11.8
Group2	7.2	5.8	9.0	4.2	10.7
Group3	7.8	5.5	10.2	3.5	12.2

Gene Summary File

Synthesis-View accepts an optional file detailing gene and location information. This gene information will be plotted above the genomic location bar for the SNPs, at the top of the plot:

Name	Chr	Start	End
ACOT11	1	55495789	55576000
PPAP2B	1	56960433	57045257
ATG4C	1	83249803	83330049

Linkage Disequilibrium File

Synthesis-View accepts an optional file input that has LD information. This file matches the LD file output of Haploview [2]. A description of this file format can be found at http://www.broadinstitute.org/science/programs/medical-and-population-genetics/haploview/output-file-formats. Synthesis-View provides a sample LD file. Figure 6 shows an example of Synthesis-View output that includes an LD plot.

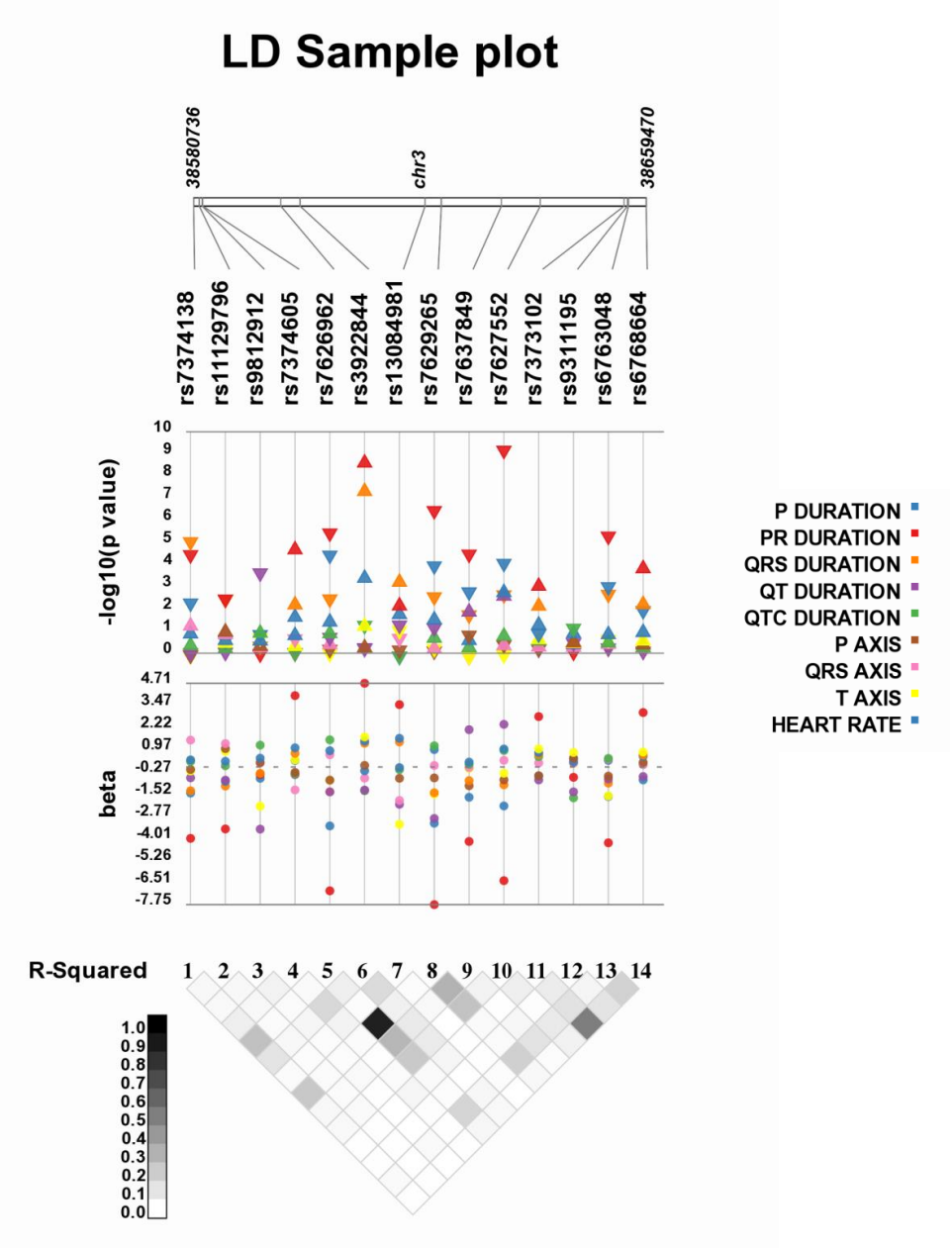


Figure 6: Synthesis–View plot utilizing the LD plot option, from Pendergrass et al. 2010 [1].

Options Log

When a Synthesis-View plot is first generated, there is log file that is produced. This file can be re-used in Synthesis-View to change the Synthesis-View settings to those used when the original log file was generated. This can aid in the replication of plots.

Abbreviation Definitions

This is an optional file that can be supplied that contains information on any abbreviations used in the labeling of groups. In this way the legend indicating what color each group corresponds to will contain more information about the abbreviations used in the plot, such as "AA" which could be an abbreviation for "African-American". The table below shows the format of an example input file.

EA	European Americans
AA	African Americans
MA	Mexican Americans
AI	Asian and Pacific Islanders

SNP Sources

This optional file is used in conjunction with the interaction plot option to create a SNP interaction plot. The file needs two columns, *SNP* and *Source*.

SNP	Source
rs1000	Add
rs1000	Rec
rs1001	Add

Plot Options

Title

The title for the plot, centered at the top of the plot.

Image format

Choose the format of the output plot, the options are PNG, JPEG, and TIFF.

General Plot Features

Rotate

Final plot is rotated 90 degrees using this option. SNPS appear on the left side of the plot.

Larger Font

This setting will create a larger font for text on the plot.

Grayscale

This setting will set the colors on the plot to shades of gray.

Axis Scaling

Axis scaling will either automatically be chosen (maximum), or axis scaling will be plotted with tick marks that are a multiple of five or ten.

Additional SNP Locations

Base pair location information for each SNP will be plotted. Otherwise the plot will contain base pair information only at either end of the graphical representation of each chromosome.

Offset Overlapping Points

When there is considerable overlapping of plotted points, a horizontal offset of overlapping points can be useful (also called jitter). Selecting this box will include a horizontal offset of overlapping points.

Phenotype Summary Plot Name

To apply a title to a phenotypic summary plot, indicate the title of the phenotypic summary plot here.

Forest/Odds-Ratio Plots

Synthesis-View will also plot forest plots when odds ratio column(s) and related information are provided. In this case, the file must include *or* (odds-ratio), *lower_ci* (lower confidence-interval) and *upper_ci* (upper confidence-interval) columns (see example input format in the below table). Figure 7 shows an example Synthesis-View forest plot.

SNP	Chromosome	Location	Group1:or	Group1:upper_ci	Group1:lower_ci	Group1:caf
rs1	1	484938	1.02	0.782	1.33	0.45
rs2	1	582839	1.035	0.861	1.244	0.22
rs3	1	786111	0.928	0.76	1.132	0.30
rs4	2	144183	1.128	0.837	1.52	0.15
rs5	2	203394	0.968	0.821	1.141	0.20

Produce Forest Plot

When used with the option –a on the command-line, or in the web interface through the selection of "Produce forest plo", a forest plot will be created. For additional forest plot examples see Pendergrass et al. 2010 [3].

Minimum Forest Plot X-axis at zero

Choose this setting to set the minimum forest plot x-axis to zero. Otherwise the x-axis will be automatically scaled for the range of the data.

Plot Case/Control Totals

To plot case/control numbers, choose "combined plot" or "split plot". If "combined plot" is chosen, cases and controls will be plotted in the same track. If "split plot" is chosen, the cases and controls will be plotted in separate tracks.

Plot Case/Control CAF

To plot case/control coded allele frequencies, choose "combined plot" or "split plot". If "combined plot" is chosen, the coded allele frequency for cases and controls will be plotted in the same track. If "split plot" is chosen, the coded allele frequency for cases and controls will be plotted in separate tracks.

Plot Significant Odds/Ratio Larger

Choose this setting to plot significant Odds/Ratio results in larger size.

Draw Legend

Choosing this setting will draw a legend indicating what colors correspond to which group.

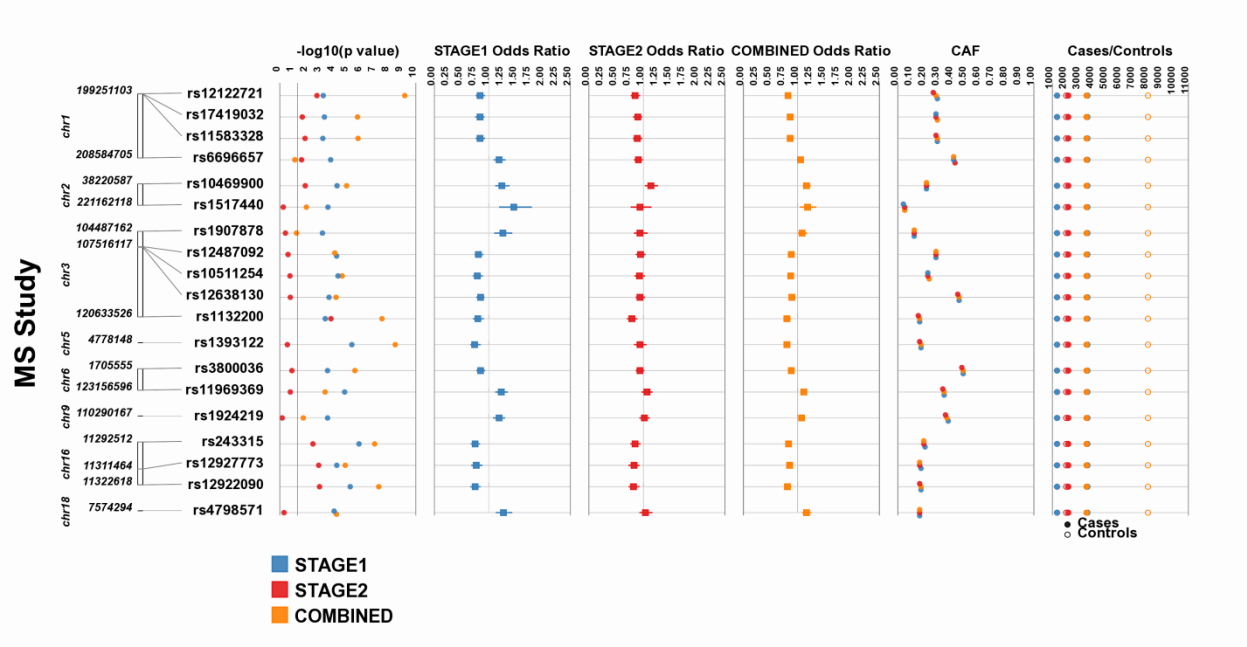


Figure 7: Synthesis–View forest plot. Figure from Pendergrass et al. 2010 [1].

P-value options

Include P-Value Plot

The plotting of the p-value plot track is optional.

Plot P-values as Circles

P-values can be plotted as circles instead of triangles. If plotted as triangle, the base of the triangle indicates the significance of the p-value, direction of effect is indicated by the direction of the triangle.

Draw Line at this P-value

A red line will be drawn on the plot at the p-value of interest indicated here. Figure 2 shows an example with a red line plotted at a p-value of interest.

Maximum Y-axis Setting for P-value Track

If there is a wide range of p-values, a p-value cutoff can be chosen whereby any p-value greater than the cutoff is plotted at the cutoff in larger size.

Highlighted Group

A diamond will be plotted for one group of interest if that group is indicated in the "Highlighted Group" box. This can aid in indicating more visual focus for one set of results compared to others.

Other Options

Include direction of effect track

Choose this setting to include the direction of effect track.

Effect Label

If the direction of effect track is included, the label can be “beta” or “effect size”.

Linkage Disequilibrium D-prime Plot

Choose this setting to include a linkage disequilibrium D’ plot. Note: an additional D’ LD plot must be supplied

Linkage Disequilibrium r-prime plot

Choose this setting to include a r^2plot. Note, an additional r^2 LD plot must be supplied.

SNP position track

Select this setting to include the SNP position track above the SNP names on the plot.

Power track

Choose this setting to include a power track on the plot.

Interaction plot

Choose this setting to create an interaction plot that includes the information from the SNP source file.

1. Pendergrass SA, Dudek SM, Crawford DC, Ritchie MD: **Synthesis-View: visualization and interpretation of SNP association results for multi-cohort, multi-phenotype data and meta-analysis.** *BioData Min* 2010, **3:**10.
2. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.** *Bioinformatics* 2005, **21:**263-265.
3. Pendergrass S, Dudek SM, Roden DM, Crawford DC, Ritchie MD: **Visual integration of results from a large DNA biobank (biovu) using synthesis-view.** *Pac Symp Biocomput* 2011**:**265-275.