

# EnigmaVis: Online Interactive Visualization of Genome-Wide Association Studies of the Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) Consortium

Nic M. Novak,<sup>1</sup> Jason L. Stein,<sup>1</sup> Sarah E. Medland,<sup>2</sup> Derrek P. Hibar,<sup>1</sup> Paul M. Thompson,<sup>1</sup> and Arthur W. Toga<sup>1</sup>

<sup>1</sup>Laboratory of Neuro Imaging, David Geffen School of Medicine, UCLA, Los Angeles, USA

<sup>2</sup>Quantitative Genetics, Queensland Institute of Medical Research, Brisbane, Australia

In an attempt to increase power to detect genetic associations with brain phenotypes derived from human neuroimaging data, we recently conducted a large-scale, genome-wide association meta-analysis of hippocampal, brain, and intracranial volume through the Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) consortium. Here, we present a freely available online interactive tool, EnigmaVis, which makes it easy to visualize the association results generated by the consortium alongside allele frequency, genes, and functional annotations. EnigmaVis runs natively within the web browser, and generates plots that show the level of association between brain phenotypes at user-specified genomic positions. Uniquely, EnigmaVis is dynamic; users can interact with elements on the plot in real time. This software will be useful when exploring the effect on brain structure of particular genetic variants influencing neuropsychiatric illness and cognitive function. Future projects of the consortium and updates to EnigmaVis will also be displayed on the site. EnigmaVis is freely available online at <http://enigma.loni.ucla.edu/enigma-vis/>

■ **Keywords:** GWAS, visualization, neuroimaging, genetics, SNP

The human brain has many heritable features (Kremen et al., 2010; Peper et al., 2007; Thompson et al., 2001). However, the genetic variants underlying these high heritability estimates are, for the most part, unknown. Genome-wide association studies (GWAS) are one way to identify common variants influencing heritable traits in large-scale population studies. GWAS have been used to identify associations between single-nucleotide polymorphisms (SNPs) and a host of different traits implicated in numerous diseases (Cichon et al., 2009; McCarthy et al., 2008). Meta-analysis has proven to be critical to our understanding of the true effects that specific genetic variants have on these traits, as most common variants have small effects. In general, individual studies — which typically assess a few hundred to a thousand individuals — are underpowered to reliably detect associations. Recently, we initiated the Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) consortium (Stein et al., 2012). The primary goal of the ENIGMA consortium is to expedite meta-analysis of large datasets and create a forum for collaboration in the field of

imaging genetics. The effort is modeled on other highly successful consortia in psychiatric genetics, which have discovered genetic loci associated with bipolar illness, schizophrenia, and attention deficit hyperactivity disorder (ADHD) (Neale et al., 2010; Ripke et al., 2011; Sklar et al., 2011), offering new leads for research at the molecular and systems levels. In several ongoing projects, the ENIGMA consortium has been analyzing the genetic influences on neuroimaging traits with data from over 20 research groups and tens of thousands of subjects. This presents a useful resource for the imaging, neuropsychiatric, and cognitive genetics communities to discover genes that influence the brain. It also facilitates the confirmation, replication, and understanding of effects of promising genetic variations and pathways.

RECEIVED 1 February 2012; ACCEPTED 28 March 2012.

ADDRESS FOR CORRESPONDENCE: Arthur W. Toga, 635 Charles Young Drive South, Suite 225, Los Angeles, CA 90095-7334, USA. E-mail: [toga@loni.ucla.edu](mailto:toga@loni.ucla.edu)

A method for user-friendly visualization and navigation of the genetic regions containing important associations is essential for demonstrating significant findings and for distributing these results. For initiatives such as ENIGMA, analyzing very large amounts of GWAS data, it is critical that researchers are able to quickly and efficiently examine the strength of the association at any desired genetic locus. Specifically, researchers may want to look up the evidence of genetic association between a gene they are interested in, and various brain measures examined by ENIGMA. As such, a visualization utility may prove to be one of the most useful methods to facilitate interpretation of ENIGMA results. Conventional publication formats for GWAS use tabular data and static graphs, which make it difficult for those not closely involved in the study to access and browse the results. Top hits are usually summarized in tables, which lose a great deal of the data that is available at other loci across the genome. Conventional formats additionally rely on non-visual methods for data presentation, often consisting of lists of SNP numbers and probabilities of association that are very hard to digest. GWAS results comprise large quantities of information, and these publication formats represent a limitation of the journal manuscript in its ability to present GWAS results. Online environments have the ability to serve rich and dynamic content, and are highly suited for presentation of these kinds of results. Existing online data visualization tools — SzGene, AlzGene, LocusZoom, Ricopili — for genetic association studies (Allen et al., 2008; Bertram et al., 2007; Pruim et al., 2010; Ripke & Thomas, 2011) are available, but have some limitations. All of these existing tools are static in nature; once a plot has been generated, it cannot be interacted with or further explored. Additionally, the SzGene and AlzGene databases are based primarily on candidate-gene studies. Also, these tools are not available for imaging genetics data. To this end, we developed EnigmaVis, an online interactive tool for visualizing unbiased genome-wide association results from ENIGMA.

## Implementation

### Features and Functionality

The main EnigmaVis webpage (<http://enigma.loni.ucla.edu/enigma-vis/>) features a query box, phenotype selection drop-down box, and file uploader. These fields allow a user to query the database of associations with neuroimaging phenotypes generated by the ENIGMA consortium and generate custom, interactive plots for a desired region of the genome. Presently, EnigmaVis allows users to select GWAS results representing eight phenotypes from Stein et al. (2012) for visualization; as the project grows, more association results to neuroimaging, neuropsychiatric disease, and neurocognitive phenotypes will be uploaded to the database and made available for use in EnigmaVis. Users can search by SNP, genomic position, or gene. Optionally, the

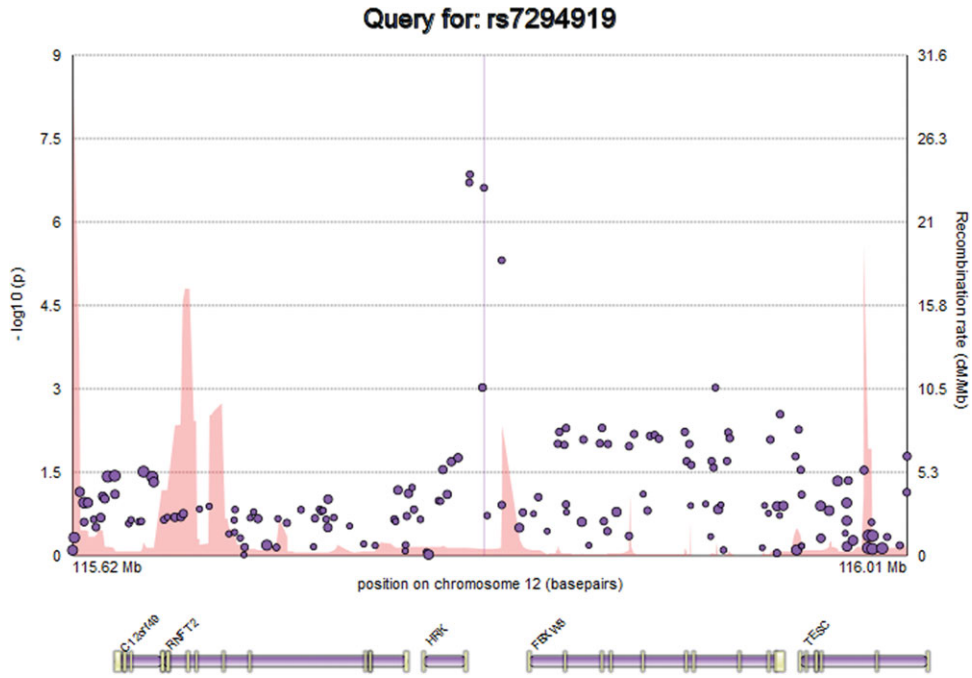
user can also specify flanking region sizes (full documentation of acceptable queries is provided on the EnigmaVis website). Multiple phenotypes can be selected at once, for simultaneous visualization. Users can additionally upload their own data for visualization within EnigmaVis, and can use the full array of features to query their own data, and explore it alone or concurrently with any of our existing data. After creating and submitting a query, an interactive plot is generated.

The plot displays the genomic position of the user's query, in basepairs, on the  $x$  axis, and  $-\log_{10}(p)$  representing the evidence for genetic association between that genomic locus and a trait (e.g., hippocampal volume) on the  $y$  axis, at the left. SNPs are represented on the plot as circles, with the minor allele frequency represented by the radius of the circle (Figure 1). Recombination rate (cM/Mb), attained from the HapMap 3 project, is represented as a filled curve beneath the SNP data points, and corresponds to the labels on the  $y$  axis at the right hand side of the plot (Altshuler et al., 2010). Below the main plot, EnigmaVis displays the position of genes and their exons from the UCSC Genome Browser (Kent et al., 2002). For direct SNP queries, the SNP of interest is indicated on the plot by a vertical line. EnigmaVis currently uses the NCBI 36/hg18 build of the human genome.

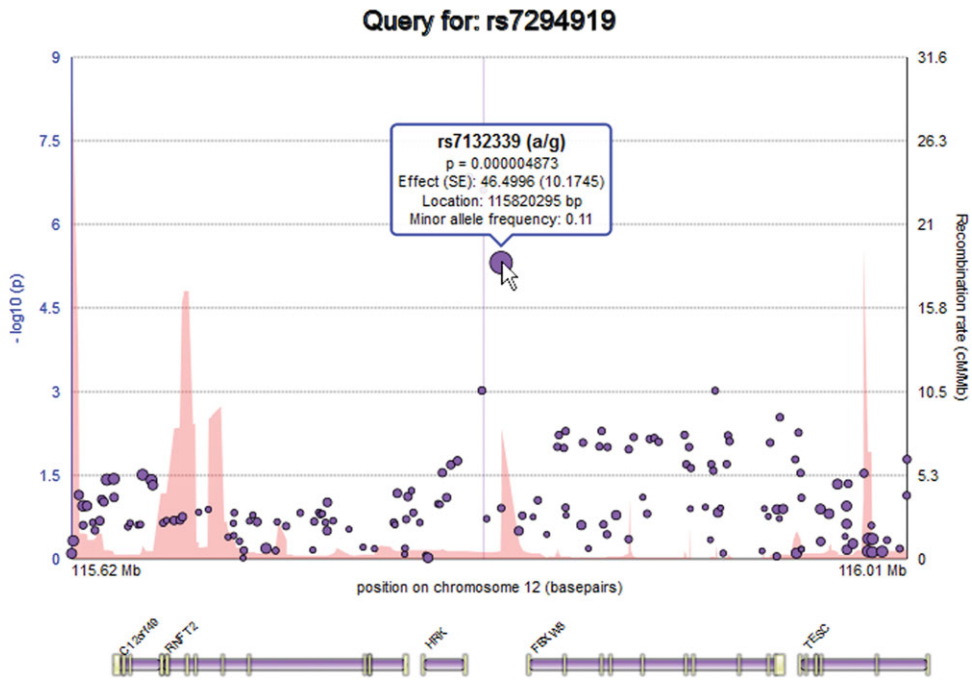
Elements on the plot are interactive. By hovering over an SNP within the plot, the user can display a tooltip providing information about the SNP, including its identifier, chromosome and basepair location, minor allele frequency (based on the ENIGMA discovery population, which is of European ancestry), meta-analytic  $p$ -value, meta-analytic effect size, and meta-analytic standard error (Figure 2). Similarly, the user can hover over a gene to reveal information about the gene. Clicking an SNP or gene generates link-outs that the user can follow to learn more information about that particular SNP or gene. EnigmaVis supports plot navigation in two ways: (i) the user can 'zoom in' (plot a subset of the currently displayed SNPs) by dragging a desired region directly onto the plot, and (ii) the user can pan the plot upstream or downstream by clicking on arrows to the left and right of the plot. EnigmaVis also provides a list of the 'top SNP hits' for each phenotype in the user's query. These are displayed below the plot, ranked in order of  $p$ -value from lowest to highest. Each top SNP can be clicked to generate a new EnigmaVis plot displaying its location.

### Implementation Details

EnigmaVis is comprised of a client-side front-end, a server-side back-end, and a MySQL database. The front-end, which contains the plotting engine and query fields, was written in JavaScript and HTML5 using the Raphaël JavaScript vector graphics library. Using custom software, the MySQL database was populated with data from ENIGMA, the UCSC Genome browser, and HapMap 3. The back-end was written in PHP; this code parses the user's query,



**FIGURE 1**  
The EnigmaVis plot.



**FIGURE 2**  
The EnigmaVis Plot, showing the tooltip that appears when the user hovers over a particular SNP. Not shown: ‘panning’ navigation buttons allow the user to navigate upstream or downstream of the current view.

retrieves all necessary data from the database, and serves it to the front-end for visualization. Most queries are typically served in less than 3 seconds, though queries of larger regions may require up to 20–30 seconds before a plot can be generated.

EnigmaVis can run natively in all major internet browsers without requiring the installation of any third party software or add-ons. It only requires that the browser version be recent enough to provide HTML5 support. The main webpage (<http://enigma.loni.ucla.edu/enigma-vis/>)

features documentation of valid query forms that the user can use to generate plots, as well as features of the application itself. No account or login information is required to use EnigmaVis.

### Future Implementations

As EnigmaVis is still an active, ongoing project, additional features are currently planned for implementation in the future. We aim to find creative ways for users to browse and filter top SNP results, display interactive Manhattan plots, and identify putative functional significance of SNPs in LD with an SNP of interest in a manner that takes advantage of the interactive nature of EnigmaVis. The user could, for example, be allowed to plot the LD to annotated SNPs with functional potential (coding non-synonymous, splice site, 5' and 3' UTR).

Presently, other groups can use EnigmaVis to browse their own GWAS results by using the data upload feature. We aim to further extend EnigmaVis in the future to allow built-in visualization of results from other groups. To accomplish this goal, EnigmaVis will accept GWAS results from other published consortia for inclusion in the EnigmaVis database; these results will be made available for all users of EnigmaVis as they are received.

### Conclusions

We present EnigmaVis, an intuitive, online tool allowing users to visualize and navigate through GWAS results by generating interactive plots. The tool provides, for the first time, a straightforward way to visually examine the large quantities of data collected through the ENIGMA consortium. EnigmaVis is OS-independent, accessible through all common web browsers, does not require installation of special software, and is extendable to groups beyond ENIGMA. Uniquely, EnigmaVis is interactive; once a plot has been generated by a user query, elements on the plot can be interacted with in a live fashion, without generating a new plot. As such, our approach enables interactive interrogating capabilities, tightly coupling the data to analysis and facilitating discovery. EnigmaVis is easily maintainable and is, at present, still an active project. In the future, support will be provided for visualizing additional phenotypes both from the ENIGMA consortium and from other consortia as the data become available, and we anticipate that new features will be implemented that build upon the unique interactive capabilities of EnigmaVis.

### Acknowledgments

We would like to thank all members of the ENIGMA consortium for this use of their data in conducting the meta-analysis. This work was supported by the National Institutes of Health (NIH) and the National Center for Research Resources (NCRR) grant P41 RR013642 to AWT. Grant funding for members of ENIGMA is detailed in

Stein et al. (in press). DH is partially supported by the NSF GRFP grant DGE-0707424. SEM is funded by an Australian Research Council Future Fellowship FT110100548. JS was supported by the NIH Postdoctoral Training Grant in Neurobehavioral Genetics. PT is supported by NIH grants U01 AG024904, EB008432, P41 RR013642, HD050735, AG036535, EB008281.

### References

- Allen, N. C., Bagade, S., McQueen, M. B., Ioannidis, J. P., Kavvoura, F. K., Khoury, M. J., Tanzi, R. E., & Bertram, L. (2008). Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: The SzGene database. *Nature Genetics*, *40*, 827–834.
- Altshuler, D. M., Gibbs, R. A., Peltonen, L., Dermitzakis, E., Schaffner, S. F., Yu, F., . . . McEwen, J. E. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, *467*, 52–58.
- Bertram, L., McQueen, M. B., Mullin, K., Blacker, D., & Tanzi, R. E. (2007). Systematic meta-analyses of Alzheimer disease genetic association studies: The AlzGene database. *Nature Genetics*, *39*, 17–23.
- Cichon, S., Craddock, N., Daly, M., Faraone, S. V., Gejman, P. V., Kelsoe, J., Lehner, T., Levinson, D. F., Moran, A., Sklar, P., & Sullivan, P. F. (2009). Genomewide association studies: History, rationale, and prospects for psychiatric disorders. *The American Journal of Psychiatry*, *166*, 540–556.
- Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, *12*, 996–1006.
- Kremen, W. S., Prom-Wormley, E., Panizzon, M. S., Eyer, L. T., Fischl, B., Neale, M. C., Franz, C. E., Lyons, M. J., Pacheco, J., Perry, M. E., Stevens, A., Schmitt, J. E., Grant, M. D., Seidman, L. J., Thermenos, H. W., Tsuang, M. T., Eisen, S. A., Dale, A. M., & Fennema-Notestine, C. (2010). Genetic and environmental influences on the size of specific brain regions in midlife: The VETSA MRI study. *NeuroImage*, *49*, 1213–1223.
- McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nature reviews. Genetics*, *9*, 356–369.
- Neale, B. M., Medland, S. E., Ripke, S., Asherson, P., Franke, B., Lesch, K. P., . . . & Nelson, S. (2010). Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*, 884–897.
- Peper, J. S., Brouwer, R. M., Boomsma, D. I., Kahn, R. S., & Hulshoff Pol, H. E. (2007). Genetic influences on human brain structure: A review of brain imaging studies in twins. *Human Brain Mapping*, *28*, 464–473.
- Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Glied, T. P., Boehnke, M., Abecasis, G. R., & Willer, C. J. (2010). LocusZoom: Regional visualization of

- genome-wide association scan results. *Bioinformatics*, 26, 2336–2337.
- Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., . . . & Gejman, P. V. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics*, 43, 969–976.
- Ripke, S., & Thomas, B. (2011). Ricopili. Retrieved from <http://www.broadinstitute.org/mpg/ricopili/>
- Sklar, P., Ripke, S., Scott, L. J., Andreassen, O. A., Cichon, S., Craddock, N., . . . & Purcell, S. M. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics*, 43, 977–983.
- Stein, J. L., Medland, S. E., Arias Vasquez, A., Hibar, D., Senstad, R. E., Winkler, A. M., . . . Thompson, P. M., for the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) Consortium. (2012). Identification of common variants associated with human hippocampal and intracranial volumes. *Nature Genetics*, 44, 552–561.
- Thompson, P. M., Cannon, T. D., Narr, K. L., van Erp, T., Poutanen, V. P., Huttunen, M., Lonnqvist, J., Standertskjold-Nordenstam, C. G., Kaprio, J., Khaledy, M., Dail, R., Zoumalan, C. I., & Toga, A. W. (2001). Genetic influences on brain structure. *Nature Neuroscience*, 4, 1253–1258.
-