Wisconsin, USA. [3]*Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, Wisconsin, USA.* [4]*Present addresses: Biomedical Statistics and Bioinformatics, Mayo Clinic, Rochester, Minnesota, USA (M.R.-M.) and Nationwide Children's Hospital, Columbus, Ohio, USA (S.M.L.).*
e-mail: hebbring.scott@mcrf.mfldclin.edu

1. Hebbring, S.J. *Immunology* **141**, 157–165 (2014).
2. Denny, J.C. *et al. Nat. Biotechnol.* **31**, 1102–1110 (2013).
3. Ye, Z. *et al. Eur. J. Hum. Genet.* doi:10.1038/ejhg.2014.123 (30 July 2014).
4. Novac, N. *Trends Pharmacol. Sci.* **34**, 267–272 (2013).
5. Plenge, R.M., Scolnick, E.M. & Altshuler, D. *Nat. Rev. Drug Discov.* **12**, 581–594 (2013).
6. Sanseau, P. *et al. Nat. Biotechnol.* **30**, 317–320 (2012).
7. Okada, Y. *et al. Nature* **506**, 376–381 (2014).
8. Power, A., Berger, A.C. & Ginsburg, G.S. *J. Am. Med. Assoc.* **311**, 2063–2064 (2014).
9. Law, V. *et al. Nucleic Acids Res.* **42**, D1091–D1097 (2014).
10. Cobanoglu, M.C., Oltvai, Z.N., Taylor, D.L. & Bahar, I. *Bioinformatics* **31**, 131–133 (2015).
11. LaBute, M.X. *et al. PLoS ONE* **9**, e106298 (2014).
12. Liu, X. *et al. J. Cheminform.* **6**, 33 (2014).
13. Aronson, A.R. & Lang, F.M. *J. Am. Med. Inform. Assoc.* **17**, 229–236 (2010).
14. A Catalog of Published Genome-Wide Association Studies. http://www.genome.gov/gwastudies. Accessed 1 February 2014.
15. Sun, X., Han, F., Yi, J., Hou, N. & Cao, Z. *Mol. Med. Rep.* **7**, 1636–1640 (2013).
16. Wang, Z.Y. & Zhang, H.Y. *Nat. Biotechnol.* **31**, 1080–1082 (2013).
17. Kilicoglu, H., Shin, D., Fiszman, M., Rosemblat, G. & Rindflesch, T.C. *Bioinformatics* **28**, 3158–3160 (2012).
18. Griffith, M. *et al. Nat. Methods* **10**, 1209–1210 (2013).

Epigenomic annotation of genetic variants using the Roadmap Epigenome Browser

To the Editor:

Advances in next-generation sequencing platforms have reshaped the landscape of functional genomic and epigenomic research as well as human genetics studies. Annotation of noncoding regions in the genome with genomic and epigenomic data has facilitated the generation of new, testable hypotheses regarding the functional consequences of genetic variants associated with human complex traits[1,2]. Large consortia, such as the US National Institutes of Health (NIH) Roadmap Epigenomics Consortium[3] and ENCODE[4], have generated tens of thousands of sequencing-based genome-wide data sets, creating a useful resource for the scientific community[5]. The WashU Epigenome Browser[6–8] continues to provide a platform for investigators to effectively engage with this resource in the context of analyzing their own data. Here, we describe the Roadmap Epigenome Browser (http://epigenomegateway.wustl.edu/browser/roadmap/), which is based on the WashU Epigenome Browser and integrates data from both the NIH Roadmap Epigenomics Consortium and ENCODE in a visualization and bioinformatics tool that enables researchers to explore the tissue-specific regulatory roles of genetic variants in the context of disease.

The browser takes advantage of the over 10,000 epigenomic data sets it currently hosts, including 346 'complete epigenomes', defined as tissues and cell types for which we have collected a complete set of DNA methylation, histone modification, open chromatin and other genomic data sets[9]. Data from both the NIH Roadmap Epigenomics and ENCODE resources are seamlessly integrated in the browser using a new Data Hub Cluster framework (**Supplementary Note** and **Supplementary Figs. 1** and **2**). Investigators can specify any number of single nucleotide polymorphism (SNP)-associated regions and any type of epigenomic data, for which the browser automatically creates virtual data hubs through a shared hierarchical metadata annotation, retrieves the data and performs real-time clustering analysis. Investigators interact with the browser to determine the tissue specificity of the epigenetic state encompassing genetic variants in physiologically or pathogenically relevant cell types from normal or diseased samples (**Supplementary Note**, **Supplementary Tutorial 1** and **Supplementary Figs. 3** and **4**).

We illustrate the epigenomic annotation of two noncoding SNPs, identified from genome-wide association studies of people with multiple sclerosis[10], by clustering the histone H3K4me1 profile of SNP-harboring

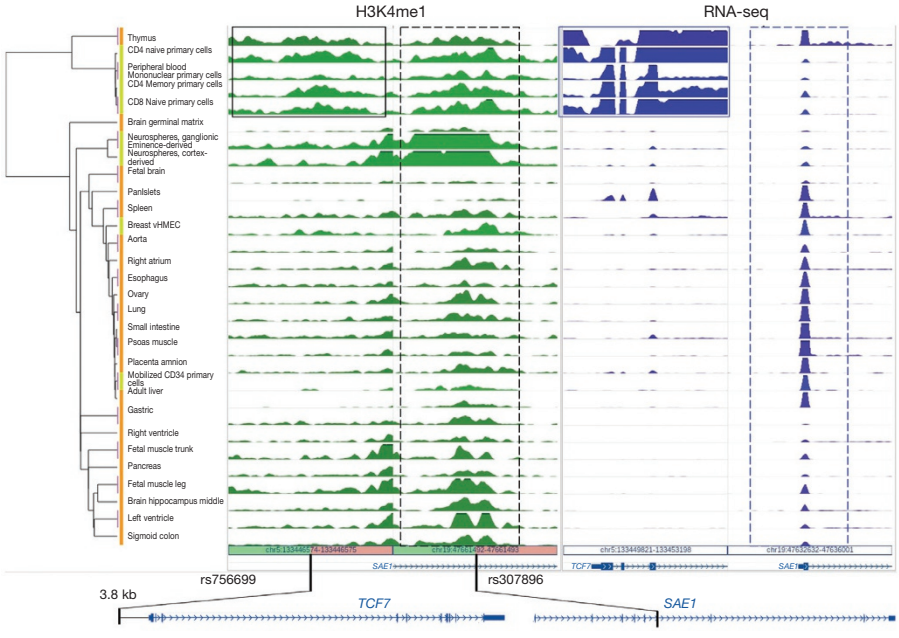

Figure 1 Epigenomic annotation of SNPs. Multiple sclerosis–associated SNPs identified through genome-wide association studies are annotated using epigenomic and expression data from 31 primary human tissues (orange) and cells (light green). H3K4me1 chromatin immunoprecipitation (ChIP)-seq read density (in green) is shown for a 6-kb region centered on each SNP. RNA-seq read density (in blue) is shown over the 5′ end of genes that are closest to these SNPs. Hierarchical clustering is applied to both H3K4me1 and RNA-seq data. The region associated with rs756699 has H3K4me1 mostly confined to immune-related cell types (solid black box). The closest gene, *TCF7* (3.8 kb downstream), also shows high expression in the same group of cell types (solid blue box; **Supplementary Fig. 9**). The region surrounding rs307896 has H3K4me1 signal in all tissues and cell types (dashed black box). SNP rs307896 lies in an intron of *SAE1*, a gene that is also expressed in all the samples (dashed blue box). Normalized gene expression values for *TCF7* and *SAE1* are included in **Supplementary Figure 5**.

regions and RNA-seq signal of their closest genes across multiple primary tissues and cells (**Fig. 1**). Both SNPs lie within putative enhancer regions. Whereas rs307896 marks an enhancer common across cell types, rs756699 is located in an enhancer specific to immune cells and is potentially targeting TCF7, a T cell–specific gene 3.8 kb downstream (**Fig. 1** and **Supplementary Fig. 5**). Thus, reference epigenomes provide important clues into the functional relevance of these genetic variants in the context of the pathophysiology of multiple sclerosis, including inflammation[11].

Investigators can also use the browser to identify covariation of epigenomic, transcriptomic and transcription factor binding profiles across cell types to predict relationships between regulatory sites and target genes (**Supplementary Note**, **Supplementary Tutorial 1** and **Supplementary Figs. 6–8**). Additionally, investigators can explore multiple complete reference epigenomes in different browser panels in parallel using synchronized genomic coordinates or independent genomic coordinates. A variety of Epigenome Browser functions, including gene set view, genome juxtaposition, chromatin interaction display and statistical testing, can be applied to better engage with this resource (**Supplementary Note**, **Supplementary Tutorial 1** and **Supplementary Fig. 9**).

We also provide the means for investigators to build their own Data Hub Clusters of different scales and clone the browser on Amazon Cloud to visualize and analyze private data in the context of public data (**Supplementary Tutorial 2**). These tools, along with the rapidly growing epigenomic data sets of human cells of different states, will facilitate the translation of genetic signals into molecular mechanisms, leading to prognostic, diagnostic and therapeutic advances.

Note: Any Supplementary Information are available in the online version of the paper (doi:10.1038/nbt.3158).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Xin Zhou[1,2], Daofeng Li[1], Bo Zhang[1], Rebecca F Lowdon[1], Nicole B Rockweiler[1], Renee L Sears[1], Pamela A F Madden[2], Ivan Smirnov[3], Joseph F Costello[3] & Ting Wang[1]

[1]Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, Missouri, USA. [2]Department of Psychiatry, Washington University in St. Louis, St. Louis, Missouri, USA. [3]Brain Tumor Research Center, Department of Neurosurgery, Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, California, USA.
e-mail: twang@genetics.wustl.edu or jcostello@cc.ucsf.edu.

Published online 18 February 2015;
doi:10.1038/nbt.3158

1. Paul, D.S., Soranzo, N. & Beck, S. *Bioessays* **36**, 191–199 (2014).
2. Ritchie, G.R., Dunham, I., Zeggini, E. & Flicek, P. *Nat. Methods* **11**, 294–296 (2014).
3. Bernstein, B.E. *et al. Nat. Biotechnol.* **28**, 1045–1048 (2010).
4. ENCODE Project Consortium. *Nature* **489**, 57–74 (2012).
5. Chadwick, L.H. *Epigenomics* **4**, 317–324 (2012).
6. Zhou, X., Li, D., Lowdon, R.F., Costello, J.F. & Wang, T. *Bioinformatics* **30**, 2206–2207 (2014)
7. Zhou, X. *et al. Nat. Methods* **10**, 375–376 (2013).
8. Zhou, X. *et al. Nat. Methods* **8**, 989–990 (2011).
9. Roadmap Epigenomics Consortium *et al. Nature* doi:10.1038/nature14248 (18 February 2015).
10. Sawcer, S. *et al. Nature* **476**, 214–219 (2011).
11. Compston, A. & Coles, A. *Lancet* **359**, 1221–1231 (2002).