



Published in final edited form as:

Epigenomics. 2012 June ; 4(3): 317–324. doi:10.2217/epi.12.18.

The NIH Roadmap Epigenomics Program data resource

Lisa Helbling Chadwick

Division of Extramural Research & Training, National Institute of Environmental Health Sciences, 530 Davis Drive, Maildrop K3-15, Morrisville, NC 27709, USA Tel.: +1 850 727 7218
chadwickL@niehs.nih.gov

Abstract

The NIH Roadmap Reference Epigenome Mapping Consortium is developing a community resource of genome-wide epigenetic maps in a broad range of human primary cells and tissues. There are large amounts of data already available, and a number of different options for viewing and analyzing the data. This report will describe key features of the websites where users will find data, protocols and analysis tools developed by the consortium, and provide a perspective on how this unique resource will facilitate and inform human disease research, both immediately and in the future.

Keywords

chromatin; data resource; data visualization; DNA methylation; epigenetic mapping; epigenetics; epigenomics; histone modification; human disease

The completion of the Human Genome Project marked a significant milestone, one which paved the way for annotation of the full catalog of human genes. This was undeniably a huge step forward for human disease research. The sequence of a gene, however, only provides some insight into its function. Given that each of our cells possesses an identical complement of genes, what differentiates a skin cell from a heart muscle cell from a neuron? Genes must be turned on, off or become expressed at different levels to effect the changes leading to the functional differences between cell types. Therefore, it is equally important to understand how these genes are regulated – when, where and how is a given gene expressed? Epigenetic mechanisms, such as DNA methylation and a variety of post-translational histone modifications, play an important role in establishing gene-expression programs, as well as in maintaining them, as cells divide.

The NIH Roadmap Epigenomics Program [101], funded through the NIH Common Fund, was developed with the goal of investigating how epigenetic mechanisms contribute to human health and disease. This multicomponent program funds research in several relevant areas, including technology development in epigenetics and epigenetic imaging, discovery and characterization of novel epigenetic marks, and investigation of how epigenetic signatures are disrupted in human disease. One key goal of this program is to gain a better understanding of the normal pattern of epigenetic modification, which will allow for

Disclaimer This article may be the work product of an employee or group of employees of the National Institute of Environmental Health Sciences (NIEHS), NIH; however, the statements, opinions or conclusions contained therein do not necessarily represent the statements, opinions or conclusions of NIEHS, NIH or the US government.

Financial & competing interests disclosure The author is an employee of the NIH. The author has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

comparisons between different tissues and cell types, and will serve as a reference for comparison to diseased samples. Recent advances in sequencing technology have made it possible to move beyond gene-by-gene analyses, allowing for truly unbiased, genome-wide mapping of epigenetic modifications. The NIH Roadmap Reference Epigenome Mapping Consortium, a group comprised of four Reference Epigenome Mapping Centers and an Epigenomics Data Analysis and Coordination Center, has been charged with generating these genome-wide epigenomic maps and assembling them into a publicly available data resource (Table 1) [1].

Types of data available: cell types & epigenetic features

The Reference Epigenome Mapping Consortium is focused on developing reference epigenomic maps for a variety of human primary cells and tissues. As is true with any epigenetic study, a number of considerations are involved when selecting samples for mapping. Each of the specific cell types that make up a tissue probably have different epigenomic profiles. However, it can often be nearly impossible to isolate enough material of a particular purified cell type for analysis. The consortium has made an effort to achieve balance by covering a wide range of disease-relevant tissues, while including more highly purified cell types when possible. Currently, a wide range of adult and fetal cells and tissues are represented, including cells from a number of distinct brain regions and a variety of purified blood cell types. In addition, several pluripotent cell lines are included, such as induced pluripotent stem cells, human embryonic stem cells, as well as some differentiated forms of these cells. Currently, over 120 unique human primary cells, tissues and pluripotent cells are represented in the database (Table 1).

Specific epigenetic modifications can often be associated with a particular function; for example, H3K9me3 is generally found in repressed regions of the genome, while H3K9ac is generally correlated with gene activation. However, simply determining the distribution of one mark is not sufficient, as the function of a given mark may vary depending upon the broader chromatin context in which it resides. Furthermore, these marks must be correlated with a functional outcome, such as altered gene expression. A key strength of this data resource is the fact that for each cell and tissue type represented, multiple features will be ultimately be mapped, including DNA methylation, post-translational histone modifications, chromatin accessibility and RNA. DNA methylation data will be made available for all cell and tissue types represented. Detailed protocols and standards for each of these analyses have been made available to the community online [102].

The consortium employs several methods for the analysis of DNA methylation, each of which have strengths and limitations (several of these were described in some detail by Bock and colleagues [2]). These include reduced representation bisulfite sequencing [3] and a combination of complementary MeDIP-seq/MRE-seq assays [4]. More recently, the consortium has moved towards using MethylC-seq [5] wherever possible, which offers full coverage DNA methyl ation maps at base pair resolution. As methods for high-throughput analysis of 5-hydroxymethylcytosine are being developed, this feature may also be added to a subset of cell types in the future. Two approaches are used for the analysis of histone modifications by chromatin immunoprecipitation with sequencing. A small number of high-value samples, including several embryonic stem cell lines and their differentiated forms, will be analyzed to significant depth, with a large panel of histone modifications. Currently, there are approximately 30 distinct modifications represented in this panel, with additional modifications added as suitable antibodies are identified. The data gained from these more comprehensive analyses are used to inform the selection of a more limited panel of histone modifications, which is applied to the majority of samples being analyzed. Currently, this panel includes H3K27me3, H3K36me3, H3K4me1, H3K4me3, H3K9me3 and H3K27ac.

These are the modifications found to be the most informative, namely the ones that are most difficult to predict based on other modifications [ERNST J, KELLIS M, PERS. COMM.]. In addition to DNA methylation and histone modifications, most samples will undergo DNase I hypersensitivity mapping [6], to provide a measure of chromatin accessibility. Finally, each sample will be analyzed for RNA content. In many cases, this will be accomplished with expression arrays, but the consortium is moving towards using RNA-sequencing (RNA-seq) [7] when possible. RNA-seq offers the most comprehensive view of RNA expression. It includes small RNAs, provides a measurement of alternative splicing events and enables allelic analyses to be carried out. The most current standards in use for chromatin immunoprecipitation with sequencing, whole-genome bisulfite sequencing, and RNA-seq can be found at the Consortium homepage, under 'Protocols and Data Standards' [102].

Navigating the data

In order to maximize community accessibility to the data and resources developed by the program, we have developed several websites for viewing and downloading data and other information, each with unique features. These are summarized in Table 1 and described below.

Reference Epigenome Mapping Consortium homepage

The centerpiece of the program is the Reference Epigenome Mapping Consortium homepage [103], a centralized landing page which houses consortium data, as well as a variety of additional information, including detailed experimental protocols and information about data quality metrics, downloadable tools for processing and analyzing sequencing data, as well as links to other websites associated with consortium data (Figure 1). A list of all consortium publications can also be found here. Tabs located at the top of the page facilitate navigation of the site. This site is continually evolving in an effort to maximize the user experience and facilitate use of the resource by the community.

This site also features an easy-to-use interface to browse consortium data. Clicking on the 'Data' tab will open a matrix-style data browser. Cell and tissue types are grouped into anatomic categories for easy navigation. Available data types are indicated as shaded squares. Clicking on any of these squares will open a track selection window on a consortium-hosted University of California-Santa Cruz (UCSC; CA, USA) Genome Browser Mirror site [104] where the user can select specific data tracks to be viewed for that cell type. Once tracks of interest are selected, the user must change the 'maximum display mode' to full, using the drop down menu, and click 'submit'. Users can also choose to view all available data for a particular epigenetic feature across all cell types by using the drop down menu found at the top of the matrix browser ('select assay to view cell/tissue data'). This site also offers a unique visual data browser, which displays cell types with available data in an anatomical context.

The consortium's goal is to provide a complete data set for each cell type analyzed. As described earlier, this data set – referred to as a 'complete epigenome' – would contain DNA methylation data, RNA expression data, a panel of histone modifications and DNase I hypersensitivity profiles where possible. Definitions of the various classes of complete epigenomes used by the consortium can be found on this site by clicking the 'Complete Epigenomes' tab at the top of the page. This page will also be updated with a list of all cell/tissue types that have reached completed status.

National Center for Biotechnology Information

The National Center for Biotechnology Information (NCBI) serves as the long-term archive for data produced by the Reference Epigenome Mapping Consortium, as well as for the

other epigenomics projects funded by the Roadmap Epigenomics Program. Rapid data release is an important goal of the consortium. While consortium data can be viewed on several websites, most are updated only after a data freeze, which occurs several times a year. By contrast, the two NCBI sites described below are updated continuously, providing a real-time picture of the data available.

Users interested in simply downloading data files for analysis with their own tools should begin at the NIH Roadmap Epigenomics page of the Gene Expression Omnibus (GEO) [105]. This site offers several options for navigating the data: listed by sample (a particular feature in a particular cell type), using a matrix-style browser, or by search terms; however, only the sample list is updated continuously. Clicking on any accession number will open the metadata associated with the sample, where the user will find specific information about the individual the sample was derived from, the experimental conditions, and data generation and processing. The data can be downloaded in several of the most commonly used file formats (e.g., .bed, .wig, .bam and Short Read Archive).

NCBI has also developed an epigenomics portal [8,106], which hosts data generated by the Roadmap Epigenomics Program, as well as other, user-submitted epigenomic data, including data from the Encyclopedia of DNA Elements (ENCODE) and ENCODE in model systems programs [9], a complementary program focused on defining the functional elements that control gene expression. Data can be browsed either by experiment (i.e., by individual data set) or by sample (i.e., by cell or tissue type). Here, data files can be downloaded or viewed on a genome browser. This allows users to view epigenomic data in the context of other genomic data housed at the NCBI, such as SNPs, clinically associated variants and significant genome-wide association studies (GWAS) associations. The epigenomics portal also features a tool to compare the epigenetic profiles of two different samples, pulling out a list of genes and associated gene-ontology terms that differ most significantly in epigenetic profiles between the samples.

Human Epigenome Atlas

The Human Epigenome Atlas [107] was developed by the consortium's Epigenomic Data Analysis and Coordination Center at the Baylor College of Medicine (TX, USA). As with the other sites described, one can quickly browse the available data using a matrix style browser, download data and metadata, and view data on a genome browser, the UCSC genome browser in this case. This site also includes more extensive explanations of the metadata collected, the flow of data from Mapping Center to Epigenomic Data Analysis and Coordination Center to NCBI, and the analysis pipelines in place. This information can all be found under the 'Informatics' tab at the top of the page.

One unique feature of this site is the Atlas Gene Browser. This allows users to focus on the epigenetic profile of the introns, exons, promoters and UTR of a specific gene(s) or pathway of interest, to assist in identifying common regulatory themes. This site also hosts the Genboree Workbench (also found under the 'Informatics' tab), which contains a number of consortium-developed tools for data analysis and quality checking of epigenomic data. Although the Genboree workbench is open to the public, one must create an account in order to use these tools. This allows the user to upload and securely store their own data for analysis with the tools provided.

UCSC Track Hub

The UCSC Genome Browser [10,108] remains one of the most popular tools for viewing genomic data. The UCSC browser is also the home to data generated by the ENCODE program [11]. In an effort to make the Roadmap Epigenomic Mapping Consortium data

available to the public within the UCSC interface, the VizHub was developed [109]. VizHub stores individual data tracks and summary tracks that can be loaded remotely onto the UCSC browser. In order to access the Roadmap Track Hub at the UCSC Human Genome Browser, the user must navigate to the UCSC Human Genome Browser Gateway [110] and click the 'Track Hubs' button. The track hub of interest can be selected and loaded, allowing Roadmap Epigenomics data to be displayed on the browser. The user can navigate the available data in the same way as one would usually do on the UCSC browser. For example, clicking on 'DNA methylation' would open a track selection window where available methylation tracks may be selected.

The Human Epigenome Browser

The Human Epigenome Browser [111] is a next-generation epigenome browser developed to tackle the challenge of navigating and viewing large amounts of epigenomic data and the associated metadata. The features of this novel browser have been thoroughly described in a recent publication [12], so it will not be covered in detail here.

Future perspective

The NIH Roadmap Epigenomic Mapping Consortium has made great strides in developing reasonably comprehensive epigenomic maps for a variety of normal human cell and tissue type [112]. The next challenge is to refine these maps even further. Each of our organs are made up of many distinct, specialized cell types. For some cell types, such as hematopoietic cells, one can obtain a highly purified cell type in sufficient quantities to allow for genome-wide epigenetic analyses, but for most cell types this remains a significant technical challenge. The consortium has made an effort to strike a balance between highly purified cell types and more heterogeneous tissues in selecting samples to be mapped, but as technologies continue to rapidly improve, so will the ability to define the epigenomes of these unique cell types.

As the NIH Roadmap Epigenomics Program nears completion in 2013, the data available to the community continues to grow. The next step will be in applying these data to understanding human disease. How are these normal epigenetic programs disrupted in disease, and if changes are observed, are they a cause or a consequence of disease? In a separate initiative, the NIH Roadmap Epigenomics Program has also funded a number of investigators tackling this question, who are seeking to clarify the role of epigenetics in the pathogenesis of a variety of complex human diseases, including Alzheimer's disease, glaucoma, atherosclerosis, bipolar disorder, asthma and autism [113]. As these and other similar studies begin to bear fruit, we will begin to be able to answer this question.

An additional challenge is, how we can leverage the data in this resource to understand the role of genetic variants in disease? Some intriguing data arising from the ENCODE program suggest how genome-wide information about chromatin state may be used to identify functional genomic elements linked to hits arising from GWAS [13]. One of the major challenges that has arisen in these studies is that SNPs strongly associated with the disease of interest are often found in regions of the genome with no obvious functional significance. Ernst and Kellis used a novel bioinformatics method to identify six general classes of chromatin states; specific combinations of epigenetic modifications associated with a functional state, such as promoter, enhancer, insulator, transcribed, repressed and inactive [14]. They found that in a number of existing GWAS data sets, disease-associated SNPs fell into regions identified as cell-type-specific enhancers in a biologically relevant cell type. Going forward, the reference epigenomic maps developed by the Roadmap Epigenomic Mapping Consortium, which cover a far broader range of human primary cells and tissues

than currently represented in ENCODE, will be incredibly valuable for investigators trying to understand how specific genetic variants contribute to disease.

Finally, once we have determined what constitutes a 'normal' epigenome, we can begin to investigate how epigenetic profiles change in response to the environment. As the agouti viable yellow (A^{VY}) mouse has demonstrated so strikingly [15–17], the epigenome is exquisitely sensitive to external factors, such as diet or environmental chemicals. The reference epigenomes developed by the consortium will form a foundation for future studies aimed at understanding how the epigenome is perturbed, to identify biomarkers of exposure, and ultimately moving towards the development of intervention strategies to minimize the health impact in exposed populations.

Acknowledgments

Without the hard work and collegiality of the investigators and NIH staff that make up the Reference Epigenome Mapping Consortium, this data resource would not be possible. The four Reference Epigenome Mapping Centers are headed by J Stamatoyannopoulos (University of Washington, WA, USA), B Bernstein and A Meissner (Broad Institute, MA, USA), B Ren (University of California San Diego, CA, USA) and J Costello (University of California San Francisco, CA, USA); the Epigenomics Data Analysis and Coordination Center (TX, USA) is led by A Milosavljevic (Baylor College of Medicine, TX, USA); the National Center for Biotechnology Information (NCBI; MD, USA) effort is led by G Schuler (NCBI). The program is directed by staff at the National Institute of Environmental Health Sciences (NC, USA) and the National Institute on Drug Abuse (MD, USA).

References

Papers of special note have been highlighted as:

■ of interest

■ ■ of considerable interest

- Bernstein BE, Stamatoyannopoulos JA, Costello JF, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* 2010; 28(10):1045–1048. [PubMed: 20944595] ■ ■ White paper describing the scope and goals of the NIH Roadmap Epigenomic Mapping Consortium.
- Bock C, Tomazou EM, Brinkman AB, et al. Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat. Biotechnol.* 2010; 28(10):1106–1114. [PubMed: 20852634] ■ A comprehensive comparison of the data generated by four popular platforms for genome-wide DNA methylation analyses, highlighting the strengths and limitations of each technology. This will be of particular interest for investigators planning their own genome-wide DNA methylation studies.
- Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 2005; 33(18):5868–5877. [PubMed: 16224102]
- Maunakea AK, Nagarajan RP, Bilenky M, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature.* 2010; 466(7303):253–257. [PubMed: 20613842]
- Lister R, Pelizzola M, Dowen RH, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature.* 2009; 462(7271):315–322. [PubMed: 19829295] ■ Describes the first human DNA methylomes.
- Hesselberth JR, Chen X, Zhang Z, et al. Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting. *Nat. Methods.* 2009; 6(4):283–289. [PubMed: 19305407]
- Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 2009; 10(1):57–63. [PubMed: 19015660]
- Fingerman IM, Mcdaniel L, Zhang X, et al. NCBI epigenomics: a new public resource for exploring epigenomic data sets. *Nucleic Acids Res.* 2011; 39(Database issue):D908–D912. [PubMed: 21075792]

9. Birney E, Stamatoyannopoulos JA, Dutta A, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007; 447(7146):799–816. [PubMed: 17571346]
10. Kent WJ, Sugnet CW, Furey TS, et al. The Human Genome Browser at UCSC. *Genome Res*. 2002; 12(6):996–1006. [PubMed: 12045153]
11. Rosenbloom KR, Dreszer TR, Pheasant M, et al. ENCODE whole-genome data in the UCSC genome browser. *Nucleic Acids Res*. 2010; 38(Database issue):D620–D625. [PubMed: 19920125]
12. Zhou X, Maricque B, Xie M, et al. The Human Epigenome Browser at Washington University. *Nat. Methods*. 2011; 8(12):989–990. [PubMed: 22127213] ■■■Describes the next-generation epigenome browser developed by the consortium.
13. Ernst J, Kheradpour P, Mikkelsen TS, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011; 473(7345):43–49. [PubMed: 21441907] ■■■Uses a novel bioinformatic method to annotate chromatin states across the genome in the nine cell types mapped by the Encyclopedia of DNA Elements Consortium, and uses these data to determine the possible functional significance of SNPs arising as top hits in several genome-wide association studies.
14. Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol*. 2010; 28(8):817–825. [PubMed: 20657582]
15. Dolinoy DC, Huang D, Jirtle RL. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proc. Natl Acad. Sci. USA*. 2007; 104(32):13056–13061. [PubMed: 17670942]
16. Dolinoy DC, Weidman JR, Waterland RA, Jirtle RL. Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. *Environ. Health Perspect*. 2006; 114(4):567–572. [PubMed: 16581547]
17. Wolff GL, Kodell RL, Moore SR, Cooney CA. Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *FASEB J*. 1998; 12(11):949–957. [PubMed: 9707167]

Websites

101. NIH Common Fund Epigenomics Program. <http://commonfund.nih.gov/epigenomics>
102. Protocols and Data Standards on the Consortium homepage. www.roadmapepigenomics.org/protocols
103. Reference Epigenome Mapping Consortium homepage. <http://roadmapepigenomics.org>
104. Epigenomics mirror of UCSC Genome Browser. www.epigenomebrowser.org
105. Roadmap Epigenomics Program data at GEO. www.ncbi.nlm.nih.gov/geo/roadmap/epigenomics
106. NCBI Epigenomics Portal. www.ncbi.nlm.nih.gov/epigenomics
107. Human Epigenome Atlas. <http://epigenomeatlas.com>
108. UCSC Genome Browser. <http://genome.ucsc.edu>
109. Roadmap Epigenomics Visualization Hub. <http://vizhub.wustl.edu>
110. UCSC Human Genome Browser Gateway. <http://genome.ucsc.edu/cgi-bin/hgGateway>
111. Human Epigenome Browser at Washington University. <http://epigenomegateway.wustl.edu>
112. Link to data on the Consortium home page. www.roadmapepigenomics.org/data
113. Link to list of grants funded through the NIH Roadmap Epigenomics Program. <http://commonfund.nih.gov/epigenomics/fundedresearch.aspx>

Executive summary

Data available

- The NIH Roadmap Reference Epigenome Mapping Consortium is assembling a community resource of reference epigenomic maps from a broad range of human primary cells and tissues.
- Currently, over 120 unique cell and tissue types are represented, including a variety of adult and fetal tissues, and pluripotent cell lines (induced pluripotent stem cells and embryonic stem cells).
- An 'epigenome' contains genome-wide data for DNA methylation, a reduced panel of histone modifications, RNA and DNase I hypersensitivity profiling when possible. A subset of cell types will be mapped with a much broader panel of histone modifications.

Navigating the data

- A centralized homepage has been established; here, users can browse available data, view data tracks on a University of California-Santa Cruz Genome Browser mirror, obtain detailed protocols and quality metrics used by the program and find consortium publications.
- The National Center for Biotechnology Information (NCBI) is the long-term archive of the data. Data and metadata can be downloaded at Gene Expression Omnibus. NCBI has also developed an epigenomics portal where users can browse data, view it in the context of other genomic data hosted at NCBI, and use an analysis tool to identify genes whose epigenetic profiles differ most significantly between two cell types.
- The Human Epigenome Atlas provides an interface by which users can view or download data and metadata. This site also hosts the Genboree Workbench, where investigators can use consortium-developed data analysis tools with their own data. Data from the consortium can also be viewed directly on the University of California-Santa Cruz Genome Browser using their 'Track Hub' feature.
- The Human Epigenome Browser is a next-generation browser that allows the user to easily manage large numbers of data tracks, and view data and metadata concurrently.

Future perspective

- The current resource represents the first step in mapping the epigenomes of different human cell types. As technologies continue to improve, similar analyses will be able to be performed in more highly purified unique cell types.
- The data developed by the consortium will provide a baseline for comparison in studies investigating how epigenetic changes contribute to complex disease.
- Epigenomic data such as these are beginning to show promise in annotating the functional significance of genetic variants associated with disease that do not map to coding regions of the genome.

- The establishment of baseline epigenomic profiles sets the stage for future studies investigating how external factors, such as diet or environmental chemicals, can impact epigenetic programming, leading to disease.

The image shows the homepage of the NIH Roadmap Epigenomics Mapping Consortium. At the top, there is a search bar and a navigation menu with links for HOME, PARTICIPANTS, Browse DATA, PROTOCOLS, COMPLETE EPIGENOMES, TOOLS, and PUBLICATIONS. Below the menu are several content blocks: 'OVERVIEW', 'PROJECT DATA', 'MAPPING CENTERS', 'PROTOCOLS & STANDARDS', 'PUBLICATIONS', and 'NEWS'. A central diagram illustrates epigenetic concepts including Chromatin, RNA, DNA methylation, DNase I hypersensitive sites, and Histone Modifications. A 'VIEW/DOWNLOAD QUICK LINKS' section provides direct access to Genome Browsers (www.epigenomebrowser.org, genomebrowser.wustl.edu, epigenomegateway.wustl.edu), Data Repositories (NCBI Epigenomics Gateway, Epigenome Atlas), and a NEWS section with recent updates.

Figure 1. The NIH Roadmap Epigenomic Mapping Consortium homepage
 Users interested in browsing data, protocols and other program resources should start here, at the consortium homepage. This site also links to additional program websites, which each offer unique features for data viewing or analysis [103].

Table 1

A reference guide to the NIH Roadmap Epigenomic data resource.

	Download data?	Browse data with...	View data on...	Updated	Other features	Can upload own data?	Ref.
Reference Epigenome Mapping Consortium Homepage	Links to data download	Clickable data matrix or visual data browser	Links to UCSC browser mirror (Epigenome Browser)	Data: at each data freeze (four-times per year) Other items: as needed	Protocols, publications, quality metrics, project and center/group information	No (but linked Epigenome Browser supports upload)	[103]
NCBI Epigenomics Hub	.wig	Sample (i.e., cell/tissue type) browser, experiment (i.e., epigenetic feature) browser, text search	NCBI epigenomics viewer or UCSC browser mirror	Continuously	'Compare Samples' tool to identify regions of greatest chromatin differences, suggests GO terms and pathways most associated	Being implemented	[106]
NCBI Gene Expression Omnibus	.bed, .wig, .bam and SRA	By sample, study or data matrix	NCBI Epigenomics viewer	Continuously		N/A	[105]
The Human Epigenome Atlas (on Genboree)	.bed, .wig by ftp or http	By sample, assay or clickable data matrix	UCSC browser, Atlas Gene/Pathway browser (read densities across single genes or pathways)	At each data freeze	Information on metadata, data flow and data quality. Tools for analysis via Genboree workbench (independent tools and Galaxy pipelines). Data and functionality exposed via HTTP REST APIs for programmatic use and extension		[107]
Roadmap Epigenomics Visualization Hub	No	ENCODE style data matrix	UCSC browser mirror, or remote display at UCSC main site (see [108], click 'load track hub')	At each data freeze	UCSC mirror hosts integrative analysis tracks and summary tracks, viewable at UCSC main site	Being implemented	[108,109]
Human Epigenome Browser at Washington University	Yes	Expandable data selection matrix and metadata matrix	Next-generation epigenome browser	At each data freeze	Google map-style zoom and pan, genomic data and metadata viewer, data collation view, pathway/gene set view, statistical analysis	Yes	[111]
Epigenome Browser UCSC mirror	.bed, .wig through table browser	UCSC data selection matrix (ENCODE style)	UCSC browser mirror	At each data freeze	High-utility UCSC mirror tracks	Yes	[104]

Download data?	Browse data with...	View data on...	Updated	Other features	Can upload own data?	Ref.
Individual reads not available						

API: Application programming interface; ENCODE: Encyclopedia of DNA Elements; GO: Gene-ontology; HTTP: Hypertext transfer protocol; N/A: Not applicable; NCBI: National Center for Biotechnology Information; REST: Representational state transfer; SRA: Short Read Archive; UCSC: University of California-Santa Cruz.