

The landscape for epigenetic/epigenomic biomedical resources

Kabita Shakya,^{1,2,3,*} Mary J. O'Connell^{2,3} and Heather J. Ruskin^{1,3}

¹Centre for Scientific Computing and Complex Systems Modelling (SCI-SYM); Dublin City University; Dublin, Ireland; ²Bioinformatics and Molecular Evolution Group; School of Biotechnology; Dublin City University; Dublin, Ireland; ³School of Computing; Dublin City University; Dublin, Ireland

Keywords: biomedical resource, data mining, epigenetics, epigenomics, primary database, secondary database, methylation

Abbreviations: BIG, Beijing Institute of Genomics; BRO, Biomedical Resource Ontology; DDBJ, DNA Data Bank of Japan; EBI, European Bioinformatics Institute; ENCODE, Encyclopaedia of DNA Elements; EPITRON, Epigenetic Treatment of Neoplastic Disease; HEP, Human Epigenome Project; HEROIC, High-Throughput Epigenetic Regulatory Organization in Chromatin; HGP, Human Genome Project; ICGC, International Cancer Genome Consortium; IHEC, International Human Epigenome Consortium; NCBI, National Center for Biotechnology Information; NHGRI, National Human Genome Research Institute; NIH, National Institute of Health; NLM, National Library of Medicine; PDB, Protein Data Bank; SCOP, Structural Classification of Proteins; UHN, University Health Network

Recent advances in molecular biology and computational power have seen the biomedical sector enter a new era, with corresponding development of bioinformatics as a major discipline. Generation of enormous amounts of data has driven the need for more advanced storage solutions and shared access through a range of public repositories. The number of such biomedical resources is increasing constantly and mining these large and diverse data sets continues to present real challenges. This paper attempts a general overview of currently available resources, together with remarks on their data mining and analysis capabilities. Of interest here is the recent shift in focus from genetic to epigenetic/epigenomic research and the emergence and extension of resource provision to support this both at local and global scale. Biomedical text and numerical data mining are both considered, the first dealing with automated methods for analyzing research content and information extraction, and the second (broadly) with pattern recognition and prediction. Any summary and selection of resources is inherently limited, given the spectrum available, but the aim is to provide a guideline for the assessment and comparison of currently available provision, particularly as this relates to epigenetics/epigenomics.

Introduction

The Human Genome Project (HGP) in 2003 led to identification of more than 20,000 genes and determined the three billion chemical base pairs of human DNA. With the tremendous advances in medical technologies, corresponding development in computational power, storage capacity, inter-connectivity and cost effectiveness, this explosive growth has resulted in the generation and collection of all aspects of biomedical data and, in

the past decade, the importance of bioinformatics has been recognized.¹ Data warehousing,² as a way of dealing with large data set size, combines databases across an entire enterprise, whereas independent or federated systems seek to integrate multiple autonomous databases into a single federation, with constituent databases interconnected via a network and often geographically decentralised.^{3,4} One example is the many bioinformatics data sources linked by the Entrez life sciences search engine.⁵

Biomedical data cover a wide range, from patient records to information from pharmaceutical studies, specific disease research and different “omics” studies, including genomics, proteomics and transcriptomics. Resource types can be classified by two key features: first, the means or method by which access is provided to entities; second, the nature of the entities themselves. The repository or web service that provides access to these data are a vital component of biomedical data resourcing.⁶ An example is PubMed, the NLM's web-based interface to MEDLINE, the premier bibliographic index to journal articles in the life sciences. In general, resource providers, such as PubMeth and MutationDB, review research papers from the domain and mine these for information relevant to the scientific audience. Typically, non-profit research institutes, such as the Sanger Institute, University of California Santa Cruz (UCSC), National Center for Biotechnology Information (NCBI), National Institute of Health (NIH), European Molecular Biology Laboratory (EMBL) and European Bioinformatics Institute (EBI), among others, make such data publicly available over the internet so that these can be further analyzed/mined for knowledge discovery.

Biological/biomedical resources may be one of several types, primary, secondary or composite. Examples of primary database containing information on biological quantities themselves indicate those for sequence or structure, e.g., SwissProt, PIR (protein sequences), GenBank and DDBJ (genome sequences). Secondary resources contain derived information from primary sources and examples include eMOTIF (Stanford) and SCOP (Cambridge). Composite resources typically draw information from a variety of

*Correspondence to: Kabita Shakya; Email: kabita.shakya@gmail.com
Submitted: 04/12/12; Revised: 07/11/12; Accepted: 07/13/12
<http://dx.doi.org/10.4161/epi.21493>

different databases, such as those of the NCBI genome browser and Genecards.⁷ The most popular genome browsers today are Ensembl, NCBI Map Viewer and UCSC, which act as gateways for access to genetic and epigenetic information.

Following completion of the Human Genome Project, increased attention has been paid to processes that lead to heritable changes in gene expression, during development or across generations, without altering the nucleotide sequence within the DNA. Both epigenetics and epigenomics, the genome-wide distribution of epigenetic changes, have become major areas of research focus. Principal epigenetic phenomena encompass DNA methylation, histone modification (methylation/demethylation, acetylation/deacetylation, phosphorylation, ubiquitylation and sumoylation), gene silencing, genomic imprinting and X-chromosome inactivation. Recently-launched large-scale initiatives include, among others, IHEC (International Human Epigenome Consortium),⁸ which plans to map up to 1,000 reference epigenomes within a decade, and the Human Epigenome Project (HEP),⁹ which aims to identify, catalog and interpret genome-wide DNA methylation patterns of all human genes in all major tissues.¹⁰

Epigenetics, cancer and other diseases. Epigenetic abnormalities have been found to be causative factors of cancer, genetic disorders and pediatric syndromes, as well as contributory factors of autoimmune diseases and aging.¹¹ The recent intensive research on cancer-epigenetics has also led to the discovery of many epigenetic markers that play an important role in disease initiation. As a consequence, cancer-related epigenetic resources preponderate over others. Two of the large-scale project initiatives for cancer research include ICGC (see “ICGC” section below) and TCGA (The Cancer Genome Atlas). TCGA has achieved comprehensive sequencing, characterization and analysis of the genomic changes in various cancers and intends to chart the genomic changes involved in more than 20 types of cancers.¹² All of the epigenetic resources are outlined in the following sections, with additional assessment of their data mining capabilities, intrinsic or externally accessed, and their adequacy provided where possible.

DNA methylation can induce “epigenetic silencing” (or loss of expression) of tumor suppressor genes, causing normal cells to be transformed into cancer cells and is the first and most common epigenetic alteration to be observed.^{13,14} A direct link also exists between DNA methylation and histone modification, since a number of proteins involved in DNA methylation (e.g., DNMTs and MBDs) directly interact with histone modifying enzymes, such as histone methyltransferases (HMTs) and histone deacetylases (HDACs).¹⁵ Epigenetic resources incorporating methylation signatures are described in the “Methylation” section below.

Resources for Epigenetic/Epigenomic Signatures

Epigenetic/epigenomic resources are inevitably less comprehensive to date but can be broadly categorized in terms of type of data content, tools and access, and are described below.

Methylation. Pubmeth,¹⁶ a cancer methylation database, provides a sorted, annotated and summarized overview of genes,

reported to be methylated in various cancers, with user query based on gene or cancer type. PubMeth draws on text-mining of Medline/PubMed abstracts, combined with manual annotation of pre-selected abstracts. The text mining approach in PubMeth is fast and intelligent, enabling search of multiple aliases and textual variants of these aliases, and querying of multiple keyword-lists simultaneously. PubMeth also provides the facility to browse a pre-computed gene list, without having to query the database directly.

MethDB¹⁷ is also a major source for experimentally confirmed DNA methylation data but is general, more sample-oriented and not optimized to cancer-related queries. The database is designed to store and annotate information on the occurrence of methylated cytosines in DNA. It currently contains 19,905 methylation content data items and 5,382 methylation patterns or profiles for 48 species, 1,511 individuals, 198 tissues and cell lines and 79 phenotypes. MethDB also has a public online submission system available.¹⁸ The resource forms part of an integrated network of biological databases through DAS (Distributed Annotation System), enabling the epigenetic data to be viewed as a layer in the human genome, and is also connected to Ensembl (for DNA sequences with available MethDB data aligned to NCBI Refseq).

A subset resource, MethPrimerDB,¹⁹ is a database of primer sequences used in PCR-based methylation methods. The database depends on submissions by users and administrators that guarantee the required quality of the database but not necessarily its completeness. To date, there are 29 primer sets. In 2006, the MethBLAST feature was added to MethPrimerDB oligonucleotide sequences. Further updates since 2006, however, are not found for this resource.

MethyCancer²⁰ is a disease-oriented database, specifically of human DNA methylation and cancer that aims to integrate methylation databases and has developed a meta-data format for data standardization, with manual curation still used for noisy data. Four main types of data are included in MethyCancer, namely, (1) CGI clones and global CGI predictions, (2) DNA methylation data, (3) cancer information, genes and mutations and (4) correlations of DNA methylation, gene expression and cancer. MethyView, a visualization tool from MethyCancer, is used to facilitate the browsing of methylation data in the context of existing human genome annotations. A search engine to query different data types and interactions from the MethyCancer database provides simple keyword search and also offers advanced options namely, “methylation,” “gene,” “cancer,” “clone” and “repeat” searches. For example, Methylation search enables the user to specify and combine query options, such as methylation type (pattern, profile, content and domain), data source (BIG/UHN, MethDB,¹⁷ HEP,⁹ Columbia University), experimental methods, sample information (tissue, sex, age and phenotype) and chromosomal positions.

On similar lines, Methylogix²¹ provides a high density DNA methylation database of human chromosomes 21 and 22, a CpG island DNA methylation database for male germ cells, enabling comprehensive analysis of DNA methylation variation between and within the germ lines of normal males, and a targeted DNA methylation database of late-onset Alzheimer disease. Similarly,

MethTools is a collection of software tools for handling and analysis of DNA methylation data, generated by the bisulfite genomic sequencing method.²²

Genomic imprinting related resources. Genomic imprinting is an important epigenetic phenomenon whereby inherited genes are “imprinted” due to one copy of the gene being epigenetically marked or imprinted in either the egg or the sperm. Thus, the allelic expression of an imprinted gene depends on whether it is inherited maternally or paternally. Imprinted expression can also vary between tissues, developmental stages and species.²³ The GeneImprint database²⁴ includes genes and related information on genomic imprinting for different animals including humans and gathered from NCBI. Genes are listed by species and sorted by chromosomal location, name and imprinting status and are provided through the web-interface. Similarly, an imprinted gene and parent-of-origin effect database²⁵ presents imprinted genes and related effects. This consists of two sections: (1) catalog of current literature on imprinted genes in humans and animals and (2) catalog of reports of parental origin of de novo mutations in humans alone. The addition of (2), showing a parent-of-origin effect, expands the scope of the database and provides a useful tool for examining parental origin trends for different types of spontaneous mutations. This second section currently includes more than 1,700 mutations, found in 59 different disorders. The 85 imprinted genes are described in 152 entries from several mammalian species. In addition, more than 300 other entries describe a range of reported parent-of-origin effects in animals.²⁶ A further resource, containing information on mouse gene imprinting,²⁷ also includes an imprinting catalog, as well as chromosome anomalies on mutant mouse lines. This represents integration of curated information from the MRC Harwell stock resource and other Harwell databases, with additional information from external data resources such as IMSR (International Mouse Stain Resource).

Histone and chromatin-related resources. The Histone Database,²⁸ of the National Human Genome Research Institute, provides a complete set of histone protein sequences. Nucleosomes, through various core histone post-translational modifications and incorporation of diverse histone variants, can serve as epigenetic markers to control processes such as gene expression and recombination. The Histone Sequence Database is a curated collection, assembled from major public databases, of sequences and structures of histones and non-histone proteins containing histone folds. A substantial increase in the number of sequences and taxonomic coverage for histone and histone fold-containing proteins is available. The database also provides comprehensive multiple sequence alignments for each of the four core histones (H2A, H2B, H3 and H4), the linker histones (H1/H5) and the archaeal histones. Also included is current information on solved histone fold-containing structures. The database is thus an inclusive resource for the analysis of chromatin structure and function.

Chromatin.us is another web portal that includes information on chromatin proteins, histones and nucleosome structures and non-histone chromatin protein structures, and provides links to the protein data bank (PDB) site, which provides further

details.²⁹ ReplicationDomain³⁰ is an online database for storing, sharing and visualizing DNA replication timing and transcription data, along with other numerical epigenetic data types. Data are typically obtained from DNA microarrays or DNA sequencing.

Gene silencing. An important epigenetic phenomenon, gene silencing, has also attracted attention and has been well reported in the literature. Collected papers are available on Bio-Tech Info-Net.³¹ Similarly, RNA induced epigenetics related papers on imprinting by non-coding RNAs are collated.³²

Other epigenetic biomedical resources. The evolution of epigenetic resources is still in its early stages, with provision associated with several specific research efforts and groups. Nevertheless, in line with genetic/genomic data examples, efforts are being made to connect information, even as new targets are emerging. The Epigenetics Database³³ includes all known epigenetics genes/proteins discovered to date. The database is arranged in hierarchical format, based upon gene ontology. While still in its developmental (β) phase, it is expected that future developments will include user-submitted meta-data, which will be freely available for use in database and flat file format. Some sites, e.g., Epigenie,³⁴ also provide bioinformatics tools (e.g., CpG Viewer, CpG and GC Plotter and tools for CpG Island detection). NCBI supported efforts include the Epigenetics Antibody Database,³⁵ providing antibody information for researchers working in the field of epigenetics/epigenomics, and Unigene,³⁶ containing same locus-of-origin transcription sequences, protein similarities, gene expression, cDNA clone reagents, genomic location and associated epigenetic information. NARNA,³⁷ supported by Newcastle University, incorporates relationships between epigenetic events, DNA methylation, gene imprinting and X-chromosome inactivation with natural antisense RNAs. Other, locally developed or supported, current resources include StatEpigen,³⁸ with an initial focus on colon cancer, although incorporating some information on other pathologies for comparison. Data are provided on simple and conditional molecular events, since many genetic and epigenetic alterations are expected to be mutually correlated and synergistic, and drive model input at the micro-layer.³⁹ Specialized resources also exist for plant data.⁴⁰

Large-Scale Epigenetic Project Initiatives

European project initiatives including HEP. A number of European initiatives exist for centralized projects on DNA methylation. The Human Epigenome Project (HEP⁹) will provide an epigenetic resource of chromosomal DNA methylation reference profiles in human tissues and cell lines. Other initiatives include chromatin profiling (HEROIC, High-Throughput Epigenetic Regulatory Organization In Chromatin), treatment of neoplastic disease (EPITRON, Epigenetic Treatment Of Neoplastic Disease⁴¹) and the SMARTER⁴² initiative, which aims to develop small inhibitors of chromatin-modifying enzymes. Another effort to provide structure to the epigenetic research landscape in Europe is that of the Epigenetic Network of Excellence, now known as Epigenesys, which aims to advance epigenetics toward Systems Biology.⁴³

Roadmap epigenomics program. The Roadmap Epigenomics Program (also known as Epigenomics Roadmap initiative),

launched by NIH (2008), seeks to create a series of epigenome maps to study epigenetic mechanisms, develop new epigenetic analytics, generate a repository and long-term data archive, standardize procedures and practices in epigenomics and support new technologies for these. As part of the \$190 million, five-year initiative, the Roadmap Epigenomics Mapping Consortium⁴⁴ was formed to provide a public database for human epigenomics data, the Human Epigenome Atlas.⁴⁵ The current release, Epigenome Atlas Release 7, includes human reference epigenomes and the results of their integrative and comparative analyses.

The NIH Roadmap Epigenomics Program has also established IHEC (International Human Epigenome Consortium),⁸ which aims to coordinate epigenome mapping and characterization worldwide, in order to ensure high data quality standards, coordination of data storage, management and analysis and free access to the epigenomes produced. To attain substantial coverage of the human epigenome, IHEC aims to decipher at least 1,000 epigenomes within the next 7–10 years. Officially launched in Paris (Jan 2010), with an initial (first phase) budget target of \$130 million, IHEC intends to coordinate the mapping of epigenomes from not only the NIH's Epigenomics Mapping Consortium but also from international efforts such as the European Epigenome Network of Excellence, the Danish National Research Foundation Centre for Epigenetics, and the Australian Epigenetic Alliance. The IHEC web portal provides links to databases, such as GEO, ARRAYEXPRESS and DDBJ, where epigenetic sequencing data will be made available.

Another significant large-scale program in epigenetics is the Encyclopedia of DNA Elements (ENCODE).⁴⁶ This is supported by the ENCODE Consortium, an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). This initiative aims to identify all functional elements, both at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active, in the human genome sequence.

ICGC. Genomic changes that occur in various types of cancer are being investigated by the International Cancer Genome Consortium (ICGC).⁴⁷ The goal is to obtain a comprehensive description of genomic, transcriptomic and epigenomic changes in 50 different tumor types and/or subtypes. Many samples from

one tumor type or subtype will be analyzed in detail so that this initiative promises to provide crucial insights on genetic-epigenetic links.

Discussion and Conclusions

The biomedical resources relating, primarily, to epigenetic data that were surveyed here are numerous and range from small-to large-scale, with considerable ongoing integration and new links still being forged. In common with many newly identified research targets, early-stage resources are often very specific and are supported locally, and this is still the case for much useful epigenetic data. Many such databases and their software tools are publicly accessible from academic/research institutions, while others are commercially available (Table S1). Major issues remain quality assurance, effective annotation and overall management, but appropriate analysis must also keep pace and is typically uneven (Table S2). Clearly, the generation of a centralized repository for epigenetics-related data is desirable and currently lacking, but new technologies offer increased potential for processing solutions down the line. Notably, biomedical needs are an important focus for federated database development, health-grid technology and, of course, Cloud computing.

Major initiatives to ensure quality and standards for genetic and epigenetic research do exist and some, such as IHEC and HEP, are described in this review. With improved technology, these should lead to improved data mining tools where those currently available for epigenetic/epigenomic analyses are limited and predominantly sequence-oriented, ranging from identification, through PCR and initial pattern matching (Table S2 presents the current summary).

Acknowledgments

The authors would like to acknowledge funding from the Daniel O'Hare Scholarship program DCU, which made it possible to carry out this study.

Supplemental Materials

Supplemental materials may be found here: www.landesbioscience.com/journals/epigenetics/article/21493

References

- Choudhary AN, Honbo D, Kumar P, Ozisikylmaz B, Misra S, Memik G. Accelerating Data Mining Workloads: Current Approaches and Future Challenges in System Architecture Design. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2011; 1:41-54; <http://dx.doi.org/10.1002/widm.9>.
- Marakas GM. *Modern Data Warehousing, mining and Visualization Core Concepts*. Pearson Education, 2003.
- Sheth AP, Larson JA. Federated Database Systems for Managing Distributed, Heterogeneous and Autonomous Databases. *ACM Comput Surv* 1990; 22:183-236. <http://dx.doi.org/10.1145/96602.96604>.
- Heimbigner D, McLeod D. A Federated Architecture for Information Management. *ACM Transactions on Office Information Systems* 1985; 3:253-78.
- Entrez gene (Internet). 24 February 2012. <http://www.ncbi.nlm.nih.gov/Entrez/>
- Tenenbaum JD, Whetzel PL, Anderson K, Borromeo CD, Dinov ID, Gabriel D, et al. The Biomedical Resource Ontology (BRO) to enable resource discovery in clinical and translational research. *J Biomed Inform* 2011; 44:137-45; PMID:20955817; <http://dx.doi.org/10.1016/j.jbi.2010.10.003>.
- Weizmann Institute of Science (Internet). Genecards. 24 February 2012. <http://www.genecards.org/>
- International Human Epigenome Consortium (Internet). IHEC. 24 February 2012: <http://www.ihec-epigenomes.org/>
- Human Epigenome Project (Internet). HEP. 24 February 2012. <http://www.epigenome.org/>
- Novik KL, Nimrich I, Genc B, Maier S, Piepenbrock C, Olek A, et al. Epigenomics: genome-wide study of methylation phenomena. *Curr Issues Mol Biol* 2002; 4:111-28; PMID:12432963.
- Rodenhiser D, Mann M. Epigenetics and human disease: translating basic biology into clinical applications. *CMAJ* 2006; 174:341-8; <http://dx.doi.org/10.1503/cmaj.050774>; PMID:16446478.
- The Cancer Genome Atlas (Internet). TCGA. 21 May 2012. <http://cancergenome.nih.gov/>
- Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 2004; 429:457-63; PMID:15164071; <http://dx.doi.org/10.1038/nature02625>.
- Feinberg AP, Tycko B. The history of cancer epigenetics. *Nat Rev Cancer* 2004; 4:143-53; PMID:14732866; <http://dx.doi.org/10.1038/nrc1279>.
- Fuks F. DNA methylation and histone modifications: teaming up to silence genes. *Curr Opin Genet Dev* 2005; 15:490-5; PMID:16098738; <http://dx.doi.org/10.1016/j.gde.2005.08.002>.
- Pubmeth. Reviewed methylation database in cancer (Internet). 24 February 2012. <http://matrix.ugent.be/pubmeth/>
- Grunau C, Renault E, Rosenthal A, Roizes G. MethDB—a public database for DNA methylation data. *Nucleic Acids Res* 2001; 29:270-4; PMID:11125109; <http://dx.doi.org/10.1093/nar/29.1.270>.

18. Negre V, Grunau C. The MethDB DAS server: adding an epigenetic information layer to the human genome. *Epigenetics* 2006; 1:101-5; PMID:17965614; <http://dx.doi.org/10.4161/epi.1.2.2765>.
19. MethPrimerDB, the DNA methylation analysis PCR primer database (Internet). 24 February 2012. <http://medgen.ugent.be/methprimerdb/>
20. He XM, Chang SH, Zhang JJ, Zhao Q, Xiang H, Kusonmano K, et al. MethyCancer: the database of human DNA methylation and cancer. *Nucleic Acids Res* 2008; 36(Database issue):D836-41; <http://dx.doi.org/10.1093/nar/gkm730>; PMID:17890243.
21. Epigenetics Group (Internet). Dr. Axel Schumacher. 24 February 2012. <http://www.methylogix.com/genetics/database.shtml.htm>
22. Methtools (Internet). 27 February 2012. <http://genome.imb-jena.de/methtools/>
23. Reik W, Walter J. Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2001; 2:21-32; PMID:11253064; <http://dx.doi.org/10.1038/35047554>.
24. Gene imprint, Imprinted Gene Database (Internet). 24 February 2012. <http://www.geneimprint.com/site/genes-by-species>
25. University of Otago (Internet). New Zealand, Catalogue of Parent of Origin Effects, Parental Origins of de novo Mutations. 24 February 2012. <http://igc.otago.ac.nz/home.html>
26. Glaser RL, Ramsay JP, Morison IM. The imprinted gene and parent-of-origin effect database now includes parental origin of de novo mutations. *Nucleic Acids Res* 2006; 34(Database issue):D29-31; <http://dx.doi.org/10.1093/nar/gkj101>; PMID:16381868.
27. Mousebook (Internet). 24 February 2012. <http://www.mousebook.org/about.php>
28. National Human Genome Research Institute (Internet). Histone Database. 24 February 2012. <http://genome.nih.gov/histones/complete.shtml>.
29. Chromatin, Chromatin Structure and Function Page (Internet). 24 February 2012. <http://www.chromatin.us/chrom.html>
30. Florida State University (Internet). Replication Domain. 24 February 2012. <http://www.replicationdomain.com/>
31. Gene Silencing (Internet). Technical papers. 24 February 2012. http://www.biotech-info.net/gene_silencing.html#silencing
32. RNA Induced Epigenetics (Internet). 24 February 2012. <http://www.euchromatin.com/RNAepi.htm>
33. The Epigenetics Database (Internet). Beta. 24 February 2012. <http://www.epidna.com/database.php>
34. The EpiGenie (Internet). EpiGenie Epigenetics Product Database. 24 February 2012. <http://www.epigenie.com/Epigenetics-Research-Products.html>
35. National Center for Biotechnology Information (Internet). NCBI. 24 February 2012. http://www.ncbi.nlm.nih.gov/guide/all/#databases_
36. Unigene, NCBI (Internet). 24 February 2012. <http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene>
37. Platform for research on natural antisense transcripts, NARNA (Internet). 24 February 2012. <http://www.narna.ncl.ac.uk>
38. Statepigen, Database of colon cancer epigenetic statistics (Internet). 24 February 2012. Available from: <http://statepigen.sci-sym.dcu.ie>
39. Raghavan K, Ruskin HJ, Perrin D, Goasmat F, Burns J. Computational micromodel for epigenetic mechanisms. *PLoS One* 2010; 5:e14031; <http://dx.doi.org/10.1371/journal.pone.0014031>; PMID:21152421.
40. Epigenomics of Plants International Consortium (Internet). EPIC. 24 February 2012. <https://www.plant-epigenome.org/groups>
41. EPIgenetic Treatment of Neoplastic disease, EPITRON (Internet). 24 June 2012. <http://www.epitron.eu/>
42. SMARTER Network (Internet). 24 June 2012. <http://www.smarter-chromatin.eu/>
43. EpiGenesys (Internet). 27 February 2012. <http://www.epigenesys.eu/>
44. NIH Roadmap Epigenomics Mapping Consortium (Internet). 24 February 2012. <http://www.roadmapepigenomics.org>
45. Human Epigenome Atlas (Internet). 24 June 2012. <http://www.genboree.org/epigenomeatlas/index.rhtml>
46. Encyclopedia of DNA Elements (Internet). 24 February 2012. <http://genome.ucsc.edu/ENCODE/>
47. ICGC Cancer Genome Projects (Internet). 24 June 2012. <http://www.icgc.org/>