

MPromDb: an integrated resource for annotation and visualization of mammalian gene promoters and ChIP-chip experimental data

Hao Sun, Saranyan K. Palaniswamy, Twyla T. Pohar, Victor X. Jin,
Tim H.-M. Huang and Ramana V. Davuluri*

Department of Molecular Virology, Immunology and Medical Genetics, Human Cancer Genetics Program, Comprehensive Cancer Center, The Ohio State University, 420 W 12th Avenue, TMRF 524, Columbus, OH 43210, USA

Received August 15, 2005; Revised and Accepted October 16, 2005

ABSTRACT

We have developed Mammalian Promoter Database (MPromDb), a novel database that integrates gene promoters with experimentally supported annotation of transcription start sites, *cis*-regulatory elements, CpG islands and chromatin immunoprecipitation microarray (ChIP-chip) experimental results with intuitively designed presentation. Release 1.0 of MPromDb currently contains 36 407 promoters and first exons (19 170 from human, 15 953 from mouse and 1284 from rat), 3739 transcription factor (TF)-binding sites (2027 from human, 1181 mouse and 531 rat) and 224 TFs with links to PubMed and GenBank references. Target promoters of TFs that have been identified by ChIP-chip assay are integrated into the database. MPromDb serves as a portal for genome-wide promoter analysis of data generated by ChIP-chip experimental studies. MPromDb can be accessed from <http://bioinformatics.med.ohio-state.edu/MPromDb>.

INTRODUCTION

Promoters located at the 5' ends of genes play a critical role in regulating transcription initiation. The basal transcriptional machinery of RNA polymerase II assembles at the core promoter, which is a minimum stretch of DNA sequence from –35 to +35 bp of the transcription start site (TSS), that is sufficient to direct transcription initiation (1). The proximal promoter region (upstream of the core-promoter region) contains the *cis*-regulatory elements of most of the transcription factors (TFs). Enhancers and silencers are located several kbp upstream of the TSS. Extensive molecular research has provided a wealth of information about experimentally

characterized proximal promoter sequences, TFs and their binding sites. This information is dispersed throughout various databases, including GenBank (2,3), PubMed (4), TRANSFAC (5) and DBTSS (6,7). The integration of such essential information with the human (8,9) and rodent (10) genome sequences is one of the major challenges of the post-genome era.

Novel high-throughput technologies, such as chromatin-immunoprecipitation followed by microarray analysis (ChIP-chip), have enabled genome-wide identification of the epigenetic mechanisms and protein–DNA interactions that affect gene expression (11). In recent years we (12–15) and others (16) have successfully used ChIP-chip assays to find the target genes of TFs in mammalian systems. A large-scale effort to map the sequence information on the microarrays to the corresponding gene promoters and integration of the ChIP-chip experimental results into a database would significantly help future studies focused on the modelling of mammalian transcriptional regulatory networks (17).

Although similar promoter databases, such as Eukaryotic Promoter database (EPD) (18), DBTSS (7) and TRED (19), have been available to the research community, these databases have their own limitations. For example, EPD and DBTSS only contain the annotation of the TSSs. Although TRED contains more than 50K promoters and seems more comprehensive, it has many putative promoters obtained by computational prediction. Computational programs, such as FirstEF (20), can predict CpG-related promoters and first exons with high accuracy but perform rather poorly in predicting non-CpG-related first exons and promoters. In view of these shortcomings, we have only considered experimentally supported data from GenBank and other primary data sources in building Mammalian Promoter Database (MPromDb). Here, we present a database of mammalian promoters that are experimentally supported with experimentally known TF-binding sites and ChIP-chip data. The data were obtained by a computational pipeline followed by manual curation to

*To whom correspondence should be addressed. Tel: +1 614 688 3088; Fax: +1 614 688 4006; Email: Davuluri-1@medctr.osu.edu

ensure a high level of quality. The overall goal of MPromDb is to integrate the ChIP-chip experimental results generated by our laboratory and others and to serve as a portal for promoter analysis of ChIP-chip experimental data.

DATA ACQUISITION AND INTEGRATION

Mapping experimentally supported promoter and first exon sequences to the genome

The position relative to the TSS is the central part of promoter annotation and is a key element to the determination of the promoter region. Since experimentally verified promoter sequences in GenBank are limited, we used an indirect approach to map the first exons and promoters to the genome. We consider promoters and first exons obtained by mapping full-length 5'-untranslated regions (5'-UTRs)/mRNAs and previously characterized promoter sequences to the genome as experimentally supported. We collected human and mouse promoters from DBTSS (7) and EPD (21) and full-length 5'-UTR sequences from the 5'-UTR database of Davuluri *et al.* (22). We then searched GenBank using composite queries including: ('*homo sapiens*'[ORGN] AND ('5'UTR'[FKEY] OR promoter[FKEY] OR exon[FKEY] OR mRNA[FKEY] OR prim_transcript[FKEY] NOT htgs[KEYWORD] NOT htg[KEYWORD])) to retrieve experimentally derived first exons, promoters, and full-length 5'-UTR sequences. We parsed these GenBank records for first exons, full-length mRNAs, full-length 5'-UTRs, and promoter sequences that are supported by experimental evidence. The Perl script scans each GenBank nucleotide record for mRNA, exon, 5'-UTR, prim_transcript, promoter and CDS annotations. If a feature is annotated as incomplete at the left end (e.g. mRNA: <1..250), or is annotated as 'putative' or 'evidence = not experimental', the record is ignored. The script also ignores the records that have identical start sites for both the mRNA (or first exon) and the CDS. We mapped all of the resultant first exons, full-length 5'-UTR/mRNA, and promoter sequences to the corresponding genome sequences by BLAT (23). The first exons that completely overlap with the internal exons at the 5' end (as determined by using alignments of mRNAs and expressed sequence tags to the genomic sequence) were removed from the dataset, since these first exons might be the result of spurious TSS annotations in GenBank and annotations of incomplete 5'-UTR/mRNA records as 5' complete. Although some of these deleted records might be real promoters, we have adopted these stringent criteria in order to minimize the errors in annotation. We then prepared a non-redundant set of mammalian promoters. Two tandem first exons from the same gene are considered redundant if they are either overlapping or the distance between their TSSs is <500 bp. We also identify bi-directional promoters of gene-pairs that are located in head-to-head orientation that are separated by an inter-genic region of length <500 bp. We then retrieved the sequence from -2K upstream to the +1K downstream of the first exon of each gene.

We regularly run the computational pipeline (once in 3 months) to query the PubMed, GenBank and other databases for retrieving the new nucleotide sequence records that contain information about experimentally validated promoters and TF-binding sites. The current version of MPromDb contains

19 170 promoters of 15 616 human genes, 15 953 promoters of 14 157 mouse genes and 1284 promoters of 1234 rat genes. Of the human and of mouse genes 22% (3554) and 12% (1796) have two or more alternative promoters. Although the promoter annotations relating to rat genes are rather limited, we expect this to increase substantially after performing the comparative genomic analysis orthologous genes in the future updates.

We calculated the distance between the TSS and translation start site of the corresponding transcript, using the coding sequence annotations of the Consensus CDS (CCDS) project (<http://www.ncbi.nlm.nih.gov/CCDS>) of NCBI, for the records that are common between MPromDb and the CCDS database. Supplementary Figure 1 shows the histogram of the length distribution. We found that the TSS of ~30% of genes in MPromDb start >1 kb upstream of the translation start site, which suggests that at least 30% of the human genes have introns in between the coding and 5'-untranslated first exons. We earlier found that ~40% of the human genes have completely non-coding first exons, based on a smaller set of first exon of 2139 genes (20).

Mapping experimentally supported *cis*-regulatory elements to the corresponding gene promoters

Searching for *cis*-regulatory elements in published papers through PubMed and collecting relevant information by mining data from each of the papers is not a trivial task, hence the initial search of PubMed and screening of the literature is done by a set of Perl scripts. This is followed by manual inspection. We also parse experimentally characterized *cis*-regulatory elements described in GenBank records. Since the entries of *cis*-regulatory elements in GenBank records are not consistent, implementing a generic parser to parse these data is a rather difficult exercise. Further, the TF-binding sites are too short to use any of the sequence alignment tools or regular expressions to map these sequences on genomes. In order to overcome these challenges, we developed a set of comprehensive parsers to parse the data from different GenBank records. During the parsing stage, the TF-binding sequences are extended by 100 bp on either side, we then use BLAT (23) to map these sequences to the genomic sequence. This approach results in a precise genomic mapping of the *cis*-regulatory elements to corresponding gene promoters.

The current version of MPromDb contains annotation of 3739 TF-binding sites (2027 from human, 1181 mouse and 531 rat) corresponding to 224 TFs with links to PubMed and GenBank references. We plotted the location of the binding sites in MPromDb relative to the corresponding TSS. Supplementary Figure 2 shows the location of the annotated binding sites relative to TSS in MPromDb. We note that ~65% of the annotated sites fall within -500 bp of the TSS.

Annotating ChIP-chip microarray data with the corresponding promoters

There are two major types of commercially available microarray platforms for conducting ChIP-chip experiments. The first one is the CpG Island (CGI) microarray (Sanger 12k ChIP) available from the UHN Microarray Center (<http://www.microarray.ca/>) (24) and the Agilent promoter array, which was initially developed by Rick Young's group at MIT (16,25). The Sanger 12k ChIP consists of 12 192 clones

constructed from various CGI arrays (26), with a median length of 500 bp for clones. The rationale for such a design was motivated by the finding that ~60% of all human genes are associated with a CGI, usually at the 5' end (27), and ~85% of CGIs have been determined to be within -500 to +1500 bp of a TSS (28). This array has been successfully used in our earlier studies (12,13). The initial Agilent promoter microarray that was used in mapping HNF TF-binding sites (16) consists of 68 704 oligonucleotide probes representing the upstream regions of 17 054 NCBI RefSeq mRNAs (29), where each gene promoter is associated with four clones with an average sequence length of 60 bp for each clone. We have mapped the sequences of the CGI clones of the CGI array and 60mer probes of the Agilent array to the corresponding promoters by BLAT (23). The experimental results of different ChIP-chip studies, such as target promoter lists of a specific TF and corresponding binding sites, were manually collected from the published articles and integrated into the MPromDb database. The current version of MPromDb has binding site annotations for the following TFs mapped by ChIP-chip studies E2F1 (15), ER α (30), Myc (12), NF- κ B (31) and E2F4 (25) (Supplementary Tables 1 and 2). In a recent study, Young and coworkers (25) have used 10-slide promoter arrays designed by Agilent technologies to map the transcriptional regulatory circuitry in human embryonic stem cells. These arrays are the latest promoter arrays that contain ~400 000 features (60mers) that span -8 kb upstream to +2 kb downstream of annotated TSSs of 17 917 genes. These features and the binding sites of the OCT4, SOX2 and NANOG TFs (25) will be incorporated in MPromDb in the near future.

DATABASE ORGANIZATION

The promoter and *cis*-regulatory sequences, corresponding attributes and annotation data are stored in a relational database. MPromDb is structured as entity relationship model (Supplementary Figure 3). The mammalian promoter information stored in the 'PromoterInfo' table represents the backbone of the database. The 'BindingSiteInfo' table stores the TF (protein)-binding site annotations, including binding sequence, genomic position and corresponding TF name. The 'GeneInfo' table stores the gene annotation data for each promoter. It uses UniGene information to annotate the promoter and contains the UniGene identifier, gene symbol and gene description. 'PromoterSeq' stores the promoter sequence itself. The 'Reference' information associated with *cis*-regulatory elements in the database such as PubMed identifiers that can be used as links to the information is stored as well. 'CpGScore' table contains the CpGscore of the all the promoters in MPromDb.

JBoss (<http://www.jboss.org>) is used as the HTTP application server with MySQL as the database server (Supplementary Figure 4). It runs on Red Hat Linux Enterprise Edition 9.0. The data acquisition pipeline for promoter annotation was written in Perl and the web interface was written in Java.

DATA ACCESS AND VISUALIZATION

MPromDb may be accessed at <http://bioinformatics.med.ohio-state.edu/MPromDb>. Users may search the database and retrieve the promoter sequence and associated annotation

information of a specified gene in several ways. For example, a user may obtain the promoter of a gene by searching with Gene Name or Symbol, LocusLink identifier, UniGene identifier or GenBank accession identifier. Alternatively, a user may obtain TF information, including binding site position, binding sequence and promoter annotation of target gene, by simply searching with the TF name. We provide not only textual but also a 2D image as a gene browser to present the search results. Currently, the available search options for species include human, mouse and rat. Users may obtain promoter and gene annotation information by querying MPromDb in following ways.

- (i) Quick search option enables users to query promoters of a gene by gene name or GenBank ID or UniGene ID or Gene ID (Figure 1).
- (ii) Advanced search option enables users to query the database for list of promoters with composite queries such as species, chromosomes, promoters regulated by TFs and users can locate promoters annotated with CpG Island microarray and Agilent promoter microarray probes by choosing from drop down list box (Figure 1). Choosing the later option would list all the promoters that are found in the above mentioned array types.
- (iii) The gene search result page contains gene name, species, symbol and function as important gene annotation information. The Web Links entry to UCSC Genome Browser, NCBI Gene Project, Stanford Source and Gene Card, of the query result provides a cross-reference annotation of that gene (Figure 2a).
- (iv) User can click on Gene ID, to access promoter annotation information page. The result contains a table presentation of all the TSSs (including alternative TSS) with their chromosome, strand and genomic position information. There are also entries, such as 'promoter sequence' and link to GenBank. Promoter sequences can be retrieved with option to pick the length of the promoter both upstream and downstream (Figure 2c).
- (v) The visual module in promoter annotations page provides depiction of the promoter with corresponding TSS, binding elements, other TFs relative to the TSS and CpG Island information. We have implemented an in house developed JAVA™ application framework called the Genome Data Visualization Tool Kit for MPromDb information presentation in the form of an image map of gene regulatory regions with interactive contextual menus for easy navigation (32). The mouse-over option is a user-friendly feature that contains the respective TF name with genomic position and binding site motif (Figure 2b).
- (vi) If a specific promoter is located in CpG Island array or Agilent array, it is displayed as an additional track with probe sequences being displayed as rectangle boxes (Figure 2b).
- (vii) The CpG score is depicted as a histogram, where the line of score 6.5 is used to represent the cut-off value to determine whether the promoter is CpG or non-CpG related.
- (viii) The textual data provides information in static form, which includes the TF name, its positions and sequence, and respective binding site reference, with link to PubMed (4) and GenBank (3).

Quick Search: [Help](#)

You can quickly locate the gene of interest by entering one of the following keywords

Species (e.g. Human, Mouse, Rat)

Keyword Type

Keyword

Advanced Search

You can get a list of genes by combining the following conditions.

1. Select species (e.g. Human, Mouse, Rat)

2. Select chromosome number:

3. Regulated by Transcription Factor (e.g. TATA, AP1)

You can also quickly locate the promoters annotated with CpGIsland Microarray and Agilent Promoter Microarray probes

1. Select chromosome number:

2. Microarray Platforms

Search Results:

- alpha-INF2
 - alpha-Pal
 - AML1/Runx1
 - ANF
 - AP-1
 - AP-2
 - AP-2A
 - AP-2-CS3
 - AP-2-CS4
 - AP-3
 - AP-4
 - AR
 - ARE
 - ATF
 - bHLH
 - C2ORF3
 - CAAT_signal
 - calcium
 - CALM1
 - CBF
 - CCTF
 - CDX1
 - CEBPZ
 - CF1
 - CFT
 - cJun
 - c-Myb
 - c-Myc
 - COUP-TF1/Ear3
 - All
- All
 - Chromosome 1
 - Chromosome 2
 - Chromosome 3
 - Chromosome 4
 - Chromosome 5
 - Chromosome 6
 - Chromosome 7
 - Chromosome 8
 - Chromosome 9
 - Chromosome 10
 - Chromosome 11
 - Chromosome 12
 - Chromosome 13
 - Chromosome 14
 - Chromosome 15
 - Chromosome 16
 - Chromosome 17
 - Chromosome 18
 - Chromosome 19
 - Chromosome 20
 - Chromosome 21
 - Chromosome 22
 - Chromosome 23
 - Chromosome X
 - Chromosome Y
- All
 - Homo Sapiens
 - Mus musculus
 - Rattus norvegicus

Figure 1. Search options for MPromDb. Users can query promoters of a gene by gene name or GenBank ID or UniGene ID or Gene ID. Advanced options enables user search for promoters by chromosome and by different array platforms (CpG Island and Agilent).

FUTURE DIRECTIONS

The long-term goal of this database is to contribute to the understanding of mammalian gene transcriptional regulation. As more data are published, we will continue to incorporate the annotations into the content of MPromDb. This database will provide the foundation for further promoter analysis such as developing novel algorithms for TF-binding site search and characterizing gene regulatory modules. Future plans

include the integration of more ChIP-chip data about TF target genes, regional histone acetylation and methylation modifications and CpG Island methylation patterns around gene promoters. In conclusion, MPromDb provides integrated transcriptional regulatory information with genomic context in an easily accessible way. We believe that its implementation will facilitate large-scale promoter analysis and contribute towards the elucidation of mammalian transcriptional regulatory networks.

Cancer Institute grant P50 CA-11300, Ohio Board of Reagent Grant (BRTT02-0003) and by funds from the Ohio State University Comprehensive Cancer Center-Arthur G. James Cancer Hospital and Richard J. Solove Research Institute. Funding to pay the Open Access publication charges for this article was provided by National Cancer Institute grant P50CA-11300.

Conflict of interest statement. None declared.

REFERENCES

- Butler, J.E. and Kadonaga, J.T. (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.*, **16**, 2583–2592.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A. and Wheeler, D.L. (2002) GenBank. *Nucleic Acids Res.*, **30**, 17–20.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Wheeler, D.L. (2004) GenBank: update. *Nucleic Acids Res.*, **32**, D23–D26.
- McEntyre, J. and Lipman, D. (2001) PubMed: bridging the information gap. *CMAJ.*, **164**, 1317–1319.
- Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Pruss, M., Reuter, I. and Schacherer, F. (2000) TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res.*, **28**, 316–319.
- Suzuki, Y., Yamashita, R., Nakai, K. and Sugano, S. (2002) DBTSS: DataBase of human Transcriptional Start Sites and full-length cDNAs. *Nucleic Acids Res.*, **30**, 328–331.
- Suzuki, Y., Yamashita, R., Sugano, S. and Nakai, K. (2004) DBTSS, DataBase of Transcriptional Start Sites: progress report 2004. *Nucleic Acids Res.*, **32**, D78–D81.
- Lander, E.S., Linton, L.M., Birren, B., Nussbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.
- Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
- van Steensel, B. (2005) Mapping of genetic and epigenetic regulatory networks using microarrays. *Nature Genet.*, **37** (Suppl), S18–S24.
- Mao, D.Y., Watson, J.D., Yan, P.S., Barsyte-Lovejoy, D., Khosravi, F., Wong, W.W., Farnham, P.J., Huang, T.H. and Penn, L.Z. (2003) Analysis of Myc bound loci identified by CpG island arrays shows that Max is essential for Myc-dependent repression. *Curr. Biol.*, **13**, 882–886.
- Yan, P.S., Shi, H., Rahmatpanah, F., Hsiao, T.H., Hsiao, A.H., Leu, Y.W., Liu, J.C. and Huang, T.H. (2003) Differential distribution of DNA methylation within the RASSF1A CpG island in breast cancer. *Cancer Res.*, **63**, 6178–6186.
- Wells, J., Yan, P.S., Cechvala, M., Huang, T. and Farnham, P.J. (2003) Identification of novel pRb binding sites using CpG microarrays suggests that E2F recruits pRb to specific genomic sites during S phase. *Oncogene*, **22**, 1445–1460.
- Weinmann, A.S., Yan, P.S., Oberley, M.J., Huang, T.H. and Farnham, P.J. (2002) Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev.*, **16**, 235–244.
- Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K. *et al.* (2004) Control of pancreas and liver gene expression by HNF transcription factors. *Science*, **303**, 1378–1381.
- Blais, A. and Dynlacht, B.D. (2005) Constructing transcriptional regulatory networks. *Genes Dev.*, **19**, 1499–1511.
- Schmid, C.D., Praz, V., Delorenzi, M., Perier, R. and Bucher, P. (2004) The Eukaryotic Promoter Database EPD: the impact of in silico primer extension. *Nucleic Acids Res.*, **32**, D82–D85.
- Zhao, F., Xuan, Z., Liu, L. and Zhang, M.Q. (2005) TRED: a Transcriptional Regulatory Element Database and a platform for in silico gene regulation studies. *Nucleic Acids Res.*, **33**, D103–D107.
- Davuluri, R.V., Grosse, I. and Zhang, M.Q. (2001) Computational identification of promoters and first exons in the human genome. *Nature Genet.*, **29**, 412–417.
- Praz, V., Perier, R., Bonnard, C. and Bucher, P. (2002) The Eukaryotic Promoter Database, EPD: new entry types and links to gene expression data. *Nucleic Acids Res.*, **30**, 322–324.
- Davuluri, R.V., Suzuki, Y., Sugano, S. and Zhang, M.Q. (2000) CART classification of human 5' UTR sequences. *Genome Res.*, **10**, 1807–1816.
- Kent, W.J. and Brumbaugh, H. (2002) BLAT—the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
- Heisler, L.E., Torti, D., Boutros, P.C., Watson, J., Chan, C., Winegarden, N., Takahashi, M., Yau, P., Huang, T.H., Farnham, P.J. *et al.* (2005) CpG Island microarray probe sequences derived from a physical library are representative of CpG Islands annotated on the human genome. *Nucleic Acids Res.*, **33**, 2952–2961.
- Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G. *et al.* (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, **122**, 947–956.
- Cross, S.H., Clark, V.H. and Bird, A.P. (1999) Isolation of CpG islands from large genomic clones. *Nucleic Acids Res.*, **27**, 2099–2107.
- Antequera, F. and Bird, A. (1993) Number of CpG islands and genes in human and mouse. *Proc. Natl Acad. Sci. USA*, **90**, 11995–11999.
- Ishikhes, I.P. and Zhang, M.Q. (2000) Large-scale human promoter mapping using CpG islands. *Nature Genet.*, **26**, 61–63.
- Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **33**, D501–D504.
- Jin, V.X., Leu, Y.W., Liyanarachchi, S., Sun, H., Fan, M., Nephew, K.P., Huang, T.H. and Davuluri, R.V. (2004) Identifying estrogen receptor alpha target genes using integrated computational genomics and chromatin immunoprecipitation microarray. *Nucleic Acids Res.*, **32**, 6627–6635.
- Testa, A., Donati, G., Yan, P., Romani, F., Huang, T.H., Vigano, M.A. and Mantovani, R. (2005) Chromatin immunoprecipitation (ChIP) on chip experiments uncover a widespread distribution of NF-Y binding CCAAT sites outside of core promoters. *J. Biol. Chem.*, **280**, 13606–13615.
- Sun, H. and Davuluri, R.V. (2004) Java-based application framework for visualization of gene regulatory region annotations. *Bioinformatics*, **20**, 727–734.