W560–W565 *Nucleic Acids Research, 2006, Vol. 34, Web Server issue*
doi:10.1093/nar/gkl158

NXSensor web tool for evaluating DNA for nucleosome exclusion sequences and accessibility to binding factors

Peter Luykx, Ivan V. Bajić[1] and Sawsan Khuri[2,*]

Department of Biology, University of Miami, Coral Gables, FL 33124, USA, [1]School of Engineering Science, Simon Fraser University, Burnaby, B.C. V5A 1S6, Canada and [2]The Dr. John T. Macdonald Foundation Center for Medical Genetics, University of Miami Miller School of Medicine, Miami, FL 33101 USA

Received February 10, 2006; Revised March 7, 2006; Accepted March 20, 2006

ABSTRACT

Nucleosomes, a basic structural unit of eukaryotic chromatin, play a significant role in regulating gene expression. We have developed a web tool based on DNA sequences known from empirical and theoretical studies to influence DNA bending and flexibility, and to exclude nucleosomes. NXSensor (available at http://www.sfu.ca/~ibajic/NXSensor/) finds nucleosome exclusion sequences, evaluates their length and spacing, and computes an 'accessibility score' giving the proportion of base pairs likely to be nucleosome-free. Application of NXSensor to the promoter regions of housekeeping (HK) genes and those of tissue-specific (TS) genes revealed a significant difference between the two classes of gene, the former being significantly more open, on average, particularly near transcription start sites (TSSs). NXSensor should be a useful tool in assessing the likelihood of nucleosome formation in regions involved in gene regulation and other aspects of chromatin function.

INTRODUCTION

Transcription factors, enhancer-binding proteins and chromatin-remodeling factors all play a role in regulating gene expression. In addition, nucleosomes, the basic structural units of chromatin, are now thought to also be involved in this process [(1,2); see also (3,4)]. The dissociation or displacement of nucleosomes from DNA by the action of other protein factors, such as DNA- and histone-modifying enzymes (methylases, kinases, acetylases and deacetylases) likely gives the basal transcription apparatus access to the promoter regions of genes (5–7). A means for assessing the distribution of nucleosomes along stretches of regulatory DNA would therefore be a useful addition to our understanding of gene regulation.

In eukaryotic cells, nucleosomes are formed by the binding of DNA to histones (8). The nucleosome consists of 147 bp of DNA wound around a histone-octamer core (9–11). In human DNA, adjacent nucleosomes are separated on average by about 50 bp of 'linker' DNA, so that successive nucleosomes occur about every 200 bp. Most DNA sequences are neutral or favorable to nucleosome formation, but some sequences, by virtue of their influence on the curvature and flexibility of the DNA double helix (12,13), are unfavorable to nucleosome formation; we have called these 'nucleosome exclusion sequences' (NXSs). Thus, nucleosomes are not uniformly or randomly spaced along DNA, allowing for variation in access to promoter and other regulatory regions by the basal transcription apparatus, transcription factors and other proteins.

We have developed a program, NXSensor (available at http://www.sfu.ca/~ibajic/NXSensor/), that locates NXSs in DNA and defines regions where nucleosomes are unlikely to be formed because of the presence of such sequences and their spacing at distances less than the number of base pairs in the nucleosome. We have applied NXSensor to sets of promoter sequences whose genes are known to differ in their expression patterns among different tissues, and were able to show, in agreement with recent reports, significant differences among promoter regions in the number and position of such sequences.

BACKGROUND AND DESCRIPTION OF NXSensor

DNA sequences chosen as NXSs were the following, based on experimental observations (14–16):

$[(G/C)_3N_2]_{\geq 3}$; e.g. GGCAACGCTTGGGTA,
GGCCGCGCAGGGGCT

$A_{\geq 10} (= T_{\geq 10})$; e.g. AAAAAAAAAA, TTTTTTTTTT

Other studies lend support to the unfavorability of these sequences for nucleosome formation because

*To whom correspondence should be addressed. Tel: +1 305 243 6069; Fax: +1 305 243 3919; Email: skhuri@med.miami.edu

© The Author 2006. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

of their hindering of DNA bending or flexibility (12,13,17).

A NXS was defined as one of the DNA sequences above, a contiguous non-overlapping DNA sequence long enough to comprise one full turn of the DNA double helix. The NXSensor program annotates individual DNA regions by marking NXSs and then examining their spacing to find sequences between NXSs that are <147 bp in length, on the assumption that nucleosomes are unlikely to form in such regions. The remaining segments of DNA are those where nucleosomes may be located.

VALIDATION AND APPLICATION

To test the general validity of the NXSs chosen, and to test the ability of the NXSensor program to find and annotate such sequences, the DNA sequences of positioned nucleosomes listed in the nucleosome positioning region database (NPRD) were examined (18). Ideally, in these nucleosome-associated DNA sequences there should be no NXSs as defined here. We found that only nine (8%) of the 112 positioned nucleosomes contained any NXSs. These sequences (a total of 12 NXSs) comprised 199 bp in a total of 16 829 nucleosome-associated base pairs (1.18%), which represents reasonable agreement with expectation. Exceptions in which a NXS is accomodated within a nucleosome may be the result of the presence of additional protein factors associated with nucleosomal DNA.

As a further test of NXSensor, the SV40 viral sequence was examined because it is known to contain a nucleosome-free segment of about 400–500 bp within its 5243 bp genome (19–21). The results showed a close correspondence between the experimentally-determined location of the nucleosome-free region and the nucleosome-free region predicted by NXSensor's analysis of DNA sequence (Supplementary Figure 1).

In a final comprehensive test, NXSensor was used to analyze the promoter regions of two sets of human genes, housekeeping (HK) and tissue-specific (TS) genes. These two sets of genes were chosen in order to assess the potential distribution of nucleosomes in promoters of genes that are used differently in different tissues, to substantiate the conclusions of other researchers who have recently investigated this question using different approaches, and to demonstrate further the usefulness of the NXSensor tool.

One hundred genes from each set, HK and TS, were selected on the basis of the tissue expression patterns given in the Novartis Research Foundation's Genomic Institute 'SymAtlas' (http://symatlas.gnf.org/) (22). The genes selected were at the two extremes of tissue expression patterns: HK genes showing significant expression in all 73 normal tissues of the SymAtlas, and TS genes showing, according to gene-specific non-cross-hybridizing probe sets (those with '_at' suffixes), significant expression in only one or two tissues. To avoid possible bias in selecting HK genes, we included as wide a variety of basic cell functions as possible, and in selecting TS genes, as wide a variety of cell- and tissue types as possible. Genes with only one region of transcription initiation were used instead of those with widely-spaced alternative transcription start sites (TSSs), as indicated in the UCSC Genome Browser (http://genome.ucsc.edu/), and double-checked in the Database of Transcriptional Start Sites (DBTSS, http://dbtss.hgc.jp/). The list of genes is available at http://www.sfu.ca/~ibajic/NXSensor/200genes.xls.

To examine the region around the TSS of each gene, we used the sequence from 1000 bases upstream (which would include the complete promoter of most genes) to 1000 bases downstream from the TSS (−1000 to +1000), and called this 2000 bp stretch the 'promoter region,' centered on the TSS. The TSS was either the predominant site in, or a site near the middle of, the cluster of oligo-capped cDNA clones shown in the DBTSS (23). In cases where no oligo-capped cDNA clones were available from the DBTSS (21 of 100 HK genes, 36 of 100 TS genes), the RefSeq (NCBI) annotated TSSs were used. All sequences tested were datamined from the UCSC Genome Browser.

The basic 'nucleosome segments' option of NXSensor annotates the submitted sequence by highlighting in gray the segments available for nucleosome binding, leaving open the segments unlikely to be bound to nucleosomes by virtue of the number and positions of NXSs (Figure 1 and Supplementary Figure 1). The annotated promoter region sequences were used to construct two basic measures. The 'NXScore' defined as the number of NXSs in a window of 147 bp at a given position within the 2000 base promoter region, estimates the likelihood that nucleosomes are excluded from promoter regions at different distances from the TSS. This measure is the obverse of the 'nucleosome formation potential' based on nucleosome binding sequences used by Levitsky *et al*. (1,24). The NXScore is averaged over all promoter regions of each class of genes, HK and TS, to generate a graph of average NXScore versus position, centered on a given window (Figure 2).

The second measure was the 'accessibility score,' measuring the overall accessibility of the 2000 base promoter region to protein factors. The accessibility score is defined as the proportion of base pairs in the region likely to be free of nucleosomes, according to the number and spacing of the NXSs. Accessibility scores can vary from 0.0 (all sequences potentially occupied by nucleosomes) to 1.0 (no sequences likely to be occupied by nucleosomes). In calculating the accessibility score ACC(10), short exclusion sequences of 10 bp or less flanked by two segments having the potential for nucleosome binding were not considered 'open' because of the unlikelihood that any protein would be able to bind effectively such a short sequence between two nucleosomes. Accessibility scores for two examples are given in Figure 1.

A direct comparison of the average number of NXSs, per window of 147 bases, in HK and TS promoter regions is shown in Figure 2. In both sets of promoter regions there were more NXSs nearer the TSS. There was a significant difference between HK-gene promoters and TS-gene promoters in the mean number of NXSs in the specific 147 bp window centered on the TSS (HK, 2.10 and TS, 0.82; $P < 0.01$, Kolmogorov–Smirnov test). Supplementary Figure 2 gives an illustrative example of 10 000 bases around the TSS of a HK gene, in which NXSensor analysis suggests that essentially only the TSS region is free of nucleosomes.

Figure 2 also shows that the mean number of NXSs in this window for the HK genes was significantly higher than that expected for random-sequence DNA with the same base composition as the HK-gene promoter regions (2.10 for HK

Figure 1. NXSensor output for two promoter regions, −1000 to +1000 relative to the major TSS, which is highlighted in green. Left column: HK gene RBPMS, RNA-binding protein; right column: TS gene F12, coagulation factor XII. Top, nucleosome-segment output: shaded regions show where nucleosomes are more likely to be located, unshaded regions are more likely to be nucleosome-free; bottom row, locations of NXSs.

genes, 0.53 for the corresponding random-sequence DNA, $P = 0.0014$). In contrast, the mean number of NXSs in this window for the TS genes was hardly different from the NXSs expected in random-sequence DNA having the TS promoter base composition (0.82 for TS genes, 0.22 for the corresponding random-sequence DNA, $P = 0.06$).

From NXSensor output we obtained accessibility scores, ACC(10), for the promoter regions in our dataset. As shown in Figure 3A, there was a significant difference between the distributions of accessibility scores of HK genes and of TS genes, the HK genes having significantly higher accessibility scores than the TS genes ($P \leq 0.01$, Kolmogorov–Smirnov test).

Further analysis was conducted using a modified accessibility score, ACC(50), which includes sequences at least 50 bp in length likely to be nucleosome-free. This is the length of sequence estimated to be occupied by the basal transcription apparatus (25,26). ACC(50) was examined for the DNA region from −110 to +60, the region in which the basal transcription apparatus would have to bind in order to initiate transcription at the majority of TSSs clustered around the modal or central TSS at +1 (27). The comparison of this TSS region (Figure 3B) showed that 58% of the HK genes, but only 22% of the TS genes, had an ACC(50) of 0.5 or higher. Of the HK genes, 45% had an ACC(50) of 1.0, while only 15% of the TS genes did. The overall distributions of HK gene and TS gene ACC(50) scores were significantly different by the Kolmogorov–Smirnov test ($P \leq 0.01$).

DISCUSSION

Work in recent years has shown that the nucleosomes are neither uniformly nor randomly positioned along the DNA double helix. In a number of instances the location of nucleosomes in control regions of DNA was found to be an important feature of gene regulation. The validity of this statement was recognized explicitly by Levitsky *et al.* (1), and is the basis for the nucleosome positioning region database (NPRD), (http://srs6.bionet.nsc.ru/srs6/) (18). The 'nucleosome formation potential' scores have recently been used to demonstrate differences in the intergenic regions (28) and more specifically in the promoter regions (29) of human HK and TS genes.

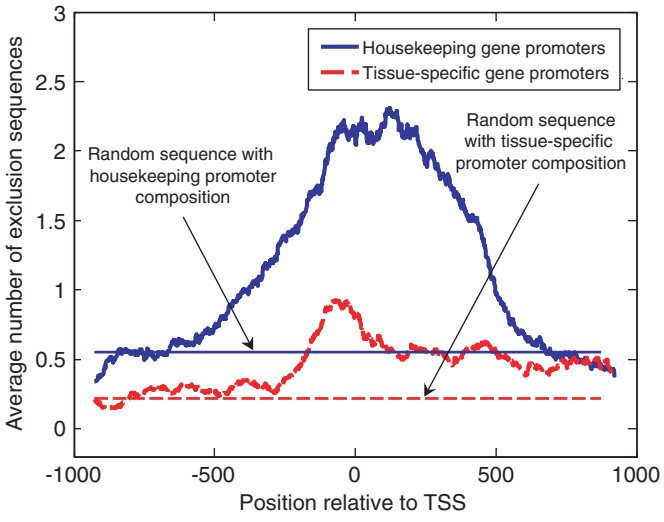

Figure 2. Average number of exclusion sequences per window of 147 bases in 100 HK (solid blue line) and 100 TS (dashed red line) promoter regions centered on the TSS. Also shown as thinner horizontal straight lines are the expected numbers of exclusion sequences in a random-sequence of the same base composition as that of the sum of the promoter regions in each class. The average number of observed exclusion sequences in a window centered at TSS is 2.10 for HK promoters, which is significantly higher ($P = 0.0014$) than the expected number of 0.53 in a random-sequence of the same composition. Similarly, for TS promoters, the average number of observed sequences in a window centered at TSS is 0.82, which is higher than the expected number of 0.22 in a random-sequence of the same composition, at the significance level of $P = 0.06$.

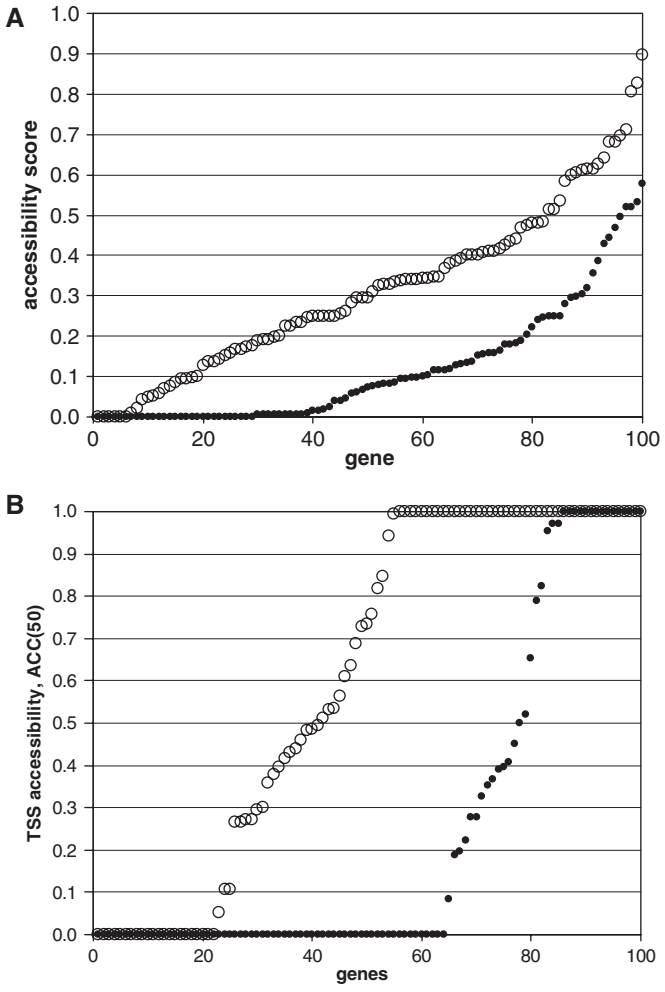

Figure 3. (**A**) Accessibility scores ACC(10) for promoter regions, −1000 to +1000 relative to the TSS. Open circles, HK-gene promoter regions; dots, TS-gene promoter regions. The gene promoter regions are ordered on the abscissa according to increasing accessibility score. (**B**) Accessibility scores ACC(50) in the vicinity of the TSS segment, −110 to +60 (accessibility to the basal transcription apparatus). Open circles, HK gene promoters; dots, TS-gene promoters. The gene promoters are ordered on the abscissa according to increasing accessibility score.

We have taken a different approach by utilizing a combination of published experimental and observational data to identify DNA sequences unfavorable for nucleosome formation because of their influence on DNA bending and flexibility. Other sequences not used here may also be unfavorable for nucleosome formation. One example is TGGA repeats (30), but these occurred too infrequently in our sample of 200 promoter regions to have any influence on the results.

Using our NXSensor program (http://www.sfu.ca/~ibajic/NXSensor/), we have shown that, as expected, these NXSs are rare in DNA sequences occupied by nucleosomes listed in the NPRD. NXSensor also accurately predicts the location of the nucleosome-free zone in viral SV40 DNA. In a set of HK and TS promoter region DNA sequences, NXSensor analysis has demonstrated that the promoter regions of HK genes are likely to be relatively nucleosome-free, a finding that may help to explain the wide tissue distribution of their expression. This is in contrast to those of TS genes, whose promoter regions contain sequences more favorable for nucleosome formation (28,29), and which are therefore likely to require additional TS transcription factors to modify or displace nucleosomes before the genes can be expressed.

The promoter regions of the two sets of genes differed in base composition and in how many of them had CpG islands (31,32). The GC content of the promoter regions of HK genes was 56.86%, compared with 50.44% for TS genes. CpG islands were characteristic of 92 of 100 HK genes, but only 19 of 100 TS genes, in agreement with the observations of others (32). The greater frequency of NXSs of the $[(G/C)_3N_2]_{\geq 3}$ type in HK-gene promoter regions compared with TS-gene promoter regions is partly a reflection of the higher GC content of the former (56.86% G + C in HK, 50.44% G + C in TS). However, the average number of $[(G/C)_3N_2]_3$ sequences in HK-gene promoter regions, 15.4, was even higher than the 7.8 ± 2*2.7 (S.D.) expected in random sequences of the same length and base composition ($P < 0.05$). In addition, in spite of their higher GC content, HK-gene promoter regions had almost as many nucleosome-unfriendly polyA and polyT tracts as did the TS-gene promoter regions (50 in HK, as compared with 52 in TS). These two observations show that the frequency of both types of NXS in HK-gene promoters is higher than would be expected from their base composition, suggesting a functional significance, possibly related to nucleosome exclusion.

Promoter regions of HK genes are typically GC-rich. Presumably this is the consequence of selection pressure for certain kinds of promoter sequences. It may be speculated that one component of such selection is for sequences that tend to exclude nucleosomes. Another possible component

is selection for GC-rich sequences that bind certain ubiquitous transcription factors, such as Sp-1. These two components of selection might be related, in that the factors that regulate expression of HK genes may have evolved to bind to nucleosome-free regions.

We were unable to find any correlation between accessibility score and expression levels of these genes as given in SymAtlas (22) (see Supplementary Figure 3). It is likely that the lower nucleosome occupancy of promoter regions of HK genes is related not to their expression levels but primarily to their ubiquity of expression in different tissues.

The NXSensor web tool is flexible enough to allow for different definitions of regions likely to be free of nucleosomes and more accessible to other protein factors. NXSensor can be set for more stringent exclusion criteria by increasing the number of tandem NXSs required for nucleosome exclusion. The minimum length of 'nucleosome-free' sequence can be increased to accommodate the space required for larger protein complexes, an approach we took to show that the region immediately surrounding TSSs of HK genes is likely to be more accessible to the basal transcription apparatus than is the corresponding region of TS genes.

Here we have applied NXSensor to the promoter regions of individual genes and classes of genes. NXSensor may also be used to investigate other control regions farther from coding sequences, such as enhancer and inhibitor regions, as well as sites of methylation, imprinting, recombination, repair, pre-mRNA splicing and indeed any DNA sequences where nucleosome location is likely to be a factor in overall chromatin organization and function (33–38).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

The authors thank Steven Green (Biology Department, University of Miami) for statistical advice, John Walker and Andrew Su (Genomics Institute of the Novartis Research Foundation) for advice on the interpretation of SymAtlas gene expression data, Terace Fletcher (Department of Biochemistry and Molecular Biology, University of Miami) for helpful discussions, and anonymous reviewers for many cogent comments. Part of the work performed while I.V.B. was with the Electrical and Computer Engineering Department, University of Miami, Coral Gables, Florida 33124, USA. Funding to pay the Open Access publication charges for this article was provided by the Dr. John T. Macdonald Foundation Center for Medical Genetics.

Conflict of interest statement. None declared.

REFERENCES

1. Levitsky,V.G., Podkolodnaya,O.A., Kolchanov,N.A. and Podkolodny,N.L. (2001) Nucleosome formation potential of eukaryotic DNA: calculation and promoters analysis. *Bioinformatics*, **17**, 998–1010.
2. Weaver,R.F. (2005) *Molecular Biology, 3rd edn.,* McGraw-Hill, Boston. Ch. 12, 13, pp. 342–423.
3. Gao,J. and Benyajati,C. (1998) Specific local histone-DNA sequence contacts facilitate high-affinity, non-cooperative nucleosome binding of both adf-1 and GAGA factor. *Nucleic Acids Res*., **26**, 5394–5401.
4. Vicent,G.P., Nacht,A.S., Smith,C.L., Peterson,C.L., Dimitrov,S. and Beato,M. (2004) DNA instructed displacement of histones H2A and H2B at an inducible promoter. *Mol. Cell*, **16**, 439–452.
5. Schrem,H., Klempnauer,J. and Borlak,J. (2002) Liver-enriched transcription factors in liver function and development. Part I: the hepatocyte nuclear factor network and liver-specific gene expression. *Pharmacol. Rev*., **54**, 129–158.
6. Morse,R.H. (2003) Getting into chromatin: how do transcription factors get past the histones? *Biochem. Cell. Biol*., **81**, 101–112.
7. Adkins,M.W. and Tyler,J.K. (2006) Transcriptional activators are dispensable for transcription in the absence of spt6-mediated chromatin reassembly of promoter regions. *Mol. Cell*, **21**, 405–416.
8. Chakravarthy,S., Park,Y.J., Chodaparambil,J., Edayathumangalam,R.S. and Luger,K. (2005) Structure and dynamic properties of nucleosome core particles. *FEBS Lett*., **579**, 895–898.
9. Luger,K., Mader,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) X-ray structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251–260.
10. Davey,C.A., Sargent,D.F., Luger,K., Maeder,A.W. and Richmond,T.J. (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J. Mol. Biol*., **319**, 1097–1113.
11. Richmond,T.J. and Davey,C.A. (2003) The structure of DNA in the nucleosome core. *Nature*, **423**, 145–150.
12. Drew,H.R., McCall,M.J. and Calladine,C.R. (1990) New approaches to DNA in the crystal and in solution. In Cozzarelli,N.R. and Wang,J.C. (eds), *DNA Topology and its Biological Effects*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Monograph 20, pp. 1–56.
13. Travers,A.A. and Klug,A. (1990) Bending of DNA in nucleoprotein complexes. In Cozzarelli,N.R. and Wang,J.C. (eds), *DNA Topology and its Biological Effects*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Monograph 20, pp. 57–106.
14. Wang,Y.H. and Griffith,J.D. (1996) The [(G/C)3NN]n motif: a common DNA repeat that excludes nucleosomes. *Proc. Natl Acad. Sci. USA*, **93**, 8863–8867.
15. Satchwell,S.C., Drew,H.R. and Travers,A.A. (1986) Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol*., **191**, 659–675.
16. Suter,B., Schnappauf,G. and Thoma,F. (2000) Poly(dA.dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters *in vivo*. *Nucleic Acids Res*., **28**, 4083–4089.
17. McConnell,K.J. and Beveridge,D.L. (2001) Molecular dynamics simulations of B′-DNA: sequence effects on A-tract-induced bending and flexibility. *J. Mol. Biol*., **314**, 23–40.
18. Levitsky,V.G., Katokhin,A.V., Podkolodnaya,O.A., Furman,D.P. and Kolchanov,N.A. (2005) NPRD: Nucleosome Positioning Region Database. *Nucleic Acids Res*., **33**, D67–D70.
19. Robinson,G.W. and Hallick,L.M. (1982) Mapping the *in vivo* arrangement of nucleosomes on simian virus 40 chromatin by the photoaddition of radioactive hydroxymethyltrimethylpsoralen. *J. Virol*., **41**, 78–87.
20. Choder,M., Bratosin,S. and Aloni,Y. (1984) A direct analysis of transcribed minichromosomes: all transcribed SV40 minichromosomes have a nuclease-hypersensitive region within a nucleosome-free domain. *EMBO J*., **3**, 2929–2936.
21. Batson,S.C., Rimsky,S., Sundseth,R. and Hansen,U. (1993) Association of nucleosome-free regions and basal transcription factors with *in vivo*-assembled chromatin templates active *in vitro*. *Nucleic Acids Res*., **21**, 3459–3468.
22. Su,A.I., Wiltshire,T., Batalov,S., Lapp,H., Ching,K.A., Block,D., Zhang,J., Soden,R., Hayakawa,M., Kreiman,G. *et al*. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.
23. 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.
24. Levitsky,V.G., Katokhin,A.V., Podkolodnaya,O.A. and Furman,D.P. (2004) Nucleosomal DNA organization: an integrated information system. In Kolchanov,N. and Hofestaedt,R. (eds), *Bioinformatics of*

Genome Regulation and Structure. Kluwer Academic Publishers, Boston, pp. 3–12.
25. Roeder,R.G. (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci*., **21**, 327–335.
26. Woychik,N.A. and Hampsey,M. (2002) The RNA polymerase II machinery: structure illuminates function. *Cell*, **108**, 453–463.
27. Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J., Hata,H., Ota,T., Isogai,T., Tanaka,T., Morishita,S. *et al*. (2001) Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites. *EMBO Rep*., **2**, 388–393.
28. Vinogradov,A.E. (2005) Noncoding DNA, isochores and gene expression: nucleosome formation potential. *Nucleic Acids Res*., **33**, 559–563.
29. Ganapathi,M., Srivastava,P., Kumar,S., Sutar,S.K.D., Kumar,K., Dasgupta,D., Singh,G.P., Brahmachari,V. and Brahmachari,S.K. (2005) Comparative analysis of chromatin landscape in regulatory regions of human housekeeping and tissue specific genes. *BMC Bioinformatics*, **6**, 126.
30. Cao,H., Widlund,H.R., Simonsson,T. and Kubista,M. (1998) TGGA repeats impair nucleosome formation. *J. Mol. Biol*., **281**, 253–260.
31. Takai,D. and Jones,P.A. (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl Acad. Sci. USA*, **99**, 3740–3745.
32. Antequera,F. (2003) Structure, function and evolution of CpG island promoters. *Cell. Mol. Life Sci*., **60**, 1647–1658.
33. Chen,C. and Yang,T.P. (2001) Nucleosomes are translationally positioned on the active allele and rotationally positioned on the inactive allele of the HPRT promoter. *Mol. Cell. Biol*., **21**, 7682–7695.
34. Baumann,M., Mamais,A., McBlane,F., Xiao,H. and Boyes,J. (2003) Regulation of V(D)J recombination by nucleosome positioning at recombination signal sequences. *EMBO J*., **22**, 5197–5207.
35. Powell,N.G., Ferreiro,J., Karabetsou,N., Mellor,J. and Waters,R. (2003) Transcription, nucleosome positioning and protein binding modulate nucleotide excision repair of the *Saccharomyces cerevisiae* MET17 promoter. *DNA Repair*, **2**, 375–386.
36. Davey,C., Fraser,R., Smolle,M., Simmen,M.W. and Allan,J. (2003) Nucleosome positioning signals in the DNA sequence of the human and mouse H19 imprinting control regions. *J. Mol. Biol*., **325**, 873–887.
37. Kogan,S. and Trifonov,E.N. (2005) Gene splice sites correlate with nucleosome positions. *Gene*, **352**, 57–62.
38. Pennings,S., Allan,J. and Davey,C.S. (2005) DNA methylation, nucleosome formation and positioning. *Brief Funct. Genomic Proteomic*., **3**, 351–361.