Transcription Factor Binding and Nucleosome Positioning Are Alternative Pathways for Transcription Start Site Selection in Eukaryotic Promoters

René Dreos[1,2], Giovanna Ambrosini[1,2] and Philipp Bucher[1,2]

[1] Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
rene.dreos@isb-sib.ch
[2] The Swiss Institute for Experimental Cancer Research (ISREC), School of Life Sciences, Swiss Federal Institute of Technology Lausanne (EPFL), 1015 Lausanne, Switzerland
{giovanna.ambrosini,philipp.bucher}@epfl.ch

Abstract. The DNA sequence determinants which direct RNA Pol-II to the correct transcription start site (TSS) are only partly understood. Conserved DNA motifs (core promoter elements) or a conserved nucleosome architecture may play a role in TSS selection. A complicating factor is that promoters are quite variable in many respects. Some have very focused while others have highly dispersed initiation site patters. Promoters also differ by the presence or absence of CPEs. Here we show that promoters without CPEs have a strong sequence–intrinsic nucleosome–positioning signal in the +1 nucleosome region, in both vertebrates and flies. The strength of the signal is inversely proportional to the degree of TSS dispersion. Interestingly, the nucleosome–positioning signal is completely absent in CPE containing promoters. Together, these findings suggest that transcription factor binding to CPEs and DNA sequence–induced nucleosome positioning are two mutually exclusive pathways of Pol-II recruitment to TSSs in eukaryotic promoters.

Keywords: transcription start site · nucleosome · promoters · TATA-box

1 Introduction

The recruitment of RNA Pol-II to specific sites in the genome called promoters is an essential step in eukaryotic gene regulation. To facilitate the interaction between Pol-II and DNA, promoter regions often contain conserved sequence motifs called core promoter elements (CPE). Over the years many classes of CPEs have been found in promoter regions [1] and their location relative to the transcription start site (TSS) is an important feature to differentiate them. CPEs are found at a fixed or nearly fixed distance from the TSS, in contrast to the so–called upstream promoter elements (UPEs) which occur at greater

and more variable distance from the TSS [2]. The best known CPEs are the TATA-box (located 25 bp upstream of the TSS) and Initiator (Inr, located at the TSS), both responsible for the correct identification of the start site by the Pol-II. Another CPE called downstream promoter element (DPE) occurs about 30 bp downstrean from the TSS [3]. For TATA-box containing promoters, the formation of an initiation–competent Pol-II complex starts with the binding of the general transcription factor TFIID to the TATA-box, followed by the recruitment of TFIIA and TFIIB [4]. This complex allows the Pol-II enzyme to interact with the DNA at a precise location and to start transcription 25 bp downstream of the TATA-box [5], often helped by the presence of the Inr motif at the TSS [6].

However, the majority of promoters in vertebrates and flies do not contain any TATA-box, Inr or DPE. It is not yet clear which molecular mechanism allows Pol-II to start transcription at specific locations in the absence of these CPEs. Chromatin may play a role in these promoters. Promoters generally have a conserved nucleosomal architecture consisting of a nucleosome–free region of around 150 bp near the TSS (enriched in CPEs and bound by the Pol-II complex) which is followed by a well–positioned nucleosome (often named +1) centered at a conserved distance of 120 bp downstream from the TSS [7–9]. In other respects, promoters are quite variable. High–throughput analysis of TSS usage has shown that some promoters have very focused initiation sites, nearly confined to a single base, while others have highly dispersed TSSs scattered over up to 100 bp [10, 11]. While the role of CPEs in recruiting Pol-II is well established, the role of chromatin in this process is poorly understood [12, 13]. Specifically, the timing and causal relationship between nucleosome binding/positioning and Pol-II binding remains unclear. *In vitro* experiments and whole genome analysis have shown that nucleosome positioning is at least partially encoded by the genome sequence [14, 15]. Sequences such as poly(dA:dT) tracts strongly disfavor nucleosome formation, whereas AA/TT dinucleotide periodicity of 10bp has high nucleosome affinity. This may be due to the intrinsic curvature given to the DNA by AA/TT dinucleotides that facilitate the wrapping of the DNA around the histone octamer, making them favored if they are located in the major groove of the double helix facing outward. Instead, GC–rich dinucleotides (CC/CG/GC/GG) are enriched in the major groove facing inward toward the histones [16]. DNA physical properties are not the only factors responsible of nucleosome positioning *in vivo* since nucleosomes can also be positioned by the interaction of DNA with non–histone DNA binding proteins [17–19]. For example, it has been shown by Fu and colleagues that the insulator protein CTCF is able to position up to 20 nucleosomes around its binding sites generating a highly ordered nucleosome array [20].

Promoters, with their tight interplay between DNA, transcription factors, Pol-II and chromatin, are regions of the genome for which both the DNA sequence and the pre–initiation complex may play a role in the chromatin organization. It has been reported that yeast promoters have nucleosome favoring sequences around them [21] and that there is a high correlation between *in vitro* and *in*

vivo nucleosome organisation in these regions [22, 23]. However, and in conflict whith these findings, another study concludes that the positioning of the +1 nucleosome in promoters is not due to nucloesome favoring sequences but rather linked to the process of transcriptional initiation itself. [24].
In this study we try to gain insights into the molecular mechanisms of TSS selection by jointly analyzing chromatin architectures and TSS patterns in different promoter classes, stratified by the presence or absence of specific CPEs. Our results show that promoters lacking CPEs have a strong sequence–intrinsic nucleosome–positioning signal in the +1 nucleosome region in both vertebrates and flies, and that the signal strength is correlated with the TSS pattern. Focused promoters have well positioned nucleosomes and a strong nucleosome–positioning signal, whereas broad promoters have fuzzy nucleosomes and a weak signal. Taken together, our data suggest that transcription factor binding and DNA sequence–induced nucleosome positioning are two mutually exclusive pathways of Pol-II recruitment and TSS selection in eukaryotic promoters.

2 Materials and Methods

2.1 Overview of Major Data Sets

The study was performed using publicly available data sets. The lists of promoters were part of the EPDnew database version 2 for human and version 1 for *D. melanogaster*. EPDnew is a collection of experimentally validated promoters that provide accurate estimates of TSS for a large fraction of protein coding genes in several organisms [25]. Human CAGE data was from the ENCODE consortium, GEO accession ID GSE34448 [27] using samples annotated as long–poly(A) cell extract (68 in total). *D. melanogaster* CAGE data was from Machibase [28] (7 samples from different developmental stages) and modENCODE preliminary study by Hoskins [11] (1 sample). Nucleosome mapping data for the GM12878 cell line produced by MNase–Seq was taken from the ENCODE consortium, GEO accession ID GSE35586 [29] and sample GEO ID GSM920558.

2.2 Position Weight Matrices for CPEs

Promoter lists were analyzed for the presence of core promoter elements using the TATA-box and Inr position weight matrices (PWMs) from [2]. The combined Inr-DPE matrix is posted on the EPD web server [25] under the link "Promoter elements". Promoter sequences were scanned with these PWMs using the cut–off values suggested in the original paper (TATA-box, Inr) or on the EPD web server (Inr-DPE).

2.3 Evaluation of Promoter Architecture

TSS patters were analyzed using publicly available CAGE data. The spread of CAGE tags in a window of 100 bp around the TSS was expressed as a Dispersion

Index (DI) using the following formula:

$$DI_k = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}}$$

Where N is the total number of tag starts in the window around promoter k, and x_i is the mapped position of the 5' end of tag i. DI values were calculated for each human promoter using CAGE data for 68 samples individually (long–poly–A samples from cell extract). A DI was calculated only if more than 5 tags mapped in the selected region. The sample–specific DI were then averaged to obtain a final unique and robust DI value for each promoter. The DI for *Drosophila* promoters was calculated in the same way using 8 CAGE samples.

2.4 Evaluation of Nucleosome Affinity Score

Promoters classified as CPE+ and CPE- were ranked from low to high DI and subsequently divided into groups of 2000 promoters for human and 1000 promoters for *D. melanogaster*. Proceeding this way, we were left with 1979 and 772 promoters in the bottom–ranked class of CPE- promoters for human and *Drosophila* respectively. Each group was analyzed for the presence of 10–bp sequence periodicities in the region +45 to +200 as follows. For the human collection, we recorded the cumulative frequency of AA/AT/TA/TT dinucleotides in a sliding window of size 4 shifted in 1 bp steps. For the *D. melanogaster* collection we used CC/CG/GC/GG instead, a window of 3 and allowing 1 mismatch. For each promoter group, the spectral density in the +1 nucleosome region was evaluated using the discrete Furrier transform (R function spec.pgram). From the resulting spectrum, the value corresponding to a frequency of 0.1 (corresponding to a period of 10 bp) was directly used as the nucleosome affinity score (NAS).

2.5 Nucleosome Distribution Around Promoters

Nucleosome mapping data for the GM12878 cell line was used to evaluate the nucleosome distribution around promoter lists using the ChIP–Seq web server [26]. Nucleosome tags were centered by 80 bp to account for the estimated fragment size of about 160 bp (centering parameter in the ChIP–Seq server). Multiple tags mapping to the same genomic location were removed from the analysis (parameter "Count cut–off" set to 1) and tag frequencies were calculated in a 10 bp sliding window.

3 Results and Discussion

3.1 The Impact of CPEs on TSS Dispersion

We sought to investigate the correlation between the CPEs and the fuzziness of TSS patterns. To this end, we classified promoters according to the presence or absence of CPEs and then looked at the TSS dispersion of the different

classes. Our analysis is based on the most recent human and *D. melanogaster* promoter collections from EPDNew [25]. As CPEs, we considered the TATA-box and Initiator (Inr) elements as described in [2]. For *D. melanogaster*, we further considered a combined Inr-DPE element, which has not been described before. The characteristics of this element (which is not the focus of this paper) will be briefly presented at the end of this Section.

A promoter was classified as TATA+ if a TATA-box was found at position -29 ± 3 relative to the TSS indicated in EPD. (The second T of TATA is used as reference position). For the Inr and Inr-DPE elements, we required exact occurrence at the TSS. Of 25988 human promoters, 2009 (8%) were thus classified as TATA+. Of the remaining promoters, 6039 (23%) were found to contain an Inr. Of 11389 *D. melanogaster* promoters, 1969 (17%) were TATA+. Of the TATA- promoters, 3892 (34%) contained an Inr and 2648 (23%) an Inr-DPE motif.

Next we computed a TSS dispersion index (DI) for all promoters (see Methods). TSS dispersion was evaluated using several CAGE data sets. CAGE (Cap Analysis of Gene Expression) is a high–throughput technique that allows the mapping of TSSs at single base resolution on a genomic scale [30]. The DI roughly reflects the spread of TSS in terms of a standard deviation around the center of gravity of the transcription initiation region. It ranges from 0 for the most focused promoters to about 30 for the most dispersed ones corresponding to broad promoters. The distribution of the DI for different promoter classes is shown in Figure 1a and d. Unsurprisingly the presence of a TATA-box, with its ability to recruit and position Pol-II to the initiation site, is correlated with a focused TSS. Both in human and *D. melanogaster*, TATA+ promoters have on average DI values smaller than TATA- promoters. In contrast and contrary to our expectation, the presence of the Inr motif was not correlated with TSS dispersion. However, in *D. melanogaster* the Inr-DPE appears to affect promoters in a similar way as the TATA-box.

The finding that the Inr element is not correlated with focused transcription initiation may be less surprising in the light of the following facts. Firstly, the Inr motif we used in this analysis is an extremely weak signal expected to occur approximately once every 25 bp. Secondly, EPDNew selects as representative position within a TSS region a base that corresponds to a local maximum in terms of CAGE tag counts. Any TSS region of 25 bp or more thus likely contains one or several Inr motifs. Being a weak motif, the Initiator may not be able to uniquely select a single site, but it may define a preferential initiation site on a local scale. This scenario would explain both the over–representation of the Inr motif at TSS positions provided by EPDNew and the absence of a correlation with focused initiation sites.

3.2 CPE- Promoters Have Nucleosome Favouring DNA Sequences in the +1 Nucleosome Region

In accordance with previously reported statistics [1, 3], [6] and [30], we were unabe to identify CPEs correlated with focused initiation in more than half of the promoters analyzed. Less than 8% of human promoters were classified as

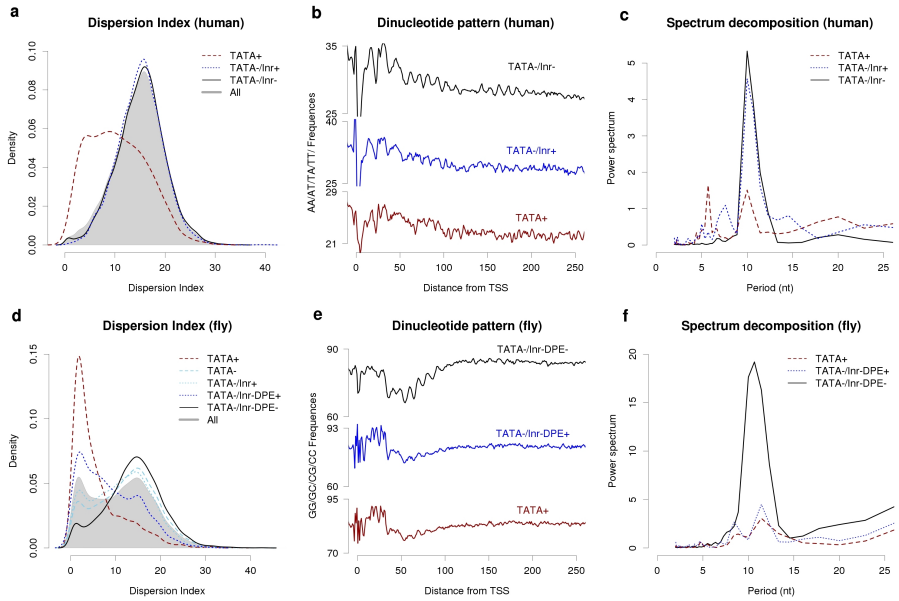

Fig. 1. Promoters architecture. (**a**) Dispersion index of human promoter classes stratified by the presence or absence of TATA–box and Inr motifs. (**b**) Frequence of AA/TT/AT/TA dinucleotides in human promoter classes aligned by the TSS. (**c**) Power spectrum analysis of the same human promoter classes for the region normally occupied by the +1 nucleosome (+50 to +200). (**d**) Dispersion index of fly promoter classes stratified by the presence or absence of TATA-box, Inr and Inr-DPE motifs. (**e**) Frequence of GG/CC/GC/CG dinucleotides in 3 fly promoter classes aligned by the TSS. (**f**) Power spectrum analysis of 3 fly promoter classes (TATA+, TATA-/Inr-DPE+ and TATA-/Inr-DPE-) for the region +50 to +200.

TATA+, and only 40% of *D. melanogaster* promoters had a recognizable TATA-box or Inr-DPE+ motif. This suggested that TF–mediated selection of TSS by Pol-II might be used only by a fraction of promoters and that the majority of them might use another mechanism to position Pol-II at the initiation site. We note in this context that promoters lacking CPEs (henceforth denoted CPE-) vary a lot in terms of TSS dispersion, including some promoters with rather focused initiation patterns.

We were wondering whether DNA sequence–driven nucleosome positioning could be the alternative mechanism of TSS selection in the CPE- promoter class. As mentioned before, nucleosomes appear to prefer sequences with a 10 bp periodic distribution of WW and SS dinucleotides (WW stays for AA, AT, TA or TT; SS for CC, CG, GC, or GG). We thus determined the local frequencies of these two dinucleotides in the downstream promoter regions of the different promoter classes (Figure 1, panels b and e). Visual inspection of the graphs suggests that CPE- promoters tend to have a strong sequence–intrinsic nucleosome–positioning

signal in the +1 nucleosome region. Consistent with previous observations, the human TATA-/Inr+ promoters constitute an exception in that they behave like CPE- promoters, whereas the *D. melanogaster* TATA-/Inr-DPE+ promoters follow the pattern of TATA+ promoters.

To evaluate the strength and the period of the dinucleotide periodicity in the +1 nucleosome region (from base 50 to base 200 relative to the TSS) a Fourier transform was applied to the dinucleotide occurrence profiles. The spectral decompositions in Figure 1 (panel c and f) confirm a strong 10–11 bp periodic signal, as well as the absence of such a periodicity for the TATA+ and Inr-DPE+ promoter classes. Moreover it reinforced the idea that only some CPEs are important in specifying promoter architecture. In human, the presence of the TATA-box completely abolished the signal. This was consistent with the idea that the TATA-box binding proteins drive the Pol-II to the TSS. In TATA+ promoters, the +1 nucleosome could play a secondary role in placing the Pol-II and could be positioned by the presence of the pre–initiation complex at the TSS and not by DNA–encoded signals. Alternatively, the absence of a strong dinucleotide periodicity could be explained by alternative rotational nucleosome positions in the same piece of DNA as recently suggested by Hapala and Trifonov [31]. Unlike the TATA-box, the Inr motif showed a pronounced signal (comparable to TATA-/Inr- promoters) suggesting that the nucleosome (and not the motif) could be important for TSS selection in this promoter class.

Promoter architecture in *D. melanogaster* showed a very similar picture. As in human, the presence of the TATA-box abolished the nucleosome signal. Moreover the Inr motif showed a strong nucleosome signal in the absence of a TATA-box and DPE motif (data not shown). Only the combined Inr-DPE motif (TATA-/Inr-DPE+ class) conferred a focused initiation and abolished the nucleosome signal (Figure 1 panel d, e and f).

These observations strengthen the idea that two mutually exclusive mechanisms could be responsible for the TSS selection in human and fly: TF–mediated and chromatin–mediated. The TF–mediated mechanism relies on the TATA-box (or Inr-DPE in fly) interacting with general transcription factors to drive Pol-II to the TSS. The chromatin–mediated mechanism relies on the sequence–specific interaction between the histone octamer and the DNA in the +1 nucleosomal region to drive Pol-II to the TSS.

3.3 The Nucleosome Signal Correlates with TSS Dispersion

It is important to recognize that the periodic dinucleotide distributions revealed by our analysis implies that all or a significant fraction of the CPE- promoters are rotationally oriented with regard to the +1 nucleosomes. In others words, RNA Pol II or the pre–initiation complex are bound to the DNA with a conserved angular position relative to the orientation of the first base–pair included in the +1 nucleosome. In principle, such rotational positioning could be achieved by a narrow TSS distribution or by a 10 bp periodic TSS distribution within a larger region.

To investigate in more detail the relationship between the DNA–encoded nucleosome signal and TSS dispersion patterns, CPE- and CPE+ promoters were ranked according to their DI and split into subsets of 2000 promoters (1000 for *D. melanogaster*). Subsequently, we computed a nucleosome affinity score (NAS) for each group, based on the Fourier transform of the dinucleotide occurrence profiles in the +1 nucleosome for each class (see Methods for details). We observe a strong inverse correlation between nucleosome affinity score (NAS) and TSS dispersion (DI) for the CPE- promoters (Figure 2 panels a and b). CPE- promoters with a focused initiation site had a strong nucleosome–positioning signal (high values of NAS). For the CPE- promoters subsets with more dispersed initiation sites we observed a weak signal (low values of NAS). Since we are analyzing the dinucleotide distribution of groups of promoters, we cannot rule out that individual promoters in these promoter classes still have strong NAS. In principle, the periodic signal could be annihilated by out–of–phase alignment of the sequences. Nevertheless, the strong correlation between DI and NAS speaks against such a scenario. Our results suggest that a +1 nucleosome strongly bound to the DNA has the capability of guiding the Pol-II complex to a specific position within a promoter region, resulting in focused initiation pattern. On the contrary, broad promoters could be the result of a weak nucleosome–DNA interaction or a yet unknown mechanism that over–rules the effect of the +1 nucleosome.

To collect additional evidence that positioned nucleosomes have the capability to select TSSs, we studied the distribution of CAGE tags and experimentally mapped (not sequence–predicted) nucleosomes in the lymphoblastoid cell line GM12878. Figure 2 (panel c and d) shows the results for two human promoters subsets with opposite characteristics: TATA- with high NAS and low DI, and TATA- with low NAS and high DI. As expected, promoters with high predicted nucleosome affinity had indeed well positioned nucleosomes whereas promoters with low nucleosome affinity showed grater fuzziness. We then looked at the distribution of minor initiation sites downstream of the major TSS indicated in EPD. The physiological role of these TSSs is of course unclear. But even if they just represent transcriptional noise, they may be able to tell us something about the molecular process underlying TSS selection. Interestingly in the high NAS/low DI class, downstream CAGE tags map preferentially to nucleosome free regions located about 240 and 420 bp downstream of the TSS. These findings reinforce the hypothesis that a strong nucleosome–DNA binding determines DNA accessibility to Pol-II and guides it to the initiation site. In fact, when nucleosome binding to the DNA is strong enough, Pol-II binding is outcompeted and transcription could start only in nucleosome–free regions.

3.4 Definition of the Inr-DPE Element

The DPE (downstream promoter element) was originally discovered in TATA-deficient *D. melanogaster* promoters and assigned the consensus sequence RGWCGTG [32]. According to a recent review [1] this 7–mer motif is located precisely at positions +27 to +33 relative to the TSS. It is further reported to be a recognition site for TFDII, which cooperatively binds to the Inr and DPE

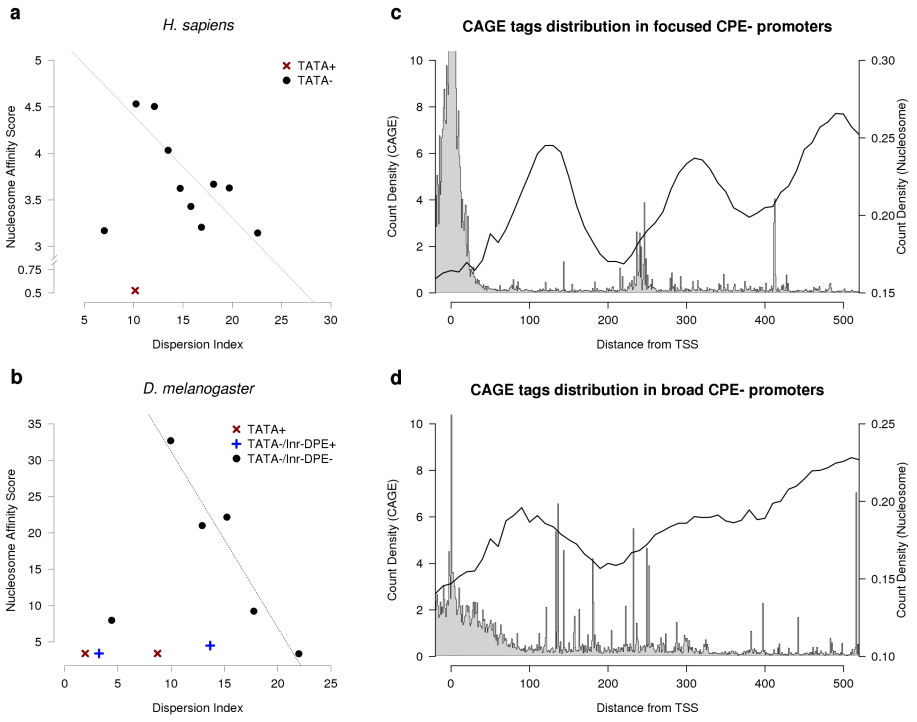

Fig. 2. Nucleosome and TSS dispersion. (a) Correlation between nucleosome affinity score (NAS) calculated in the region +50 tp +200 bp from the TSS and TSS dispersion index (DI) in human promoters stratified by the presence of the TATA-box. (b) Correlation between NAS calculated in the region +50 tp +200 bp from the TSS and TSS DI in fly promoters stratified by the presence of the TATA-box and the Inr-DPE motifs. (c) CAGE tag start position distribution and nucleosome (MNase) tag distribution in human GM12878 cell line for 2000 promoters with the higher NAS and lower DI. (d) CAGE tag start position distribution and nucleosome tag distribution in human GM12878 cell line for 1988 promoters with lower NAS and higher DI.

motifs. By analysis of the *D. melanogaster* promoter collection from EPD, we were able to confirm the co–occurrence of the Inr and DPE motifs, as well as the strict spacing requirement between them. In view of these observations, and taking into account that both elements interact with the same protein complex, we thought it would be more logical to view them as components of a single motif. Hence we set out to develop a new position weight matrix (PWM) for the combined Inr-DPE. We used the the PATOP algorithm [2] for this purpose, the same algorithm that was used to derive the TATA and Inr PWMs used in this work. The new Inr-DPE matrix, which is posted on the EPD web site [25], is 35 bp long and covers the promoter region from –2 to +32. Besides the conserved Inr and DPE regions, the matrix also exhibits conserved positions in the central area around +15, possibly reflecting the MTE motif identified by

Ohler and coworkers [33]. As it is assumed that DPE–like elements function in human promoters as well [34], we screened our human promoter collection for over–represented DPE motifs in the expected downstream region but were unsuccessful. We also scanned human promoters with the downstream part of the newly derived Inr-DPE matrix, again without obtaining any evidence for local motif enrichments. In view of these negative results, we did not use any DPE-like element for the stratification of human promoters.

References

1. Juven-Gershon, T., Hsu, J.H., Theisen, J.W.M., Kadonaga, J.T.: The RNA Plymerase II core promoter - the gateway to transcription. Curr. Opin. Cell Biol., 20(3), 253-259 (2008)
2. Bucher, P.: Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J. Mol Biol., 212(4), 563-578 (1990)
3. Kutach, A.K., Kadonaga, J.T.: The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. Mol. Cell. Biol. 20(13), 4754-4764 (2000)
4. Rhee, H. S., Pugh, B. F.: Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature, 483(7389), 295-301 (2012)
5. He, Y., Fang, J., Taatjes, D.J., Nogales, E.: Structural visualization of key steps in human transcription initiation. Nature, 495, 481-486 (2013)
6. Yang, C., Bolotin, E., Jiang, T., Sladek, F.M., Martinez, E.: Prevalence of the initiator over TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene, 389, 52-65 (2007)
7. Schmid, C.D., Bucheri, P.: ChIP-Seq data reveal nucleosome architecture of human promoters. Cell, 131(5), 831-832 (2007)
8. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., Wang, W., Weng, Z., Green, R.D., Crawford, G.E., Ren, B.: Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. 39, 311 - 318 (2007)
9. Belmont, A.S.: Large-scale chromatin organization: the good, the surprising, and the still perplexing. Cur. Op. Cell Biol., 26, 69-78 (2014).
10. Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A.M, Taylor, M.S., Engström, P.G., Frith, M.C., Forrest, A.R.R., Alkema, W.B., Lam Tan, S., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S.M., Wells, C.A., Orlando, V., Wahlestedt, C., Liu, E.T., Harbers, M., Kawai, J., Bajic, V.B., Hume, D.A., Hayashizaki, Y.: Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet., 38(6), 626-635 (2006)
11. Hoskins, R. A., Landolin, J. M., Brown, J. B., Sandler, J. E., Takahashi, H., Lassmann, T., Yu, C., Booth, B.W., Zhang, D., Wan, K.H., Yang, L., Boley, N., Andrews, J., Kaufman, T.C., Graveley, B.R., Bickel, P.J., Carninci, P., Carlson, J.W., Celniker, S. E.: Genome-wide analysis of promoter architecture in Drosophila melanogaster. Gen. Res., 21(2), 182-192 (2011)

12. Vavouri, T., Lehner, B.: Human genes with CpG island promoters have a distinct transcription-associated chromatin oerganisaition. Gen. Biol., 13, R110, (2012)
13. Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., Hume, D. A.: Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nature Reviews Genetics, 8(6), 424-436 (2007)
14. Trifonov, E.N., Sussman, J.L.: The pitch of chromatin DNA is reflected in its nucleotide sequence. PNAS, 77(7):3816-20 (1980)
15. Struhl, K., Segal, E.: Determinants of nucleosome positioning. Nature Structural and Molecular Biology, 20(3), 267-273 (2013)
16. Albert, I., Mavrich, T. N., Tomsho, L. P., Qi, J., Zanton, S. J., Schuster, S. C., Pugh, B. F.: Translational and rotational settings of H2A. Z nucleosomes across the Saccharomyces cerevisiae genome. Nature, 446(7135), 572-576 (2007)
17. Fedor, M.J., Lue, N.F., Kornberg, R.D. Statistical positioning of nucleosomes by specific protein binding to an upstream activating sequence in yeast. J. Mol. Biol. 204, 109-127 (1988)
18. Gaffney, D.J., McVicker, G., Pai, A.A., Fondufe-Mittendorf, Y.N., Lewellen, N., Michelini, K., Widom, J., Gilad, Y., Pritchard, J.K.: Controls of nucleosome positioning in the human genome. Genome research 8(11), e1003036 (2012)
19. Valouev, A., Johnson, S. M., Boyd, S. D., Smith, C. L., Fire, A. Z., Sidow, A.: Determinants of nucleosome organization in primary human cells. Nature, 474(7352), 516-520 (2011)
20. Fu, Y., Sinha, M., Peterson, C. L., Weng, Z.: The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. PLoS genetics, 4(7), e1000138 (2008)
21. Mavrich, T. N., Ioshikhes, I. P., Venters, B. J., Jiang, C., Tomsho, L. P., Qi, J., Schuster, S.C., Albert, I., Pugh, B. F.: A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome research, 18(7), 1073-1083 (2008)
22. Kaplan, N., Moore, I., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., Hughes, T.R., Lieb, J.D., Widom, J., Segal, E.: Nucleosome sequence preferences influence in vivo nucleosome organization. Nature structural and molecular biology, 17(8), 918-920 (2009)
23. Field, Y., Fondufe-Mittendorf, Y., Moore, I. K., Mieczkowski, P., Kaplan, N., Lubling, Y., Lieb, J.D., Widom, J., Segal, E.: Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. Nature genetics, 41(4), 438-445 (2009)
24. Zhang, Y., Moqtaderi, Z., Rattner, B. P., Euskirchen, G., Snyder, M., Kadonaga, J. T., Liu, S., Struhl, K.: Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. Nature Structural and Molecular Biology, 16(8), 847-852 (2009)
25. Dreos, R., Ambrosini, G., Périer, R.C. and Bucher, P.: EPD and EPDnew, high-quality promoter resources in the next-generation sequencing era. NAR, 41(D1), D157-D164 (2013)
26. The ChIP-Seq Web Server, `http://ccg.vital-it.ch/chipseq/`
27. Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G.K.. Khatun, J., Williams, B.A., Zaleski, C., Rozowsky, J., Roder, M., Kokocinski, F., Abdelhamid, R.F., Alioto, T., Antoshechkin, I., Baer, M.T., Bar, N.S., Batut, P., Bell, K., Bell, I., Chakrabortty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M.J., Gao, H., Gonzalez, D., Gordon,

A., Gunawardena, H., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O.J., Park, E., Persaud, K., Preall, J.B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L., Shahab, A., Skancke, J., Suzuki, A.M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S.E., Hannon, G., Giddings, M.C., Ruan, Y., Wold, B., Carninci, P., Guigo, R., Gingeras, T.R.: Landscape of transcription in human cells. Nature, 489(7414), 101-108 (2012)

28. Ahsan, B., Saito, T. L., Hashimoto, S. I., Muramatsu, K., Tsuda, M., Sasaki, A., Matsushima, K., Aigaki, T., Morishita, S.: MachiBase: a Drosophila melanogaster 5'-end mRNA transcription database. NAR, 37(suppl 1), D49-D53 (2009)
29. Kundaje, A., Kyriazopoulou-Panagiotopoulou, S., Libbrecht, M., Smith, C.L., et al.: Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. Genome Res, 22(9), 1735-1747 (2012)
30. Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., Fukuda, S., Sasaki, D., Podhajska, A., Harbers, M., Kawai, J., Hayashizaki, Y.: Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. PNAS, 100(26), 15776-15781 (2003)
31. Hapala, J., Trifonov, E.N.: Nucleosomal TATA-switch: competing orientations of TATA on the nucleosome. Gene, 527(1), 339-343 (2013)
32. Burke, T.W., Kadonaga, J.T.: Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes and Dev., 10, 711-724 (1996)
33. Ohler, U., Liao, G., Niemann, H., Rubin, G.M.: Computational analysis of core promoters in the Drosophila genome. Genome Biol., 3, RESEARCH0087 (2002)
34. Juven-Gershon, T., Kadonaga, J.T.: Regulation of gene expression via the core promoter and the basal transcriptional machinery. Developmental biology, 339(2), 225-229 (2010)