Looking for Tumor Specific Promoters In Silico

K. N. Kashkin[1]

Shemyakin−Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, 117997 Russia
Received April 6, 2022; revised May 18, 2022; accepted May 20, 2022

Abstract—Previously we demonstrated the tumor-specific activity of several human native and chimeric promoters. Here we have analyzed the DNA sequences of experimentally tested tumor-specific promoters for the presence of recognition matrices of transcription factors and for de novo motif discovery. CiiiDER and MEME Suite software tools were used for this purpose. A number of transcription factor matrices have been identified, which are present more often in tumor-specific promoters than in the promoters of housekeeping genes. New promoter−TF regulatory relationships have been predicted by pathway analysis. A motif of 44 bp characteristic of tumor-specific promoters but not of housekeeping gene promoters has been discovered. The search through 29 598 human promoters from the EPDnew promoter database has revealed a series of promoters with this motif, their genes being associated with unfavorable prognoses in cancer. We suppose that some of these promoters may possess a tumor specific activity. In addition, a close similarity in nucleotide motifs between the promoters of the *BIRC5* and *MCM2* genes has been shown. The results of the study may contribute to understanding the peculiarities of gene transcription in tumors, as well as to searching for native tumor-specific promoters or creating artificial ones for cancer gene therapy, as well as in the development of anticancer vaccines.

Keywords: transcription, regulation of transcription, cancer, tumor-specific, promoter, transcription factor

DOI: 10.1134/S1068162022060127

INTRODUCTION

The functioning of a eukaryotic promoter is a complex process including the interaction between *cis*- and *trans*-regulatory elements of the promoter and transcription factors (TF), the enzymes of the transcription complex, and regulatory RNA [1]. Epigenetic mechanisms such as conformational changes in chromatin, DNA methylation, histone modifications [2], as well as physicochemical processes related to phase separation [3] play a substantial role in transcription regulation. However, the overall regulation of promotor activities is based on TF binding with specific sequences in DNA and interaction between the factors.

At present, more than 1600 human TFs are known altogether, and their number is continuously increasing, but the sequences of DNA binding sites are unknown for many of them (more than 500 factors) [4]. The essential features of TF binding sites are their degeneracy and cross affinity of different TFs to the same DNA sequence [5]. This results in the probabilistic and competitive regulation of transcription.

The peculiarities of genes transcription, together with metabolic rearrangements in transformed cells, determine the ability of tumors to grow slipping out of control of the surrounding tissues and a whole body; therefore, the study of these processes is of theoretical and practical importance for oncology. Some of the known TFs are considered as diagnostic markers for tumors or their metastases. For example, CDX2 is a sensitive marker of colorectal adenocarcinomas, TTF1 is a significant marker for adenocarcinomas of the lung, PAX8 can be a marker for gynecologic tumors, etc. [6]. That is why the study of TFs, in particular, the presence of their recognition profiles in the promoters, remain relevant.

At present, the accumulated data on the structure of promoters have already opened up great prospects for using them in practical medicine, e.g., as components of genetic engineering vectors of gene therapy for tumors [7, 8]. However, the choice of optimal promoters, as well as the search of new promoters for genetic engineering therapeutic vectors, are still nontrivial problems. Great interest in promoters has been recently renewed due to the need of designing vaccine vectors for prevention and treatment of infectious diseases and cancer [9, 10].

Previously, we cloned a number of gene promoters regulating the proliferation of cells exhibiting tumor-specific activity [11, 12] (Table 1). The present work was aimed at revealing the peculiarities of these pro-

---

Abbreviations: TF, transcription factor; EPDnew, Eucaryotic Promoter Database; TCGA, The Cancer Genome Atlas; THPA, The Human Protein Atlas; TSS, Transcription Start Site.
[1] Corresponding author: phone: +7 (495) 330-63-29; e-mail: kachkine@yandex.ru.

Table 1. The promoters studied in the present work

Promoter	EPDnew index
Tumor-specific promoters*	
BIRC5	FP024000
CDC6	FP023205
CKS1B	FP001855
MCM2	FP005950
PLK1	FP021553
POLD1	FP026208
TERT	FP007749
Promoter PCNA*	
PCNA	FP026732
Chimeric promoters	
CH2-CH26	[13]
Promoters of housekeeping genes	
ACTB	FP010679
ALAS1	FP005471
B2M	FP020416
GAPDH	FP017185
GUSB	FP011089, FP011090
HMBS	FP016933, FP016934
HPRT1	FP029346, FP029347
PGK1	FP029028
PPIA1	FP011011
RPL13A	FP026153
RPLP0	FP018538, FP018539
SDHA	FP007723
TBP	FP010603, FP010604
TFRC	FP006502
TUBB	FP009374
YWHAZ	FP012696, FP012697, FP012698

*Only the promoters analyzed in the experiments [11−13].

moters, as well as chimeric promoters derived from the latter [13], which determine their enhanced activity in tumor cells. For this purpose, we have compared these promoters with a set of housekeeping gene promoters with respect to composition of the known TF recognition profiles, as well as searched for nucleotide motifs de novo in the promoters.

RESULTS

Predominant profiles in tumor-specific promoters. Previously we have shown that relatively short (hundreds of nucleotides) promoters of some genes (Table 1) involved in the regulation of proliferation have a tumor-specific activity [11, 12]. It is known that the most important regulatory elements of the analyzed promoters are concentrated within a region of approx-

imately up to −500 bp from the transcription start site (TSS) [12, 14, 15]. Therefore, we chose seven DNA sequences from the EPDnew eukaryotic promoter database with the coordinates [−499; +100] relative to the transcription start site, which corresponded to the previously studied tumor-specific promoters. Since there is more than one promoter known for some of these genes, we selected only the promoters that had been studied in direct experiments (Table 1). For comparison, we used 23 housekeeping gene promoters of the same length. Using the CiiiDER software [16] and the JASPAR bases of transcription factor profiles (http://jaspar.genereg.net/) [5], we have performed the comparison (enrichment) of tumor-specific promoters with housekeeping gene promoters with respect to the composition of TF recognition profiles (Fig. S1, Supplementary Information). At the same time, we compared the numbers of promoters in both groups with the respective TF profiles; the result was estimated using the Mann−Whitney test. The complete results of promoter comparison are given in Table S1 in Supplementary Information. Tables 2 and S2 (see Supplementary Information) show the TF profiles, which are statistically significantly ($p < 0.05$) predominant in tumor-specific promoters compared to housekeeping gene promoters (hereinafter, conditionally tumor-specific TFs).

Among the seven promoters, the MCM2, CKS1B and PLK1 promoters are more enriched in the TF profiles of this group than other promoters (13, 10 and 9 profiles, respectively) (Table S3 in Supplementary Information). More than half of tumor-specific promoters were shown to contain the recognition profiles of factors SREBF2, ZNF75D, Zfx (each in six promoters), RUNX2 and ETS2 (each in five promoters), and Creb3l2 (in four promoters) (Table 2). Preferential locations in the promoters were successfully identified for three TFs (Fig. 1). For example, the SREBF2 profiles in four out of seven promoters are localized in the region [−458;−384] relative of TSS on the noncoding DNA strand. The ZNF75D profiles take the region [−253;−184] relative to TSS also on the noncoding DNA strand in four promoters and nearby TSS [−24;+30] on the coding strand in three promoters. The RUNX2 profiles are localized on the coding strand in the region [−414;−247] in four promoters.

Interaction between transcription factors and promoters. Using the Pathway Commons resource [17], we have constructed a network of direct regulatory interactions between the revealed TFs and the products of the studied tumor-specific genes (Fig. 2). The scheme also shows the proliferating cell nuclear antigen (PCNA), because it interacts with the products of the genes, the promoters of which have been studied. PCNA is involved in the regulation of cell proliferation and has a marked tumor-specific expression [18]. However, the PCNA promoter that we cloned [19] was active both in tumor and normal cells [20] and, hence,

Table 2. The transcription factors with the profiles predominant in tumor-specific promoters compared to the promoters of housekeeping genes

Transcription factor	JASPAR nomenclature	Class	Family	Contained in tumor-specific promoters
RUNX2*	MA0511.2	Runt domain factors	Runt-related factors	*POLD1, BIRC5, PLK1, CDC6, CKS1B*
Creb3l2*	MA0608.1	Basic leucine zipper factors (bZIP)	CREB-related factors	*MCM2, PLK1, CDC6, TERT*
SREBF2*	MA0596.1	Basic helix-loop-helix factors (bHLH)	bHLH-ZIP factors	*MCM2, POLD1, BIRC5, PLK1, CKS1B, TERT*
ETS2*	MA1484.1	Tryptophan cluster factors	Ets-related factors	*MCM2, PLK1, CDC6, CKS1B, TERT*
CENPB	MA0637.1	CENP-B		*MCM2, CDC6, CKS1B*
HEY1	MA0823.1	Basic helix-loop-helix factors (bHLH)	Hairy-related factors	*MCM2, PLK1, TERT*
HEY2	MA0649.1	Basic helix-loop-helix factors (bHLH)	Hairy-related factors	*MCM2, PLK1, TERT*
RARA::RXRG	MA1149.1	Nuclear receptors with C4 zinc fingers::Nuclear receptors with C4 zinc fingers	Thyroid hormone receptor-related factors (NR1)::RXR-related receptors (NR2)	*MCM2, POLD1, CKS1B*
SREBF1(var.2)	MA0829.2	Basic helix-loop-helix factors (bHLH)	bHLH-ZIP factors	*MCM2, BIRC5, PLK1*
ZNF75D*	MA1601.1	C2H2 zinc finger factors	More than 3 adjacent zinc finger factors	*MCM2, POLD1, PLK1, CDC6, CKS1B, TERT*
Zfx*	MA0146.2	C2H2 zinc finger factors	More than 3 adjacent zinc finger factors	*MCM2, POLD1, BIRC5, CDC6, CKS1B, TERT*
HOXD11	MA0908.1	Homeo domain factors	HOX-related factors	*POLD1, CKS1B*
KLF11	MA1512.1	C2H2 zinc finger factors	Three-zinc finger Kruppel-related factors	*POLD1, BIRC5*
MEF2A	MA0052.4	MADS box factors	Regulators of differentiation	*MCM2, BIRC5*
RORB	MA1150.1	Nuclear receptors with C4 zinc fingers	Thyroid hormone receptor-related factors (NR1)	*MCM2, POLD1*
SIX1	MA1118.1	Homeo domain factors	HD-SINE factors	*PLK1, CKS1B*
Stat6	MA0520.1	STAT domain factors	STAT factors	*MCM2, CKS1B*

The presented factors have $p < 0.05$ according to Mann−Whitney test for the number of tumor-specific promoters relative to the number of promoters of housekeeping genes containing this profile. The complete results of comparison are presented in Table S1 in Supplementary Information. The TF recognition profiles are presented in Tables S2 and S4 in Supplementary Information.
*The profiles found in four and more tumor-specific promoters.

initially was not included in the differential analysis of promoters. In the present work, the *PCNA* (FP026732) promoter is considered as nonspecific.

As one can see from Fig. 2a, MCM2 and PLK1 take the central position in this regulatory network, which is in agreement with the presence in their promoter of a great number of profiles of conditionally tumor-specific TFs. The Creb3l2, CENPB, SIX1, ZNF75D, Zfx, HOXD11 and KLF11 factors were not included into the regulatory network though their recognition profiles are present in tumor-specific promoters much more often than in housekeeping gene promoters. There are no direct data on the involvement of these seven TFs in the regulation of promoters under study. However, their selection does not seem to be random, because these TFs, according to the GEPIA2 database, demonstrate differential expression or are associated with specific prognosis in some tumors (Table S2 in Supplementary Information) and, consequently, can play important roles in carcinogenesis.

Nonspecific transcription factors. Additionally, we have determined the TF profiles that more often occur in the promoters of housekeeping genes compared to tumor-specific promoters (Tables S4 and S5 in Supplementary Information; hereinafter, conditionally
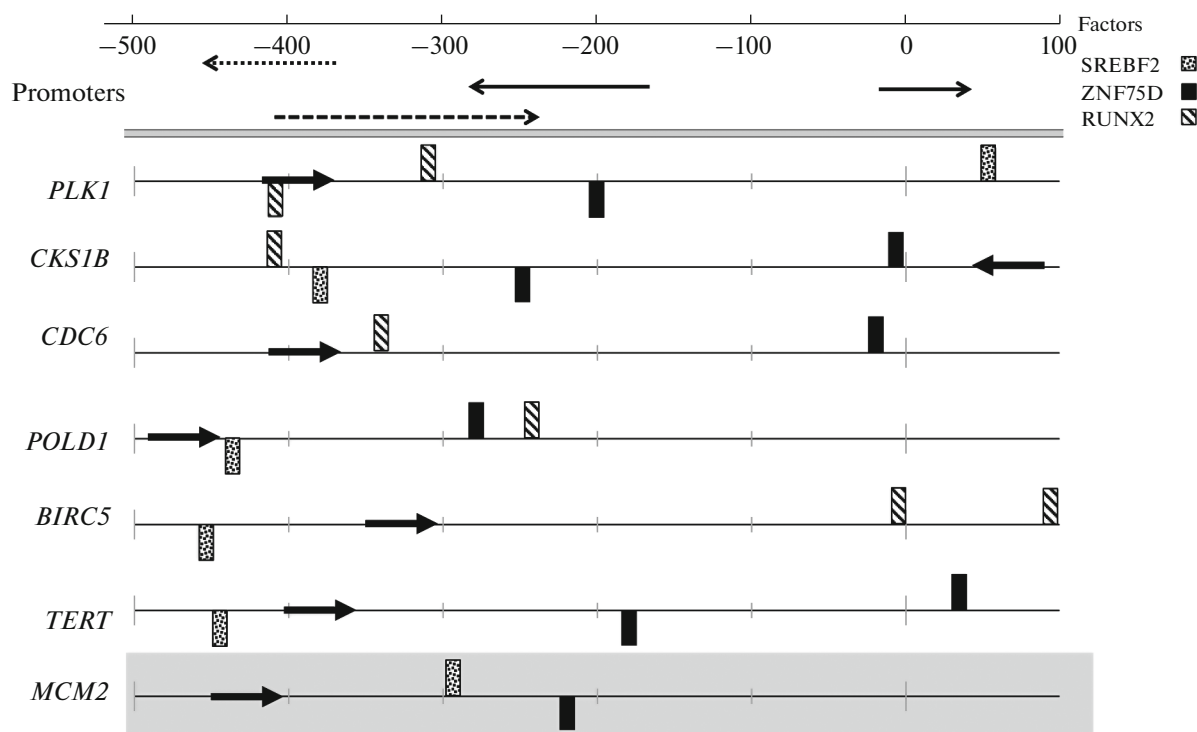
Fig. 1. The positions of recognition profiles of transcription factors SREBF2, ZNF75D and RUNX2 in promoters. Transcription start sites (0) are indicated according to the EPDnew eukaryotic promoter database. The bold arrows indicate Motif 1.

nonspecific factors). It should be noted that these factors also demonstrate differential expression and can be prognostic markers in some cancers. At the same time, six out of seven TFs of this group participate in the development and differentiation of various types of cells (Mafb [21], Arid3a [22], MEIS3 [23], BHLHA15 [24], BSX [25], E2F1 [26]). The TBP factor is related to the general transcription mechanisms [27]. According to Pathway Commons, only TBP and E2F1 directly interact with the products of tumor-specific genes under study (Fig. 2b).

Search of nucleotide motifs de novo. For improving the search of regulatory DNA sequences typical of tumor-specific promoters, we searched for motifs de novo using the MEME Suite software [28].

We have performed discriminative search of nucleotide motifs in seven tumor-specific promoters relative to housekeeping gene promoters. Motif 1 of 44 bp found thereby is present in all tumor-specific promoters (Fig. S2, Supplementary Information). According to the GOMO test, the functions of Motif 1 are associated with the activity of olfactory receptors, the sensory perception of smell and the G-protein coupled receptor protein signaling pathway. In six out of seven promoters under study, Motif 1 is localized approximately in the [−490;−210] region relative to TSS. In the *CKS1B* promoter, Motif 1 is localized in the [−47;−4] region, in the opposite direction relative to the promoter. It is probably due to the fact that the

CKS1B/SHC1 promoter is bidirectional. It can be supposed that Motif 1 consists of three shorter submotifs. Therefore, we have analyzed these submotifs separately using the Tomtom program from MEME Suite and the database of vertebrate TFs (Table S6, Supplementary Information).

Submotif 1a contains the recognition profiles of myocyte enhancer factors MEF2, which can activate the genes induced by growth factors and stress and are involved in both the suppression and progression of cancer under various conditions [29]; the profile of the forkhead E1 factor associated with thyroid cancer [30] and the profile of the POU homeobox transcription factor, which is one of the master regulators of small cell lung cancer [31]. The NR2E1 factor is also important for carcinogenesis, because it is associated, in particular, with metastasis of breast cancer [32]. PHOX2B has been defined as the key regulator of differentiation and maintenance of stem cell properties of neuroblastoma [33]. Sp100 is a component of promyelocytic leukemia (PML) nuclear bodies [34].

Submotif 1b contains the recognition profile for retinoid X receptors (RXRs) and retinoic acid receptors (RARs), the nuclear receptors mediating the biological effects of retinoids associated with tumor development [35]. Activation of the *DMRT1* gene was found in testicular germ cell neoplasia [36]. The interferon-regulating factor 3 (IFN3) is involved in import-
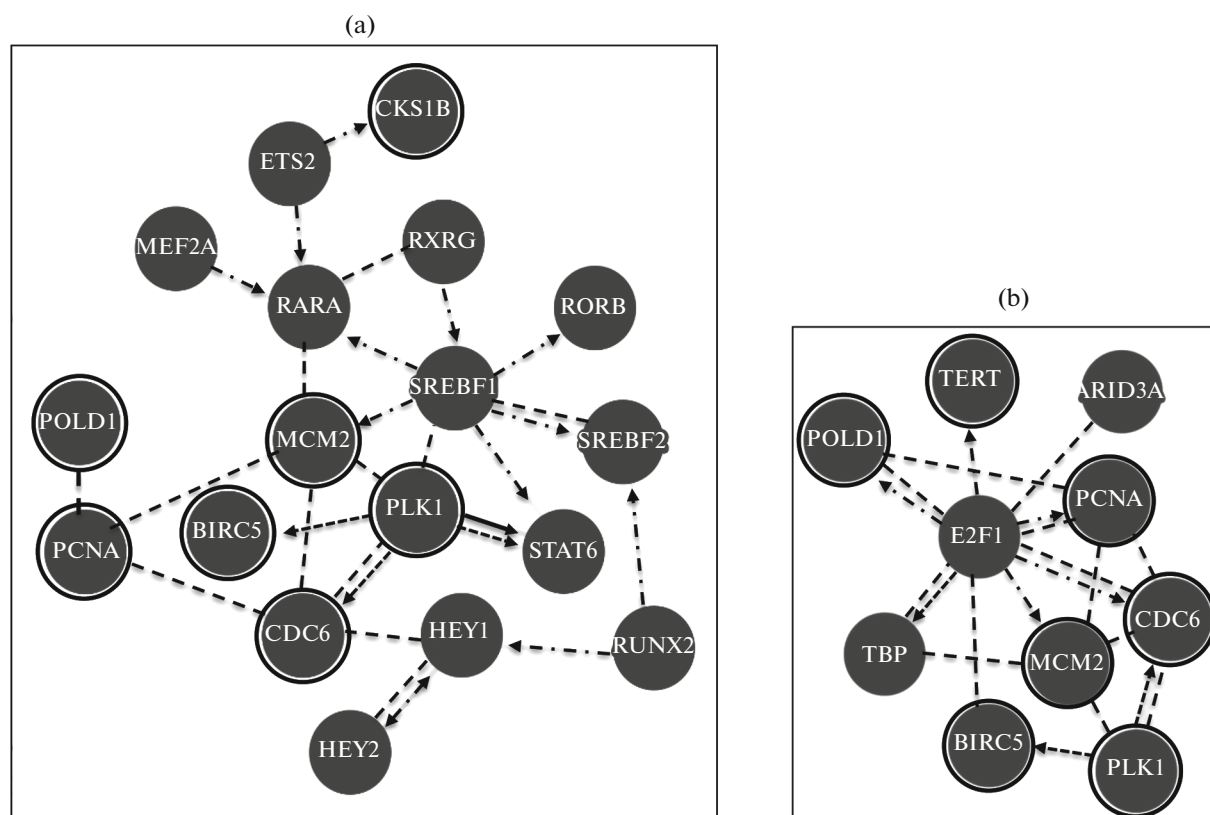
Fig. 2. The direct regulatory interactions between the products of genes under study (in black circles) and transcription factors (Pathway Commons): (a) for conditionally tumor-specific transcription factors; (b) for conditionally nonspecific transcription factors. The dashed−dotted arrows show expression regulation; the dashed arrows show modifications; the solid arrow shows transport control; the dashed lines without arrows show binding into a complex.

ant processes such as anticancer immunity and resistance to some bacterial and viral infections [37].

Submotif 1c includes the sequence responsible for cAMP (cAMP-response element) and capable of binding TFs of the CREB3 family regulating the proliferation and migration of cancer cells [38], as well as the profile of the ATF6 factor involved in the unfolded protein response; the overexpression of ATF6 correlates with tumor aggressiveness [39]. ZNF135 is a target of regulation by differentially expressed microRNA in nasopharyngeal carcinoma [40].

Certainly, Motif 1 may be shown to contain the recognition profiles of other TFs, in addition to those listed above, as well as unknown TFs, due to degeneracy. Nevertheless, our data suggest that Motif 1 can also be typical of other tumor-specific promoters.

We have tried to find other promoters containing Motif 1. For this purpose, we have scanned 29 598 human promoters from the EPDnew promoter database (all in the [−499;+100] coordinates relative to TSS) using the FIMO program from MEME Suite. We have found 4733 sequences in both orientations with $p < 0.0001$, including all of the seven promoters under study. Ten promoters containing the sequences

closest to Motif 1 are presented in Table 3. According to The Human Protein Atlas, the respective nine (90%) genes are prognostic markers for tumors (mainly unfavorable). It is notable that the position of Motif 1 in the presented promoters relative to TSS approximately corresponds to that in the seven analyzed tumor-specific promoters.

Among the seven promoters under study, the *MCM2* and *BIRC5* promoters proved to be the closest ones to Motif 1. Separate analysis of the motifs de novo in these two promoters using MEME Suite revealed a consensus Motif 2 of 184 bp in the coordinates [−579;−396] for *MCM2* and [−496; −303] for *BIRC5* relative to TSS (Fig. S3, Supplementary Information). Interestingly, the region of similarity between the *MCM2* and *BIRC5* promoters is noticeably wider; it includes 296 bp and has the same order of shorter motifs. The motif similar to Motif 2 can also be found in another five promoters using FIMO, but with very low estimates (data not shown). Therefore, we consider such similarity to be the peculiar feature of the *MCM2* and *BIRC5* promoters.

Analysis of chimeric promoters. We have studied the chimeric promoters possessing the higher activity in

Table 3. The ten promoters most similar to Motif 1 from the EPDnew database

No.	EPDnew index	Gene promoter	Start	Stop	Strand	Score	p-Value	q-Value	Prognostic significance in tumors*
1	FP024000	BIRC5_1	−357	−314	+	48.9817	2.22E-19	7.32E-12	Kidney (−), liver (−), lung (−) cancer
2	FP020135	CDCA4_2	−437	−394	−	48.4404	5.93E-19	9.78E-12	Kidney (−), liver (−), pancreatic (−), lung (−) cancer
3	FP014177	WDR37_1	−423	−380	+	48.0183	1.24E-18	1.35E-11	Pancreatic (+), thyroid (−), head and neck (+) cancer
4	FP025645	YIF1B_2	−487	−444	−	47.6055	2.45E-18	1.35E-11	Liver (−), pancreatic (+) cancer
5	FP027891	RSPH14_1	−478	−435	−	47.6055	2.45E-18	1.35E-11	Kidney (+) cancer
6	FP011458	SRRT_2	−403	−360	+	47.6055	2.45E-18	1.35E-11	Kidney (−), liver (−) cancer
7	FP003432	PLEK_3	−470	−427	+	47.4495	3.21E-18	1.35E-11	Kidney (−) cancer
8	FP013453	S1PR3_2	−448	−405	−	47.4404	3.28E-18	1.35E-11	Kidney (−), colorectal (−), breast (+) cancer
9	FP026025	C5AR1_1	−482	−439	−	47.1927	4.74E-18	1.62E-11	Kidney (−), cervical (+), testicular (−), ovarian (−) cancer
10	FP009477	MSH5-SAPCD1_5	−300	−257	+	47.156	5.2E-18	1.62E-11	NP

Start, stop—the positions of Motif 1 relative to the transcription start site; score—estimation of the correspondence of a promoter sequence to the preset motif; *p*-value—the probability that a random sequences of the same length corresponds to the preset motif with the same or better result; *q*-value—false positive rate, if the occurrence frequency of the motif is taken as significant (FIMO).* Prognostic markers in tumors according to The Human Protein Atlas database: (+)—favorable prognosis, (−)—unfavorable prognosis, NP—no relation to prognosis.

A431 epidermoid carcinoma cells compared to normal fibroblasts (Fig. S4, Supplementary Information) [13]. CiiiDER was used in the search of conditionally tumor-specific and nonspecific TF profiles in chimeric promoters, as well as in nonspecific CMV and *PCNA* promoters. The profiles of Creb3l2, ETS2, HEY1, HEY2 and SREBF1 (var. 2) have been found only in the chimeric promoters with enhanced activity in the A431 cells (CH2, CH20, CH26, Fig. S4 in Supplementary Information), but they are absent in nonspecific CH10, CMV, *PCNA* promoters (Table S7, Supplementary Information). All profiles of conditionally nonspecific TFs prevailing in housekeeping gene promoters (Table S5, Supplementary Information) occur in at least one of the three nonspecific promoters (Table S8, Supplementary Information).

Though this observation can hardly be assessed statistically, we are inclined to consider it as a tendency that confirms our conclusions but requires further verification. Only two chimeric promoters, CH8 and CH16, contain Motif 1 of 44 bp in length within a promoter fragment *POLD1* (data not shown).

DISCUSSION

The accumulated data on the structure and functions of promoters have already opened up great prospects for their use in practical medicine, e.g., as components of gene engineering vectors in the gene therapy for tumors [7, 8]. The clinical trials of gene therapy constructs containing viral and human promoters have been completed (ClinicalTrials.gov Identifier: NCT01455259, NCT00891748, NCT00197522, NCT00051480; see the review by Ginn et al. [8], etc.). In addition to the gene therapy for cancer, native promoters can be used in gene engineering vectors for treating other diseases, in industry and in various biotechnological processes. Recently, there has been a renewed interest in promoters due to need for vaccine vectors for preventing or treating infectious diseases and cancer [9, 10]. The approaches are being developed with construction of hybrid promoters [41, 42] and chimeric promoters [13]. Therefore, the study of the primary structure of promoters, in particular, tumor-specific promoters, is still relevant.

In the present work we have studied seven tumor-specific promoters (Table 1) with respect to the content of recognition profiles of the known TFs. In spite of the fact that transcription regulation involves DNA sequences both adjacent to TSS and remote, it has been shown that certain regulatory functions are maintained in relatively short proximal promoters being the minimum (core) promoters with adjacent *cis*-regulatory elements, with a total length of several hundreds of nucleotides [43]. This is also true for the promoters under study, where many important regula-

tory elements are concentrated within a region of approximately up to −500 bp from TSS [12, 14, 15]. Such size is convenient for genetic manipulations with these promoters.

The comparison of the promoters revealed 17 TF recognition profiles, more frequent in tumor-specific promoters than in housekeeping gene promoters (conditionally tumor-specific TFs, Table 2). Table S2 (see Supplementary Information) presents recognition profiles of these factors according to the JASPAR database [5]. The profiles of the SREBF2, ZNF75D, Zfx, RUNX2, ETS2 and Creb3l2 factors were found in more than half of tumor-specific promoters. The rest of 17 TFs occurred in some tumor-specific promoters but in none of the 23 promoters of the reference group. We have also succeeded in determining the preferential locations of recognition profiles of the SREBF2, ZNF75D and RUNX2 factors in tumor-specific promoters (Fig. 1). It should be noted that the *CKS1B* and *PLK1* promoters containing 10 and 9 TF profiles of this group, respectively (Table S9, Supplementary Information), previously have shown the maximum tumor specificity compared to other promoters [12].

Some tumor-specific promoters contain the recognition profiles of TFs with not enough data on their involvement in the regulation of these promoters. Such factors are Creb3l2, CENPB, SIX1, ZNF75D, HOXD11, Zfx and KLF11.

Probably, our results will be an argument in favor of the study of the role of these TFs in regulation of the respective genes.

In addition, we have detected seven profiles of TFs represented to a greater extent in housekeeping gene promoters than in the promoters under study (Tables S4 and S5, Supplementary Information). Six out of seven TFs of this group participate in the development and differentiation of cells of various types (see above). Cell differentiation is often considered as a process opposite to malignant transformation and proposed to be used for treating tumors [44]. Our observations show that TF selection with respect to nonspecificity is nonrandom in this case. At the same time, the differential expression of these TFs in some tumors and their association with prognosis (Table S4, Supplementary Information) demonstrate the need for more thorough investigation of this problem.

Obviously, the binding of a transcription factor to DNA depends not only on the nucleotide sequence but also on many other parameters: profile orientation, adjacent sequences, interaction between this TF and other regulatory molecules and sequences, etc. [43]. In addition, conditionally tumor-specific TFs can be related not to the malignant transformation of cells but rather to the proliferative functions of genes,

and this difference should be studied in each particular case. Nevertheless, the findings provide an opportunity to investigate promoter regulation more precisely in direct experiments.

Using a de novo discriminatory search of motifs, we have found a sequence of 44 bp (Motif 1), which is present in all seven tumor-specific promoters but is not characteristic of housekeeping gene promoters. In the EPDnew database, we have found a group of promoters with such a motif, the respective genes being the prognostic markers for tumors. We believe that at least some of these promoters may have a tumor-specific activity.

We have shown that two promoters, *BIRC5* and *MCM2*, contain a highly similar motif of 184 bp. Moreover, the region of similarity between these two promoters spreads to 296 bp. We assume that such similarity is a peculiar feature of these two promoters associated with their regulation rather than the typical common feature of tumor-specific promoters. Though this observation is interesting, it is beyond the framework of our research.

Our conclusions can be partially verified by the study of TF profiles in chimeric promoters. Five out of 17 profiles of conditionally tumor-specific TFs, namely, the Creb3l2, ETS2, HEY1, HEY2 and SREBF1 (var. 2) profiles, have undergone selection on A431 cells and were chosen in chimeric promoters with respect to tumor specificity [13]. Only two chimeric promoters, CH8 and CH16, contain Motif 1 of 44 bp in length, which has been found in seven native tumor-specific promoters. Together with other data, it probably means that the presence of Motif 1 in a promoter is a sufficient but not necessary condition for tumor specificity. Obviously, these issues require more comprehensive investigation.

EXPERIMENTAL

Human promoter sequences were obtained from the EPDnew database (Eucaryotic Promoter Database, https://epd.epfl.ch/EPDnew_database.php) [45] in the [−499;+100] coordinates relative to the transcription start site (TSS) (Table 1). The search and comparison of TF profiles in promoters and the search of motifs de novo were performed using CiiiDER [16] and MEMESuite [28] (https://meme-suite.org/meme/) with default parameters and the bases of transcription factor binding profiles: JASPAR (http://jaspar.genereg.net/) [5] and Jolma, 2013 [46]. The differences between the groups of promoters with respect to the frequency of TF profiles were estimated by the statistical Mann−Whitney test, primarily taking into account the "significance score" and "gene *p*-value" parameters (CiiiDER). TF expression in tumors is presented according to the GEPIA2 database

(http://gepia2.cancer-pku.cn/). The prognostic significance of TFs in tumors was determined using The Human Protein Atlas (https://www.proteinatlas.org/) [47]. The networks of regulatory interactions were constructed using the Pathway Commons resource (http://www.pathwaycommons.org/) [17].

CONCLUSIONS

It has been shown that the tumor-specific promoters under study differ from nonspecific housekeeping gene promoters in the presence of recognition profiles for 17 transcription factors, as well as a 44-bp motif, which can be promising objects for studying tumor-specific regulation of gene expression. The presence of these sequences in any unknown promoter may be indicative of its tumor specificity. We believe that our results can contribute to the choice of promising native promoters to be used in gene engineering antitumor constructs such as the vectors with killer genes or vector vaccines. Among the seven promoters under study, the *MCM2*, *CKS1B* and *PLK1* promoters are more enriched with the profiles of conditionally specific TFs than other promoters, which is in agreement with the data on considerable tumor specificity of these promoters [12]. Some chimeric promoters that we have obtained previsouly can also be of practical importance [13].

It should be emphasized that the conclusions of this work are based on the theoretical analysis of nucleotide sequences of the promoters. We considered TF recognition profiles only as nucleotide motifs with certain variability. Therefore, all conclusions relate only to the presence of such sequences in the promoters. The actual role of transcription factors in the tumor-specific regulation of promoters should be studied in direct experiments. We hope that our work will contribute to better understanding of how tumor-specific gene transcription works and will open up new opportunities for creating artificial promoters and gene engineering vectors used in the gene therapy for tumors.

ACKNOWLEDGMENTS

FUNDING

COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that they have no conflict of interest. This article does not contain any studies involving animals or human participants performed by any of the authors.

OPEN ACCESS

SUPPLEMENTARY INFORMATION

The online version contains supplementary material available at https://doi.org/10.1134/S1068162022060127.

REFERENCES

1. Zabidi, M.A. and Stark, A., *Trends Genet.*, 2016, vol. 32, pp. 801−814.
https://doi.org/10.1016/j.tig.2016.10.003

2. Corso-Díaz, X., Jaeger, C., Chaitankar, V., and Swaroop, A., *Prog. Retin. Eye Res.*, 2018, vol. 65, pp. 1−27.
https://doi.org/10.1016/j.preteyeres.2018.03.002

3. Gurumurthy, A., Shen, Y., Gunn, E.M., and Bungert, J., *BioEssays*, 2019, vol. 41, p. e1800164.
https://doi.org/10.1002/bies.201800164

4. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T., *Cell*, 2018, vol. 172, pp. 650−665.
https://doi.org/10.1016/j.cell.2018.01.029

5. Castro-Mondragon, J.A., Riudavets-Puig, R., Rauluseviciute, I., Berhanu Lemma, R., Turchi, L., Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Manosalva Pérez, N., Fornes, O., Leung, T.Y., Aguirre, A., Hammal, F., Schmelter, D., Baranasic, D., Ballester, B., Sandelin, A., Lenhard, B., Vandepoele, K., Wasserman, W.W., Parcy, F., and Mathelier, A., *Nucleic Acid Res.*, 2022, vol. 50, pp. D165−D173.
https://doi.org/10.1093/nar/gkab1113

6. Moll, R., *Der Pathologe*, 2018, vol. 39, pp. 285−290.
https://doi.org/10.1007/s00292-018-0518-7

7. Alekseenko, I.V., Snezhkov, E.V., Chernov, I.P., Pleshkan, V.V., Potapov, V.K., Sass, A.V., Monastyrskaya, G.S., Kopantzev, E.P., Vinogradova, T.V., Khramtsov, Y.V., Ulasov, A.V., Rosenkranz, A.A., Sobolev, A.S., Bezborodova, O.A., Plyutinskaya, A.D., Nemtsova, E.R., Yakubovskaya, R.I., and Sverdlov, E.D.,

J. Transl. Med., 2015, vol. 13, p. 78.
https://doi.org/10.1186/s12967-015-0433-0

8. Ginn, S.L., Amaya, A.K., Alexander, I.E., Edelstein, M., and Abedi, M.R., *J. Gene Med.*, 2018, vol. 20, p. e3015.
https://doi.org/10.1002/jgm.3015

9. Watanabe, M., Nishikawaji, Y., Kawakami, H., and Kosai, K.I., *Viruses*, 2021, vol. 13, p. 2502.
https://doi.org/10.3390/v13122502

10. Morse, M.A., Gwin, W.R., 3ᵈ, and Mitchell, D.A, *Target Oncol.*, 2021, vol. 16, pp. 121−152.
https://doi.org/10.1007/s11523-020-00788-w

11. Alekseenko, I.V., Pleshkan, V.V., Kopantzev, E.P., Stukacheva, E.A., Chernov, I.P., Vinogradova, T.V., and Sverdlov, E.D., *PLoS One*, 2012, vol. 7, p. e46474.
https://doi.org/10.1371/journal.pone.0046474

12. Kashkin, K., Chernov, I., Stukacheva, E., Monastyrskaya, G., Uspenskaya, N., Kopantzev, E., and Sverdlov, E., *J. Cell. Biochem.*, 2015, vol. 116, pp. 299−309.
https://doi.org/10.1002/jcb.24968

13. Kashkin, K.N., Chernov, I.P., Didych, D.A., and Sverdlov, E.D., *BioTechniques*, 2017, vol. 63, pp. 107−116.
https://doi.org/10.2144/000114586

14. Takakura, M., Kyo, S., Kanaya, T., Hirano, H., Takeda, J., Yutsudo, M., and Inoue, M., *Cancer Res.*, 1999, vol. 59, pp. 551−557.

15. Xu, R., Zhang, P., Huang, J., Ge, S., Lu, J., and Qian, G., *Biochem. Biophys. Res. Commun.*, 2007, vol. 356, pp. 286−292.
https://doi.org/10.1016/j.bbrc.2007.02.140

16. Gearing, L.J., Cumming, H.E., Chapman, R., Finkel, A.M., Woodhouse, I.B., Luu, K., Gould, J.A., Forster, S.C., and Hertzog, P.J., *PLoS One*, 2019, vol. 14, p. e0215495.
https://doi.org/10.1371/journal.pone.0215495

17. Rodchenkov, I., Babur, O., Luna, A., Aksoy, B.A., Wong, J.V., Fong, D., Franz, M., Siper, M.C., Cheung, M., Wrana, M., Mistry, H., Mosier, L., Dlin, J., Wen, Q., O'Callaghan, C., Li, W., Elder, G., Smith, P.T., Dallago, C., Cerami, E., Gross, B., Dogrusoz, U., Demir, E., Bader, G.D., and Sander, C., *Nucleic Acid Res.*, 2020, vol. 48, pp. D489−D497.
https://doi.org/10.1093/nar/gkz946

18. Tang, Z., Kang, B., Li, C., Chen, T., and Zhang, Z., *Nucleic Acid Res.*, 2019, vol. 47, pp. W556−W560.
https://doi.org/10.1093/nar/gkz430

19. Sverdlov, E.D., Alekseenko, I.V., Bezborodova, O.A., Vinogradova, T.V., Kashkin, K.N., Kopantsev, E.P., Monastyrskaya, G.S., Nemtsova, E.R., Uspenskaya, N.Ya., Chernov, I.P., and Yakubovskaya, R.I., Patent RU 2551784 S1, 2015.

20. Kondratyeva, L.G., Kashkin, K.N., Chernov, I.P., Stukacheva, E.A., Dydich, D.A., Kopantzev, E.P., and Sverdlov, E.D., *Mol. Genet. Microbiol. Virol.*, 2017, vol. 32, pp. 137−140.
https://doi.org/10.3103/S089141681703003X

21. Miyai, M., Hamada, M., Moriguchi, T., Hiruma, J., Kamitani-Kawamoto, A., Watanabe, H., Hara-Chikuma, M., Takahashi, K., Takahashi, S., and Kataoka, K., *J. Invest. Dermatol.*, 2016, vol. 136, pp. 1848−1857.
https://doi.org/10.1016/j.jid.2016.05.088

22. Popowski, M., Lee, B.K., Rhee, C., Iyer, V.R., and Tucker, H.O., *J. Stem Cell Ther. Transplant.*, 2017, vol. 1, pp. 52−62.
https://doi.org/10.29328/journal.jsctt.1001005

23. Elkouby, Y.M., Polevoy, H., Gutkovich, Y.E., Michaelov, A., and Frank, D., *Development*, 2012, vol. 139, pp. 1487−1497.
https://doi.org/10.1242/dev.072934

24. Dekaney, C.M., King, S., Sheahan, B., and Cortes, J.E., *Cell. Mol. Gastroenterol. Hepatol.*, 2019, vol. 8, pp. 549−560.
https://doi.org/10.1016/j.jcmgh.2019.07.003

25. Schredelseker, T. and Driever, W., *Development*, 2018, vol. 145, p. dev163477.
https://doi.org/10.1242/dev.163477

26. Xu, X., Zhang, T., Zhang, M., Li, L., Deng, G., Lu, Z., Zhang, Z., Du, Y., Feng, Y., Feng, X., Peng, X., and Chen, F., *Cell. Signal.*, 2022, vol. 93, p. 110280.
https://doi.org/10.1016/j.cellsig.2022.110280

27. Kramm, K., Engel, C., and Grohmann, D., *Biochem. Soc. Trans.*, 2019, vol. 47, pp. 411−423.
https://doi.org/10.1042/bst20180623

28. Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S., *Nucleic Acid Res.*, 2015, vol. 43, pp. W39−W49.
https://doi.org/10.1093/nar/gkv416

29. Di Giorgio, E., Hancock, W.W., and Brancolini, C., *Biochim. Biophys. Acta Rev. Cancer*, 2018, vol. 1870, pp. 261−273.
https://doi.org/10.1016/j.bbcan.2018.05.007

30. Credendino, S.C., Moccia, C., Amendola, E., D'Avino, G., Di Guida, L., Clery, E., Greco, A., Bellevicine, C., Brunetti, A., De Felice, M., and De Vita, G., *Int. J. Mol. Sci.*, 2020, vol. 22, p. 25.
https://doi.org/10.3390/ijms22010025

31. Huang, Y.H., Klingbeil, O., He, X.Y., Wu, X.S., Arun, G., Lu, B., Somerville, T.D.D., Milazzo, J.P., Wilkinson, J.E., Demerdash, O.E., Spector, D.L., Egeblad, M., Shi, J., and Vakoc, C.R., *Genes Dev.*, 2018, vol. 32, pp. 915−928.
https://doi.org/10.1101/gad.314815.118

32. Park, S.B., Hwang, K.T., Chung, C.K., Roy, D., and Yoo, C., *Clin. Exp. Metastas.*, 2020, vol. 37, pp. 657−674.
https://doi.org/10.1007/s10585-020-10060-0

33. Yang, L., Ke, X.X., Xuan, F., Tan, J., Hou, J., Wang, M., Cui, H., and Zhang, Y., *Cancer Biother. Radiopharm.*, 2016, vol. 31, pp. 44−51.
https://doi.org/10.1089/cbr.2015.1952

34. Guion, L.G. and Sapp, M., *Front Cell. Infect. Microbiol.*, 2020, vol. 10, p. 35.
https://doi.org/10.3389/fcimb.2020.00035

35. le Maire, A., Alvarez, S., Shankaranarayanan, P., Lera, A.R., Bourguet, W., and Gronemeyer, H., *Curr. Top. Med. Chem.*, 2012, vol. 12, pp. 505−527.
https://doi.org/10.2174/156802612799436687

36. Al-Hussain, T., Bakshi, N., and Akhtar, M., *Adv. Anat. Pathol.*, 2015, vol. 22, pp. 202−212.
https://doi.org/10.1097/pap.0000000000000066

37. Petro, T.M., *J. Immunol.*, 2020, vol. 205, pp. 1981−1989.
https://doi.org/10.4049/jimmunol.2000462

38. Kim, H.C., Choi, K.C., Choi, H.K., Kang, H.B., Kim, M.J., Lee, Y.H., Lee, O.H., Lee, J., Kim, Y.J., Jun, W., Jeong, J.W., and Yoon, H.G., *Cell. Mol. Life Sci.*, 2010, vol. 67, pp. 3499−3510.
https://doi.org/10.1007/s00018-010-0388-5

39. Urra, H., Dufey, E., Avril, T., Chevet, E., and Hetz, C., *Trends Cancer*, 2016, vol. 2, pp. 252−262.
https://doi.org/10.1016/j.trecan.2016.03.007

40. Xi, T. and Zhang, G., *Pathol. Res. Pract.*, 2017, vol. 213, pp. 83−88.
https://doi.org/10.1016/j.prp.2016.12.007

41. Alekseenko, I.V., Pleshkan, V.V., Sass, A.V., Filyukova, O.B., Snezhkov, E.V., and Sverdlov, E.D., *Dokl. Biochem. Biophys.*, 2018, vol. 480, pp. 158−161.
https://doi.org/10.1134/s1607672918030092

42. Farokhimanesh, S., Rahbarizadeh, F., Rasaee, M.J., Kamali, A., and Mashkani, B., *Biotechnol. Prog.*, 2010, vol. 26, pp. 505−511.

43. Haberle, V. and Stark, A., *Nat. Rev. Mol. Cell Biol.*, 2018, vol. 19, pp. 621−637.
https://doi.org/10.1038/s41580-018-0028-8

44. Prasad, S., Ramachandran, S., Gupta, N., Kaushik, I., and Srivastava, S.K., *Biochim. Biophys. Acta Mol. Basis D*, vol. 1866, no. is. 2020, p. 165424.
https://doi.org/10.1016/j.bbadis.2019.02.019

45. Dreos, R., Ambrosini, G., Groux, R., Cavin, PérierR., and Bucher, P., *Nucleic Acid Res.*, 2017, vol. 45, pp. D51−D55.
https://doi.org/10.1093/nar/gkw1069

46. Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., Palin, K., Vaquerizas, J.M., Vincentelli, R., Luscombe, N.M., Hughes, T.R., Lemaire, P., Ukkonen, E., Kivioja, T., and Taipale, J., *Cell*, 2013, vol. 152, pp. 327−339.
https://doi.org/10.1016/j.cell.2012.12.009

47. Sjöstedt, E., Zhong, W., Fagerberg, L., Karlsson, M., Mitsios, N., Adori, C., Oksvold, P., Edfors, F., Limiszewska, A., Hikmet, F., Huang, J., Du, Y., Lin, L., Dong, Z., Yang, L., Liu, X., Jiang, H., Xu, X., Wang, J., Yang, H., Bolund, L., Mardinoglu, A., Zhang, C., von Feilitzen, K., Lindskog, C., Ponten, F., Luo, Y., Hökfelt, T., Uhlén, M., and Mulder, J., *Science*, 2020, vol. 367, p. 1090.
https://doi.org/10.1126/science.aay5947

Translated by E. Makeeva