

Functional analysis and identification of *cis*-regulatory elements of human chromosome 21 gene promoters

Hans-Jörg Warnatz^{1,*}, Robert Querfurth¹, Anna Guerasimova¹, Xi Cheng¹, Stefan A. Haas², Andrew L. Hufton^{1,2}, Thomas Manke², Dominique Vanhecke³, Wilfried Nietfeld¹, Martin Vingron², Michal Janitz⁴, Hans Lehrach¹ and Marie-Laure Yaspo^{1,*}

¹Department for Vertebrate Genomics, ²Department for Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, 14195 Berlin, Germany, ³Department of Biomedicine, University Basel, Mattenstrasse 28, 4058 Basel, Switzerland and ⁴School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney NSW 2052, Australia

Received November 26, 2009; Accepted April 29, 2010

ABSTRACT

Given the inherent limitations of *in silico* studies relying solely on DNA sequence analysis, the functional characterization of mammalian promoters and associated *cis*-regulatory elements requires experimental support, which demands cloning and analysis of putative promoter regions. Focusing on human chromosome 21, we cloned 182 gene promoters of 2500 bp in length and conducted reporter gene assays on transfected-cell arrays. We found 56 promoters that were active in HEK293 cells, while another 49 promoters could be activated by treatment of cells with Trichostatin A or depletion of serum. We observed high correlations between promoter activities and endogenous transcript levels, RNA polymerase II occupancy, CpG islands and core promoter elements. Truncation of a subset of 62 promoters to ~500 bp revealed that truncation rarely resulted in loss of activity, but rather in loss of responses to external stimuli, suggesting the presence of *cis*-regulatory response elements within distal promoter regions. In these regions, we found a strong enrichment of transcription factor binding sites that could potentially activate gene expression in the presence of stimuli. This study illustrates the modular functional architecture of chromosome 21 promoters and helps to reveal

the complex mechanisms governing transcriptional regulation.

INTRODUCTION

Gene expression in eukaryotic organisms requires coordinated regulation of thousands of genes. The challenge is to unravel the components and function of complex genetic networks and underlying regulatory processes. Although these processes are integrated at many different levels of the cellular machinery, the regulation of the initiation of transcription is essential and often the rate-limiting step (1) and involves mainly promoter regions located immediately upstream of the gene transcription start sites (TSSs). These regions can integrate various signals to control the transcription rates of associated genes, such as spatial and temporal cues during development, or in response to hormonal, physiological and environmental signals (2). Promoter regions usually entail a core promoter within 50–100 bp surrounding the TSS (3), proximal response elements located up to 250 bp upstream of the TSS, and distal response elements, which can reside several kilobases upstream and downstream of the TSS. The core promoter contains transcription factor binding sites (TFBSs) recognized by the general transcription factors and regulates basal transcription levels, whereas proximal and distal promoter regions are believed to harbor gene-specific TFBSs integrating additional signals for fine-tuning of transcription rates (4).

*To whom correspondence should be addressed. Tel: +49 30 8413 1356; Fax: +49 30 8413 1128; Email: yaspo@molgen.mpg.de
Correspondence may also be addressed to Hans-Jörg Warnatz. Tel: +49 30 8413 1225; Fax: +49 30 8413 1128; Email: warnatz@molgen.mpg.de

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

Mammalian promoter regions have been investigated mainly on a gene-by-gene basis using various reporter gene assays. Attempts to map promoter elements by genome-wide computational analyses of the DNA sequence made use of TFBS predictions and evolutionary conservation of short DNA stretches (5), but these approaches have been limited by the heterogeneous nature of promoter regions, and to some extent by the sequence divergence of regulatory regions among mammals (2). Large-scale experiments addressing the architecture and activity of predicted promoters remain essential. Of the few studies available so far, transient reporter assays testing several hundred putative human promoters detected activity for 60–68% of those in the given experimental conditions (6,7). In a study by the ENCODE consortium, 25% of putative promoters predicted *de novo* from cDNA analysis were reported to be functional (6). These and other studies converge in postulating that mammalian upstream regulatory regions represent a heterogeneous group of modular elements with disparate structural features and cell-type specific activities (2,7,8). Additional systematic experimental analyses are necessary to gain broader insight into promoter structure and function.

We recently established a procedure based on transfected-cell arrays for the functional characterization of promoters (9). Here, we expanded this approach to characterize the promoters of the human chromosome 21 (HSA21) genes, since HSA21 serves as a model for pilot genomic studies due to its small size (48 Mb) and because of its association with Down syndrome. Based on the 231 well-annotated protein-coding genes on HSA21, we cloned 182 promoter fragments of 2.5 kb in size upstream of the predicted TSS, including the TSS itself, as well as a set of shorter fragments (500 bp upstream of the TSS) for 62 promoters. We used transfected-cell arrays (9–15) to carry out promoter reporter assays in HEK293 cells under normal growth conditions and after treatments known to alter gene expression. We correlated the measured activities with the presence of core promoter elements, RNA polymerase II occupancy (16), endogenous transcript levels, and expression profiles derived from EST data for 45 different human tissues. We show that data collected after treatment of cells with different stimuli can provide insight into the presence and identity of functional *cis*-regulatory elements. Taken together, we generated the first chromosome-scale reference data set on the structure, function and responses of human gene promoters.

MATERIALS AND METHODS

Promoter annotation, primer design and vector construction

Promoter annotation and primer design was based on human gene annotations from the Ensembl database v30. Promoter regions were defined relative to the most 5' TSS of all annotated transcripts of a gene. PCR primers for all 231 genes of chromosome 21 were designed using the software PRIDE (17). The optimal target region for

the design was defined as ranging from –2450 to +50 bp relative to the respective TSS. Each primer pair was required to flank the TSS. The downstream part was shortened if an ATG appeared within the +50 bp downstream region. In total, 223 primer pairs were obtained, to which 12 bases of adapter sequences were added for recombination cloning after two-step PCR amplification of the fragments (Gateway technology, Invitrogen). The same approach was used for cloning of truncated promoter fragments of ~500 bp upstream of the TSS.

Promoter cloning

Touch-down PCR from genomic DNA was performed according to a protocol optimized for amplification of GC-rich promoter regions (18), using as templates genomic DNA as well as available genomic BAC and fosmid clones. Then, a secondary PCR was performed with Gateway adapter primers (Invitrogen), followed by PEG-8000 precipitation of PCR products. The modified reporter gene vector pZsGreen1-1 and the control plasmid pHcRed1-N1 were used as described before (9). PCR products were cloned into the pZsGreen vector using Gateway BP Clonase II Enzyme Mix (Invitrogen) and transformed into competent TOP10 cells following the manufacturer's recommendations. Resulting colonies were screened by colony PCR, plasmids were isolated from positive clones using a QIAprep Spin Miniprep Kit (Qiagen), and inserts were confirmed by 5' and 3' end sequencing. Promoter coordinates and primer sequences for 182 cloned promoter fragments are listed in Supplementary Table S1.

Microarray spotting, cell culture and reverse transfection

Samples for array spotting were prepared as previously described (9). Effectene reagent (Qiagen) with Enhancer (Qiagen) was used as transfection agent. Spotting solutions containing 32 ng/μl of promoter construct and 7.5 ng/μl of reference plasmid were kept at 4°C until arraying. Automated spotting was performed with a high-speed non-contact dispensing system (instrumentONE, M2 Automation). Arrays were printed onto home-made poly-L-lysine (Sigma) coated microscope glass slides using a 500 μm outlet port solenoid valve, which delivered 20 nl of sample per spot. Average spot to spot center distance was 1.5 mm. Samples were arrayed in triplicates. After arraying, slides were maintained in low humidity condition at 4°C. Human embryonic kidney cells (HEK293T from ATCC) were cultured in Dulbecco's modified Eagle's medium (DMEM, Gibco Invitrogen) supplemented with 10% (v/v) fetal calf serum (Biochrom) at 37°C in a humidified 6% CO₂ incubator. One day prior to transfection, cells were seeded in a 60 cm² culture plate in 10 ml of medium. On the day of transfection, cells were washed with PBS, detached with Accutase (PAA Laboratories) and seeded at 3.5 × 10⁶ per slide onto printed slides, which were placed into a QuadriPerm chamber (Greiner) for reverse transfection. For each treatment, two slides were used in parallel, so that for each construct, six replicate spots could be analyzed. For treatments after 24 h of incubation at 37°C

with 6% CO₂ in DMEM supplemented with 10% fetal calf serum, the medium was changed to DMEM/FCS with 200 nM Trichostatin A (Sigma) or fetal calf serum-free DMEM. After 48 h of transfection, slides were washed with PBS, fixed in 3.7% formaldehyde with 4M sucrose in PBS for 30 min, stained with DAPI and mounted with Fluoromount-G (Southern Biotech). The slides were kept in the dark at 4°C until analysis.

Image acquisition, object detection and scoring of promoter activity

Microscopy images were acquired and fluorescent objects were detected as previously described (9). The average total number of cells per image frame was 649 cells, as controlled by DAPI staining. In this area, the maximum number of cells that could theoretically be transfected, i.e. cells found in the area of the spotted DNA, was measured to be 370 cells. For each scanning position, the number of HcRed-expressing cells, ZsGreen-expressing cells and co-transfected cells (HcRed- and ZsGreen-positive) was determined. The average transfection efficiency for HcRed alone was 14.1%, while an average of 7.8% of the cells expressed ZsGreen, depending on promoter reporter activity. The average total number of transfected cells (positive for either fluorophore) was 68 cells, resulting in 18.4% combined average transfection efficiency. To determine promoter reporter activities from numbers of fluorescent cells, two selection criteria were taken into account. First, the fraction of green-fluorescent cells among all red cells in a spot had to exceed 16% (transfection threshold). Second, the number of cells both green and red had to exceed the number of cells both green and red in the negative control spots (empty pZsGreen1-1 spotted in 10 replicates) by three standard deviations (reporter activity threshold). A promoter was classified as active if both thresholds were exceeded in at least four out of six replicates on two different cell array slides. Thus, a binary promoter activity index (with 0 for inactive and 1 for active promoters) was generated for each promoter region under investigation.

Computational analyses

We used known position-weight matrices for TATA box, INR and DPE elements (19) together with the TransFac MATCH tool (20) for detection of common promoter motifs under default parameters. Genome-wide coordinates of CpG islands (21) were intersected with the coordinates of cloned promoters to identify CpG islands. In this, we required 500 bp immediately upstream of the TSS to overlap with at least 10% of the total sequence of a CpG island. Genome-wide coordinates of RNA polymerase IIA-bound regions in HEK293 (16) were intersected with the coordinates of cloned promoters to assess occupancy of the hypophosphorylated form of Pol IIA in promoter regions.

We retrieved associations of expressed sequence tag (EST) identifiers to UniGene cluster identifiers in 45 tissues generated for 5 799 931 human ESTs clustered into 116 190 UniGene clusters from the UniGene FTP site (Hs.profiles.gz for Homo sapiens build #207). EST

expression profiles for these UniGene clusters were extracted from the 'Body Sites' category of the original file. The resulting EST set for 156 HSA21 genes consisted of 32 450 ESTs from 45 tissues. We then calculated for each gene with corresponding cloned promoter the number of different tissues where corresponding ESTs could be found.

For the set of promoter sequences that showed specific response patterns in our experiments, we searched for common TFBSs that might explain these responses. To score transcription factor binding, we used a physical affinity-based model described in previous publications (22,23), and matrices describing 610 vertebrate transcription factor binding preferences from TRANSFAC version 12.1 (20). For each binding matrix, we calculated the affinity of the matrix for each sequence, and then transformed these affinities into *P*-values as described before (22). These *P*-values represent the probability that the observed binding affinity is greater than would be expected from a random sequence from a human-promoter-based background model. The *P*-values for each sequence can then be combined using Fisher's method. *P*-values are not multiple test-corrected, and are merely used to rank the factors. Each binding matrix is then ranked according to its combined *P*-value, giving a natural ranking of the transcription factors that have the most enriched binding within the sequence set as a whole.

Enriched gene ontology (GO) terms were identified using the DAVID functional annotation tool (24). Entrez GeneIDs for 40 promoters activated by serum depletion and 28 promoters activated by Trichostatin A were compared to a background set of 126 inactive promoters within the GO category 'biological process'.

RESULTS

Cloning and reporter activity of HSA21 gene promoters in HEK293 cells

Promoter fragments were selected for 231 genes on HSA21, based on their most upstream annotated TSS (see 'Materials and methods' section). Primer pairs encompassing 2.5 kb of DNA sequence upstream of the TSS could be designed for 223 promoter fragments. Of these, 182 fragments were successfully amplified and cloned into a reporter vector upstream of the ZsGreen fluorescent reporter gene. Promoter coordinates, primer sequences and supporting data for each TSS from RNA-seq and Pol IIA ChIP-seq data (16) as well as CAGE and EST/cDNA data (25,26) are listed in Supplementary Table S1. Retrospective comparison with endogenous TSS coordinates determined by RNA-seq from HEK293 indicated that the use of Ensembl TSS annotations yielded slightly more correct TSS positions than the use of predominant TSSs derived from CAGE/EST data from other tissues or cell types. Screenshots from the UCSC genome browser showing all data sets for genes with contradictory TSS data can be found in Supplementary Figure S2. HEK293 cells were co-transfected on cell arrays spotted with promoter reporter constructs and a normalization plasmid expressing red fluorescent protein HcRed. The

combined mean transfection efficiency was calculated to be 18.4% (see 'Materials and methods' section). For assessing reporter gene activity, we measured the ZsGreen and HcRed fluorescence signals in a cell-number-based quantification approach with stringent thresholds ensuring a reliable readout (see 'Materials and methods' section and Supplementary Table S1). The mean numbers of co-transfected cells were 32.9 ± 5.9 and 6.2 ± 4.8 for promoter fragments scored as active and inactive, respectively, whereas we found 1.7 ± 2.0 transfected cells in the negative controls. Figure 1A shows that this scoring scheme allowed a clear-cut distinction to be made between the active cloned promoter fragments (56 out of 182 tested) and the silent fragments (126 out of 182).

Promoter reporter activities correlate with endogenous gene expression

We compared the capacity of the cloned promoters to drive the transcription of a reporter gene with the endogenous transcript levels previously determined for HEK293 cells by RNA-seq (16). For the 56 active promoters, 50 corresponding genes were found expressed (Figure 1B), indicating a rate of 89% true positives in the assay (Supplementary Table S1). In contrast, only 37 of 126 inactive promoters were associated with expressed genes. The enrichment of expressed genes in the set of active promoters was highly significant ($P = 1.2 \times 10^{-14}$).

In order to distinguish promoters of ubiquitously expressed genes from those which might confer tissue specificity, we made use of expression data compiled in UniGene EST clusters (27). We observed a strong correlation between promoter reporter activity in HEK293 cells and the number of different tissues in which the associated genes were transcribed (Figure 1C). Out of 56, 38 active promoters were associated with broadly expressed genes (ESTs found in >25 different tissues). In contrast, only 29 of the 126 inactive promoters controlled genes with a broad expression pattern (Supplementary Table S1). The enrichment of broadly expressed genes among active promoters was highly significant ($P = 1.1 \times 10^{-8}$).

For all genes analyzed, we correlated promoter and gene activities in HEK293 cells with the presence or absence of hallmarks of TSSs and of key regulatory elements. An overview of the different features analyzed for the HSA21 promoters is summarized in Figure 2.

Active promoters are enriched for RNA Pol II-bound regions and activating core promoter elements

To correlate promoter activity with the presence of functional TSSs, we relied on ChIP-seq data that we previously reported for hypophosphorylated RNA polymerase II polypeptide A (Pol II_A) used as a landmark of transcription initiation in HEK293 cells (16). A large fraction of active promoter fragments (35 out of 56) contained or overlapped Pol II_A-bound regions (Figures 2 and 3A), whereas inactive promoter fragments were strongly depleted of Pol II_A (12 of 126; Supplementary Table S1). The enrichment of Pol II_A occupancy in active promoters was highly significant ($P = 2.9 \times 10^{-13}$).

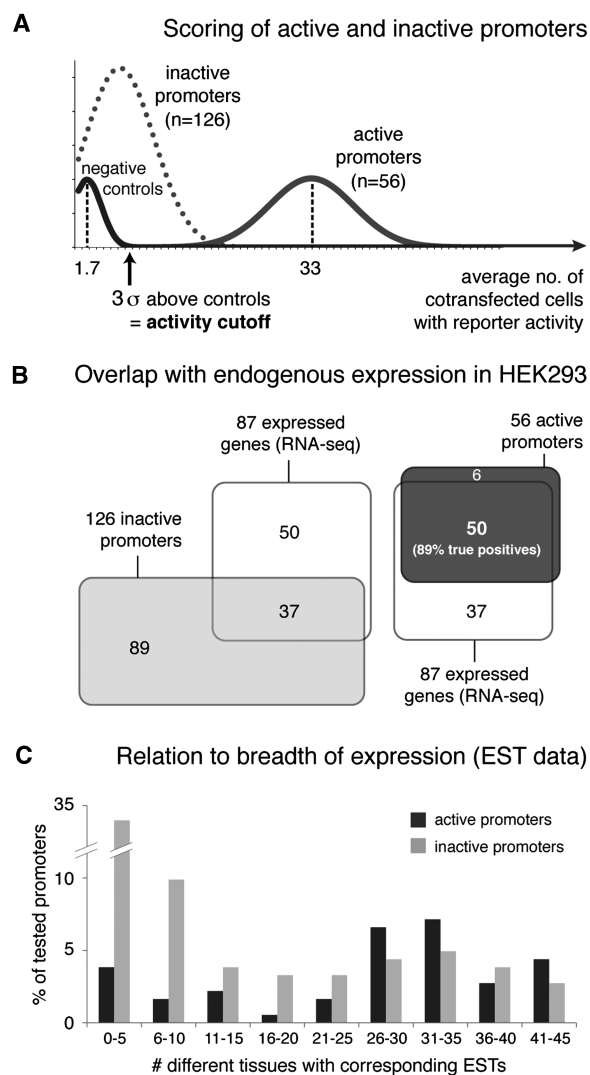


Figure 1. Active promoters can be reliably distinguished from inactive promoters using reporter gene assays on transfected-cell arrays. Promoter reporter activities strongly correlate with endogenous HEK293 gene expression levels and with breadth of expression according to EST data. (A) A promoter was scored active if a cell number threshold for reporter activity (3 SD over the mean of the negative controls) and a co-transfection threshold for the two plasmids (16%) were both exceeded (see 'Materials and methods' section for details). A plot of the mean numbers and standard deviations of co-transfected cells with reporter activity shows the significant differences between active and inactive promoter fragments. (B) Active promoter fragments are mostly derived from genes with endogenous expression in HEK293, as detected by RNA-seq (16). In contrast, most inactive promoters correspond to genes without expression in HEK293. (C) The majority of inactive promoters is derived from genes with restricted expression patterns, whereas most active promoters correspond to genes with broader expression patterns (ESTs found in >25 tissues). For each gene, the number of different human tissues with corresponding ESTs according to the UniGene database was determined and plotted against active and silent promoter fractions.

Core promoters are known to be associated with promoter-specific sequence elements controlling the initiation of transcription of downstream genes. For instance, CpG islands, the TATA box, initiator (INR) and downstream promoter elements (DPE) are functionally important, although their presence is not always required for

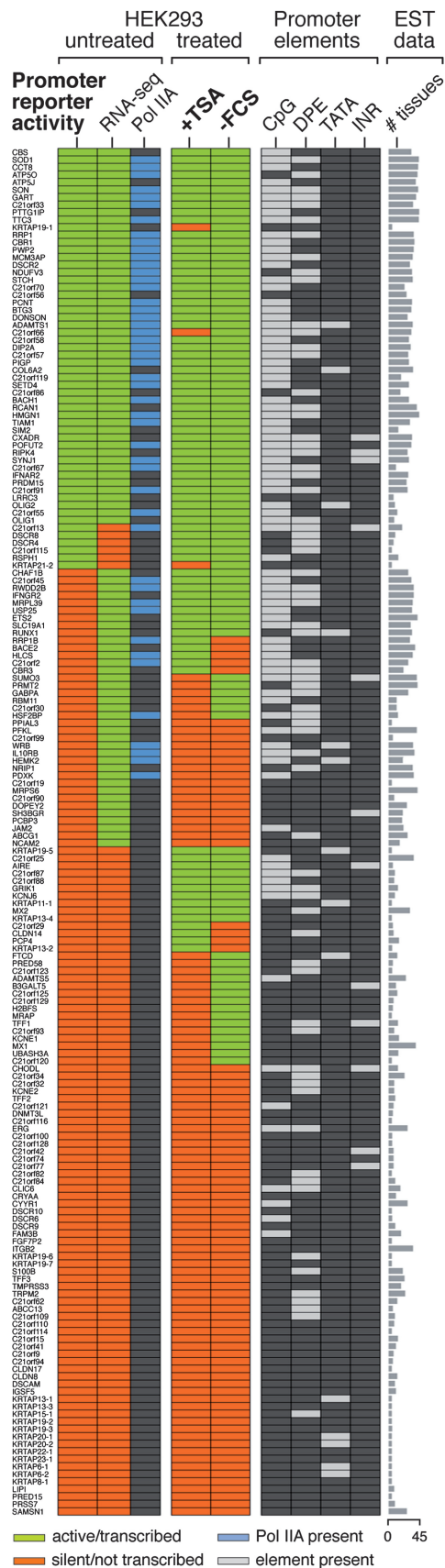


Figure 2. Overview of promoter reporter activities, gene expression data and core promoter elements for 182 tested 2.5 kb promoters. Each row in the panels represents one tested promoter, with gene

promoter activity (7,8,28). We analyzed the occurrence of these four elements within the 500 bp near the TSSs of all 182 cloned promoter fragments (Supplementary Table S1). Based on a genome-wide reference map of CpG islands (21), we observed that 83 of the 182 cloned promoters (46%) overlapped with a CpG island over a sequence length of at least 50 bp, and that these islands were almost always located at the TSS (98% of the cases). Active promoters were highly enriched for CpG islands ($P = 2.3 \times 10^{-11}$) when compared to the silent ones (46 out of 56, and 37 out of 126, respectively). The TATA box, located 28 to 34 bp upstream of the TSS (29), is the best-known core promoter element. TATA boxes are often associated with strong tissue-specific promoters and result in clearly defined TSSs (8). TATA boxes were present in only 14 of the 182 cloned promoters (7.7%), which is slightly below the previously reported portion of TATA-containing promoters among all known human promoters, estimated to be from 10 to 20% (30). TATA boxes occurred almost twice as frequently in silent promoters (9% with TATA) than in active promoters (5% with TATA), but this enrichment was not significant. No trend was observed for the INR element, which was present in 7% of the active fragments (4 of 56) and in 6% of inactive promoters (8 of 126). On the other hand, as shown in Figure 3A, a DPE element was found in half of the active promoters (28 of 56) but in only in one-third of silent promoters (42 of 126; $P = 0.025$; Figure 3A). Three elements occurred together in a significant number of cases. Of the 47 promoters with Pol IIA occupancy, 45 contained a CpG island and 24 a DPE element. Lastly, it is notable that 53 inactive fragments did not overlap with any of the promoter elements or Pol IIA-bound regions.

External stimuli modulate the activities of different sets of promoters

To assess further the functionality of the cloned promoter fragments, we monitored promoter activities after

symbols indicated on the left. The panel on the left shows promoter reporter gene activities for untreated HEK293 cells measured on transfected-cell arrays, corresponding endogenous gene expression (RNA-seq) and RNA polymerase IIA occupancy (Pol IIA). The central panel shows promoter reporter activities after treatment on transfected-cell arrays with Trichostatin A (+TSA) and after depletion of fetal calf serum (-FCS). The panel on the right shows the presence or absence of core promoter elements residing in the cloned fragments. Promoter reporter activity: active promoters driving reporter gene expression are represented by green boxes, inactive promoters by red boxes. RNA-seq: transcripts detected in HEK293; red—no transcripts detected or uncertain; Pol IIA: Blue boxes indicate RNA Polymerase IIA-bound regions in promoters identified by chromatin immunoprecipitation (ChIP-seq) from HEK293 cells (16). Promoter elements: Light gray boxes indicate the presence of CpG islands, downstream promoter elements (DPE), TATA boxes and initiator elements (INR). EST data: The number of different tissues (out of 45) in which corresponding ESTs are present in the UniGene database is represented here as the length of a horizontal bar. All rows are sorted (i) by promoter reporter activity in untreated conditions, (ii) by endogenous expression in HEK293 cells, and (iii) by promoter activities in treated conditions.

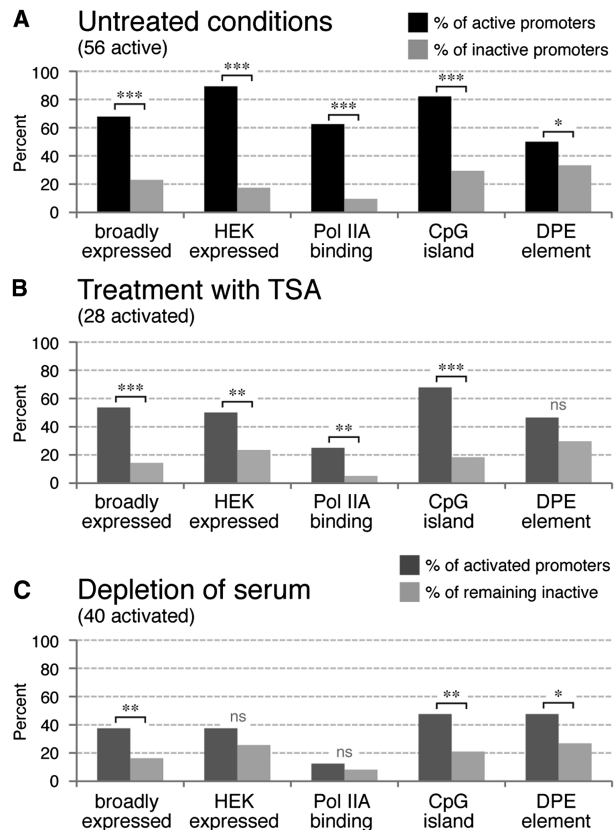


Figure 3. Gene expression features and sequence elements associated with active and silent promoter fragments. (A) The 56 promoters active under untreated conditions are significantly enriched for genes with expression in >25 different tissues (broadly expressed), for genes endogenously expressed in HEK293 cells (HEK expressed), for occupancy of RNA polymerase IIA in the promoter fragment (Pol IIA binding), for presence of CpG islands and for downstream promoter elements (DPE). Dark grey bars represent the fraction of 56 active promoters associated with the indicated feature; light grey bars represent the fraction of 126 promoters inactive in untreated conditions. (B) Treatment of cells with Trichostatin A activates 28 promoters that are significantly enriched for broadly expressed genes, expression in HEK293, Pol IIA binding and CpG islands. (C) Depletion of serum from transfected cells activates 40 promoters that are significantly enriched for broadly expressed genes, CpG islands and DPE elements. Significance levels of enrichments were calculated by the hypergeometric test and are indicated as *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$ and ns: not significant.

challenging the cells by treatment with Trichostatin A (TSA) or by depletion of fetal calf serum (FCS) from the culture medium. The effects of TSA (an inhibitor of class I and II histone deacetylases) on cell function are complex and include triggering the activation of transcription from repressed regions of the chromosomes (31). Here, TSA treatment activated 28 of the 126 inactive promoters, whereas only three of the 56 previously active promoters were silenced (Figure 2). The genes activated by TSA did not belong to any specific category of biological function (data not shown). A large fraction of the TSA-activated genes (15/28) exhibited a broad expression profile (ESTs in >25 tissues), in contrast to those which remained silent (14/98; $P = 5 \times 10^{-5}$, see Figure 3B). Similarly, we observed significant enrichments of

endogenously expressed genes and Pol IIA binding regions (Figure 3B). Interestingly, CpG islands represented the most frequently observed class of elements, found present in 68% of TSA-activated promoters (Figure 3B), whereas no enrichment for either TATA, INR or DPE elements was detected.

Serum depletion from the cell culture medium is known to elicit stress responses and apoptosis through activation of several factors, such as *NFκB* and *CREB* (32–34). Here, after the cells were deprived of serum for 24 h, no promoter was silenced. However, 40 promoters were activated, of which 19 were also found to be activated by TSA (Figure 2). Comparison of the specific features of promoters activated by TSA or serum depletion (Figure 3B and C) showed, in both cases, enrichment of genes with broad expression patterns and with CpG islands, albeit the latter feature was more pronounced among promoters responding to TSA. However, in contrast to the TSA treatment, promoters activated by serum depletion were not enriched for genes endogenously expressed in HEK293 cells or marked by RNA Pol IIA occupancy. Instead, a significant enrichment of DPE elements was observed. CpG islands and DPE elements occurred together in 28% of promoters activated by serum depletion (11/40), but only in 8% of promoters that remained silent (7/86). We found that 20% of the activated promoters corresponded to genes associated with cellular responses to the environment (Supplementary Table S1). In contrast, only 8% of the promoters remaining inactive belonged to this category.

Altogether, monitoring HSA21 promoter gene activity on transfected-cell arrays revealed that 56/182 promoters of 2500 bp in length were able to drive reporter gene expression in HEK293 cells under normal growth conditions. Assays in the presence of different external stimuli showed that an additional 49 promoter fragments have the capacity to induce reporter gene expression. Figure 4 shows the associated EST data from 45 different tissues and gene expression levels in HEK293 cells. Those promoters that were active under standard conditions were expressed in a large number of tissues and exhibited relatively high endogenous expression levels in HEK293 cells. The set of promoters that could be activated by TSA exhibited slightly lower expression levels in both data sets, while the promoters activated through serum depletion tend to be expressed in significantly fewer tissues and showed only weak expression in HEK293 cells. For the 77 silent fragments that could not be activated, only sparse EST data were available, and only 17 promoters were associated with genes expressed in HEK293 cells. A closer inspection of these 17 inactive fragments, together with integration of Pol IIA ChIP-seq and RNA-seq data, revealed that in four cases the core promoter was missed by 10–30 bp (for *C21orf19*, *C21orf90*, *HEMK2* and *PFKL*), and that in five cases an alternative TSS was used for these genes in HEK293 cells (*ABCG1*, *MRPS6*, *NCAM2*, *NRI1* and *PCBP3*). Apart from these cases, the majority of cloned HSA21 promoters recapitulated their function in living cells.

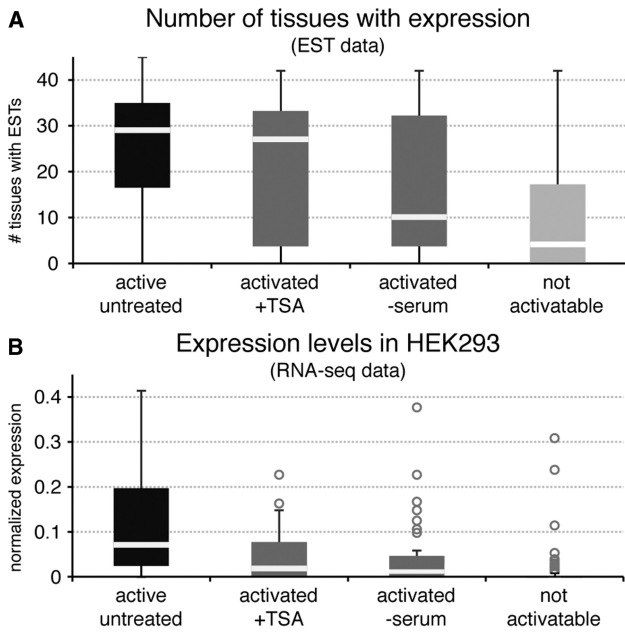


Figure 4. Promoter activities and responses to external stimuli are correlated with endogenous gene expression levels. The distributions of expression values are represented by box-and-whiskers plots for the promoters active in untreated cells ($n = 56$), for the promoters activated by TSA treatment ($n = 28$) or by depletion of serum ($n = 40$) and for the promoters that stay inactive under all tested conditions ($n = 77$). (A) Distributions of the numbers of tissues with expression according to UniGene EST data. (B) Box-and-whiskers plot showing the distribution of normalized expression values in HEK293 according to RNA-seq data (16). Boxes in the plots stretch from the 25th percentile to the 75th percentile of the distribution (inter-quartile range). The median number of tissues is shown as a line across the box. The whiskers stretch to the minimum and maximum data values, unless outliers are present, in which case the whiskers extend to a maximum of 1.5 times the inter-quartile range. Outliers located beyond the whiskers are shown as circles.

Truncation of promoters indicates the presence of distal regulatory elements

To investigate the influence of distal promoter regions on transcription, we cloned a subset of 62 truncated promoter fragments of ~ 500 bp, thus removing the distal ~ 2000 bases. Under standard conditions, 29 of the 62 short fragments could drive transcription in reporter assays (Figure 5). Compared to assays performed with the long fragments, truncation of promoter length resulted in loss of activity for only three cases (*DSCR2*, *OLIG1* and *SIM2*). However, six promoters gained activity in their truncated form, while their longer version was inactive (*MRPL39*, *RBM11*, *CHAF1B*, *HLCS*, *C21orf45* and *SH3BGR*), suggesting the presence of inhibitory regulatory regions in the distal ~ 2000 bp sequences.

Regarding the responses of short fragments to treatments with TSA and depletion of serum, we found that 40 short promoters (66%) recapitulate, under all conditions, the activity patterns observed for the long fragments (Figure 5, lower part), while the remaining 21 behaved differently (Figure 5, upper part). In contrast to their longer counterpart, 14 truncated promoters could not be activated by any treatment, indicating the

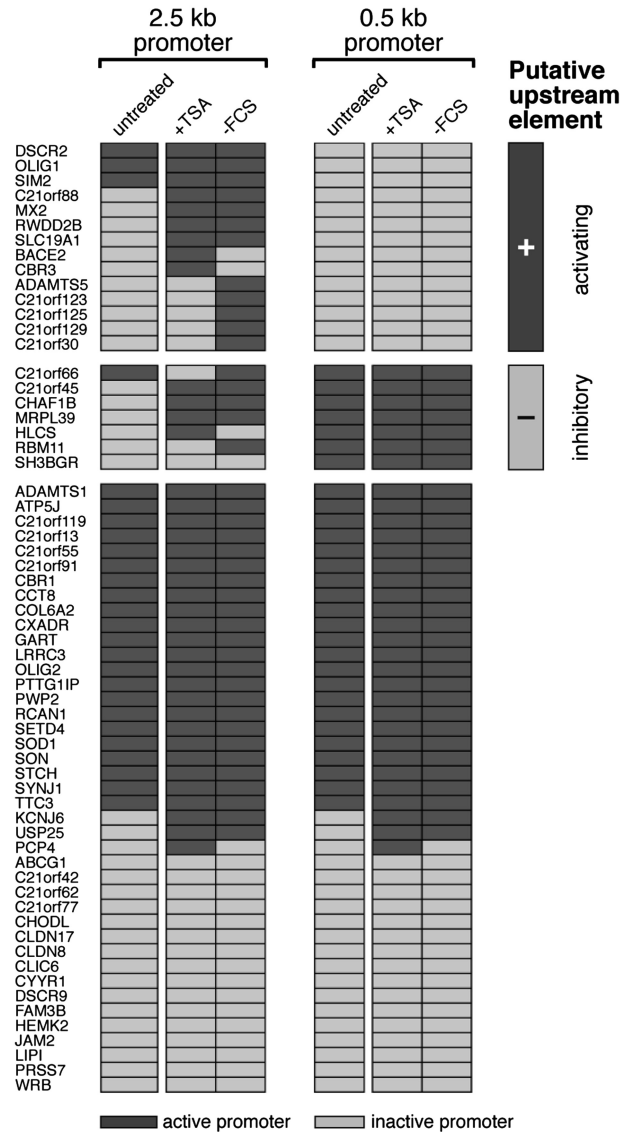


Figure 5. Truncation of promoter fragments can result in loss of responsiveness to external stimuli. The panels show an overview of reporter activities of 62 promoters that were assayed as both long (2.5 kb) and truncated fragments (~ 500 bp). Each row represents one tested pair of long and short promoter, with gene symbols indicated on the left. Active promoters are represented by dark grey boxes, inactive promoters by light grey boxes. Reporter assays were carried out under standard growth conditions (untreated), after treatment of cells with Trichostatin A (+TSA) and after depletion of fetal calf serum (-FCS). Promoters are sorted by the result of truncation, which is either loss of response to external stimuli (21 promoters, upper part) or no change in the response to stimuli (41 promoters, lower part). The presence of upstream *cis*-regulatory elements in distal promoter regions (-2500 to -500 bp) can be inferred from the observed results of truncation, namely activating upstream elements (14 promoters) or inhibitory elements (seven promoters).

loss of activating *cis*-regulatory upstream elements. Conversely, seven truncated promoters could be activated under more conditions than their longer counterparts (*C21orf66*, *C21orf45*, *CHAF1B*, *MRPL39*, *HLCS*, *RBM11* and *SH3BGR*), suggesting that they have lost inhibitory elements located in the distal ~ 2000 bp sequences.

Table 1. Top-enriched transcription factor binding matrices among promoters responding to external stimuli

Promoter response category (no. of promoters; analyzed regions)	Enriched TF binding matrix	Combined <i>P</i> -value	Promoters with motif	Transcription factor	References to stimulus
Activation by serum depletion (<i>n</i> = 40; full 2.5 kb regions)	V\$USF_C	0.000104	48%	<i>USF1</i>	(54)
	V\$OSF2_Q6	0.000155	80%	<i>RUNX2</i>	–
	V\$NFKAPPAB_01	0.00279	85%	<i>NFκB</i>	(32)
	V\$MYC_Q2	0.00355	48%	<i>MYC</i>	(56,63)
Loss of activation by serum depletion after truncation to 500 bp (<i>n</i> = 9; 2 kb distal regions)	V\$NFKAPPAB_01	0.00287	89%	<i>NFκB</i>	(32)
	V\$NERF_Q2	0.007	78%	<i>ELF2</i>	–
	V\$SETS1_B	0.00938	100%	<i>ETS1</i>	(55)
	V\$MAF_Q6	0.00979	89%	<i>MAF</i>	–
Activation by Trichostatin A (<i>n</i> = 28; full 2.5 kb regions)	V\$POU3F2_01	0.00127	86%	<i>OCT7</i>	–
	V\$TCF11MAFG_01	0.00355	93%	<i>MAFG</i>	(57)
	V\$API_Q2	0.00372	71%	<i>API</i>	(61)
	V\$MEF2_01	0.00432	86%	<i>MYEF2</i>	–
Loss of activation by Trichostatin A after truncation to 500 bp (<i>n</i> = 6; 2 kb distal regions)	V\$P53_01	0.015	50%	<i>p53</i>	(58,59)
	V\$OCT1_01	0.0258	67%	<i>OCT1</i>	(60)
	V\$MAF_Q6	0.0345	83%	<i>MAF</i>	–
	V\$TEF1_Q6	0.0396	83%	<i>TEAD1</i>	–

Affinities for 610 known vertebrate TF-binding matrices (20) were ranked in the entire sequences of the promoters activated by serum depletion or TSA treatment, and in the distal 2000 bp of those promoters that lost their activation by stimuli after these regions had been removed by truncation. The *P*-values in each set were combined using Fisher's method and represent the probability that the observed binding affinities are obtained by chance from a random set of sequences from a human-promoter-based background model. The table also lists references describing known connections between the identified TFs and the applied stimuli. All identified binding motifs in the analyzed promoter regions are listed in Supplementary Table S1.

Characterization of *cis*-regulatory elements among promoters responding to external stimuli

We aimed to identify *cis*-regulatory elements that might contribute to the response patterns observed after treatment with external stimuli. We ranked affinities for 610 known vertebrate transcription factor (TF) binding matrices (20) for the 2.5 kb promoters activated by serum depletion or TSA treatment, and for the distal 2000 bp region of promoters which had lost their capacity to be activated by either of those stimuli upon truncation (see 'Materials and methods' section). Of the TF binding matrices enriched in the sequences from each class of promoter, we retained only data pertaining to TFs expressed in HEK293 cells. We found that the highest ranked factors tend to be related to serum responses or TSA treatment (Table 1). All binding sites detected for these TFs in the analyzed promoter fragments are listed in Supplementary Table S1. Results are in accordance with previous reports that demonstrated the importance of *USF1*, *NFκB*, *MYC* and *ETS1* activities in serum response. For the TSA responses and associated histone deacetylase inhibition, we found reports describing responses to TSA treatment for *MAFG*, *API* (*FOS/JUN*), *p53* and *OCT1*. Thus, four of seven TFs with enriched binding sites in serum-sensitive promoters and four of eight TFs with enriched sites in TSA-responsive promoters have been previously implicated in corresponding signal transduction pathways.

DISCUSSION

Using a transfected-cell array procedure, we monitored the activities of 182 cloned promoters corresponding to ~80% of all HSA21 genes in HEK293 cells. Compared to previous studies, where the length of promoters was no

larger than 1000 bp (6,35,36), we aimed here at a more comprehensive coverage of potential upstream regulatory elements by cloning 2500 bp fragments. In addition, we challenged the cells with external stimuli to identify the regulatory nature of the elements in the cloned promoters. In this, the cell array format proved reliable and represents a cost-efficient alternative to conventional reporter gene assays in microtiter plates.

Correlations between promoter reporter activities and endogenous expression

Our data showed that promoter reporter activities recapitulated endogenous gene expression to a great extent (89% concordance). Nevertheless, no transcripts were detected for six genes whose promoters were active on the cell arrays (*C21orf13*, *C21orf115*, *DSCR4*, *DSCR8*, *KRTAP21-2* and *RSPH1*). Four of these genes are expressed in only a few tissues according to EST data. The activity of these promoters in our assay might indicate that elements repressing the expression of those genes in HEK293 were not included in the promoter reporter constructs, or that tight chromatin structures or DNA methylation occurring in the natural genomic context were not recapitulated in the reporter constructs. Conversely, for 37 out of 126 promoters inactive under standard conditions, we found expression of the corresponding genes in HEK293 cells. In fact, the majority of these promoters were activated by treatment with TSA or serum depletion, suggesting that the corresponding cloned fragments contained a functional core promoter but lacked binding sites for activating factors essential to lift reporter gene expression above the detection threshold. The larger part of the remaining 17 inactive promoters missed the core promoter required for transcription; either due to inadequate TSS annotation or use of an alternative TSS in HEK293 cells

(see Supplementary Figures S2). Future promoter studies in HEK293 and other cultured cells will benefit from the wealth of data that is now becoming available from RNA-seq and TSS-seq (37). The TSA-activated promoters generally exhibited broader expression patterns and higher expression levels in HEK293 cells than the promoters activated after serum depletion, suggesting that TSA enhanced the activity of promoters that were already active at low levels, while serum depletion leads to activation of previously inactive promoters.

We observed that the co-occurrence of core promoter elements and Pol IIA occupancy were hallmarks of promoter activity. The finding of significant enrichments of CpG islands and DPE elements, but not TATA or INR elements in active promoters confirmed previous observations (35,38). CpG islands are present within more than 80% of the promoters active in HEK293 cells, a feature previously reported in other cell types, such as primary fibroblasts (39). Moreover, the strong correlation between gene expression levels and promoter activity in our reporter assays concerned mostly genes containing CpG islands. RNA polymerase IIA-binding could either correspond to Pol II stalling at genes poised for activation (40,41), where CpG islands are also known to be enriched (42), or to active TSSs (16,43). As expected, we found the presence of Pol IIA-bound regions strongly associated with active promoter fragments.

Influences of chromatin states and CpG methylation on promoter reporter activity

We observed that a subset of the silent promoters was activated by treatment with Trichostatin A. An expected effect of TSA, a specific inhibitor of mammalian class I and II histone deacetylase enzymes (31), is the activation of transcription from repressed chromosomal regions through chromatin remodeling (44). Transiently transfected plasmids are not entirely subjected to the same regulatory mechanisms that affect native chromatin, but it has been shown that chromatin structures can be formed on plasmid DNA (45). Subsequently, it should be possible to reverse histone deacetylase-dependent silencing mechanisms by TSA (46,47). Indeed, TSA was able to activate 28 of the silent cloned promoters. Apart from the effects of TSA on histone acetylation, it might also be possible that TSA has an indirect influence on CpG methylation. We found that 68% of the TSA-activated promoters contain a CpG island. It has been shown that in some cell lines, TSA downregulates the expression of the DNA methyltransferase DNMT1 (48), and that CpG demethylation occurs at genes with low expression levels, which are then fully activated upon demethylation (49). Taken together, it is likely that the promoter-reporter constructs are sensible to histone deacetylation, or to DNA methylation-mediated silencing (50,51), or to both mechanisms.

Cis-regulatory elements involved in cellular responses to TSA and serum depletion

The comparison of the activities of long and short promoter fragments showed that only three promoters lost their activity upon truncation. This finding implies

that in general, proximal promoter regions are sufficient to drive gene expression. A different picture emerges when activity changes through external stimuli are taken into account. Here, 29% of all tested long fragments changed their activity upon stimulation by TSA or serum depletion, while only 5% of the tested short fragments responded to these stimuli. The observed difference is evidence for the presence of *cis*-regulatory response elements in the distal promoter regions of these genes.

Depletion of serum triggers cell type-specific responses affecting cell cycle regulation, cell growth, differentiation and apoptosis (33,52,53). Indeed, among the 40 promoters activated by serum depletion, one-fifth was annotated within the GO category 'response to stimulus'. Analysis of TFBSs in the promoters of the genes activated by serum depletion showed strong enrichment of binding sites for seven transcription factors. Notably, four of the top-ranking factors (*USF1*, *NFκB*, *ETS1* and *MYC*) have already been shown to be involved in responses to serum starvation. For instance, this treatment enhanced *USF1* expression and binding of *USF1* in the promoter of the target gene lipocalin-type PGD synthase in brain-derived cells (54). *NFκB* has been found potently activated upon serum starvation in HEK293 cells, leading to apoptosis (32). Ets domain-containing transcription factors, such as *ETS1* identified here, are implicated in the response to serum in endothelial cells (55). Finally, *MYC* has been previously implied in responses to growth factor-deprived conditions, where it is involved in induction of apoptosis (56).

Analysis of TFBSs in the promoters of the genes activated by TSA showed strong enrichment of binding sites for eight transcription factors. Remarkably, similar to serum depletion, four of these factors (*MAFG*, *p53*, *OCT1* and *API1*) have previously been shown to be responsive to TSA treatment. TSA can abolish *MAFG*-mediated repression of gene expression via Maf recognition elements in reporter gene assays in HEK293 cells (57), leading to gene activation. TSA can induce *p53*-mediated cell cycle arrest or apoptosis, depending on the cell type (58,59). TSA can also induce gene expression via *OCT1*, independently of *p53* (60). In addition, TSA can promote the binding of *API1* to a recognition site and activate the expression of the osteoporin gene in a mouse mesenchymal cell line (61). The activating effects of TSA, originally shown for specific genes as reported above, seem to represent a more general mechanism targeting the genes whose promoters contain binding sites for these transcription factors.

We observed similarities in the responses of promoters to serum depletion with that to TSA treatment, given that 19 promoters were activated by both types of treatment in our reporter assays. One explanation for this overlap might involve the TSA-responsive *API1* complex, which is composed of members of the *JUN*, *FOS* and *CREB/ATF* families. *JUN* is also a mediator of the *MYC*-induced apoptotic signaling following serum starvation (56). Thus, *JUN* is involved in mediating the responses to both TSA treatment and serum depletion, which might explain the observed overlap between the sets of promoters activated by these stimuli.

CONCLUSION

Using reporter gene assays on transfected-cell arrays, we showed that cloned promoter fragments largely recapitulate gene expression driven by the endogenous promoters, with integration of endogenous signaling pathways into reporter gene expression. Under standard conditions, cloned promoters of ~500 bp in size, spanning core and proximal promoter regions, are generally sufficient to drive gene expression. Comparative analysis of the activities of long and short promoters highlighted the presence of inhibitory *cis*-regulatory elements within -500 to -2500 bp for six promoters. Extended promoter regions were found necessary in many cases to integrate cellular signaling into reporter gene expression. Upon truncation, a significant number of promoters lost their response to cellular signaling. Monitoring the activity of long and short promoter fragments upon stimulation showed evidence for both inhibitory and activating response elements upstream of the 500 bp proximal promoter. Although it has been reported that promoter truncation reveals predominantly the presence of negative regulatory elements within 1000 to 500 bp upstream of the TSS (35,62), we found here that activating distal promoter elements were significantly more frequent than negative elements in our experimental set-up.

This chromosome-scale promoter study sheds some light upon both general and specific regulatory response elements controlling gene activity. The identified promoter activities and responses to stimuli constitute a valuable resource for further investigations. This collection of cloned HSA21 promoters could be used in future promoter activity studies in different cell lines, or in combination with transcription factor overexpression or knock-down, thereby allowing researchers to build a detailed understanding of the mechanisms of transcriptional regulation across an entire chromosome.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Sabine Thamm and Irina Girnus for assistance in preparation of plasmid constructs.

FUNDING

The Max Planck Society for the Advancement of Science, Munich, Germany, and the German Federal Ministry of Education and Research (BMBF) in the framework of the National Genome Research Network 2 (NGFN2, SMP-DNA) [grant number 01GR0414]. Funding for open access charge: The Max Planck Society for the Advancement of Science, Munich, Germany.

Conflict of interest statement. None declared.

REFERENCES

- Asturias, F.J. (2004) Another piece in the transcription initiation puzzle. *Nat. Struct. Mol. Biol.*, **11**, 1031–1033.
- Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V. and Romano, L.A. (2003) The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.*, **20**, 1377–1419.
- Juven-Gershon, T., Hsu, J.Y., Theisen, J.W. and Kadonaga, J.T. (2008) The RNA polymerase II core promoter - the gateway to transcription. *Curr. Opin. Cell. Biol.*, **20**, 253–259.
- Novina, C.D. and Roy, A.L. (1996) Core promoters and transcriptional control. *Trends Genet.*, **12**, 351–355.
- Miller, W., Makova, K.D., Nekrutenko, A. and Hardison, R.C. (2004) Comparative genomics. *Annu. Rev. Genomics Hum. Genet.*, **5**, 15–56.
- Trinklein, N.D., Karaoz, U., Wu, J., Halees, A., Force Aldred, S., Collins, P.J., Zheng, D., Zhang, Z.D., Gerstein, M.B., Snyder, M. *et al.* (2007) Integrated analysis of experimental data sets reveals many novel promoters in 1% of the human genome. *Genome Res.*, **17**, 720–731.
- Muller, F., Demeny, M.A. and Tora, L. (2007) New problems in RNA polymerase II transcription initiation: matching the diversity of core promoters with a variety of promoter recognition factors. *J. Biol. Chem.*, **282**, 14685–14689.
- Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y. and Hume, D.A. (2007) Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat. Rev. Genet.*, **8**, 424–436.
- Cheng, X., Guerasimova, A., Manke, T., Rosenstiel, P., Haas, S., Warnatz, H.J., Querfurth, R., Nietfeld, W., Vanhecke, D., Lehrach, H. *et al.* (2009) Screening of human gene promoter activities using transfected-cell arrays. *Gene*, **450**, 48–54.
- Baghdoyan, S., Roupioz, Y., Pitaval, A., Castel, D., Khomyakova, E., Papine, A., Soussaline, F. and Gidrol, X. (2004) Quantitative analysis of highly parallel transfection in cell microarrays. *Nucleic Acids Res.*, **32**, e77.
- Erfle, H., Neumann, B., Liebel, U., Rogers, P., Held, M., Walter, T., Ellenberg, J. and Pepperkok, R. (2007) Reverse transfection on cell arrays for high content screening microscopy. *Nat. Protoc.*, **2**, 392–399.
- Fiebitz, A., Nyarsik, L., Haendler, B., Hu, Y.H., Wagner, F., Thamm, S., Lehrach, H., Janitz, M. and Vanhecke, D. (2008) High-throughput mammalian two-hybrid screening for protein-protein interactions using transfected cell arrays. *BMC Genomics*, **9**, 68.
- Hu, Y.H., Warnatz, H.J., Vanhecke, D., Wagner, F., Fiebitz, A., Thamm, S., Kahlem, P., Lehrach, H., Yaspo, M.L. and Janitz, M. (2006) Cell array-based intracellular localization screening reveals novel functional features of human chromosome 21 proteins. *BMC Genomics*, **7**, 155.
- Vanhecke, D. and Janitz, M. (2005) Functional genomics using high-throughput RNA interference. *Drug Discov. Today*, **10**, 205–212.
- Ziauddin, J. and Sabatini, D.M. (2001) Microarrays of cells expressing defined cDNAs. *Nature*, **411**, 107–110.
- Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D. *et al.* (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*, **321**, 956–960.
- Haas, S.A., Hild, M., Wright, A.P., Hain, T., Talibi, D. and Vingron, M. (2003) Genome-scale design of PCR primers and long oligomers for DNA microarrays. *Nucleic Acids Res.*, **31**, 5576–5581.
- Ralsler, M., Querfurth, R., Warnatz, H.J., Lehrach, H., Yaspo, M.L. and Krobtsch, S. (2006) An efficient and economic enhancer mix for PCR. *Biochem Biophys. Res. Commun.*, **347**, 747–751.
- Jin, V.X., Singer, G.A., Agosto-Perez, F.J., Liyanarachchi, S. and Davuluri, R.V. (2006) Genome-wide analysis of core promoter elements from conserved human and mouse orthologous pairs. *BMC Bioinformatics*, **7**, 114.
- Wingender, E., Chen, X., Fricke, E., Geffers, R., Hehl, R., Liebich, I., Krull, M., Matys, V., Michael, H., Ohnhauser, R. *et al.* (2001) The

- TRANSFAC system on gene expression regulation. *Nucleic Acids Res.*, **29**, 281–283.
21. Bock, C., Walter, J., Paulsen, M. and Lengauer, T. (2007) CpG island mapping by epigenome prediction. *PLoS Comput. Biol.*, **3**, e110.
 22. Manke, T., Roeder, H.G. and Vingron, M. (2008) Statistical modeling of transcription factor binding affinities predicts regulatory interactions. *PLoS Comput. Biol.*, **4**, e1000039.
 23. Roeder, H.G., Kanhere, A., Manke, T. and Vingron, M. (2007) Predicting transcription factor affinities to DNA from a biophysical model. *Bioinformatics*, **23**, 134–141.
 24. Dennis, G. Jr, Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. and Lempicki, R.A. (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.*, **4**, P3.
 25. Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C. *et al.* (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.*, **38**, 626–635.
 26. Yamashita, R., Suzuki, Y., Wakaguri, H., Tsuritani, K., Nakai, K. and Sugano, S. (2006) DBTSS: DataBase of Human Transcription Start Sites, progress report 2006. *Nucleic Acids Res.*, **34**, D86–D89.
 27. Wheeler, D.L., Church, D.M., Federhen, S., Lash, A.E., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Sequeira, E., Tatusova, T.A. *et al.* (2003) Database resources of the National Center for Biotechnology. *Nucleic Acids Res.*, **31**, 28–33.
 28. Gardiner-Garden, M. and Frommer, M. (1987) CpG islands in vertebrate genomes. *J. Mol. Biol.*, **196**, 261–282.
 29. Ponjavic, J., Lenhard, B., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y. and Sandelin, A. (2006) Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. *Genome Biol.*, **7**, R78.
 30. Gershenzon, N.I. and Ioshikhes, I.P. (2005) Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics*, **21**, 1295–1300.
 31. Khan, N., Jeffers, M., Kumar, S., Hackett, C., Boldog, F., Khramtsov, N., Qian, X., Mills, E., Berghs, S.C., Carey, N. *et al.* (2008) Determination of the class and isoform selectivity of small-molecule histone deacetylase inhibitors. *Biochem. J.*, **409**, 581–589.
 32. Grimm, S., Bauer, M.K., Baeuerle, P.A. and Schulze-Osthoff, K. (1996) Bcl-2 down-regulates the activity of transcription factor NF-kappaB induced upon apoptosis. *J. Cell. Biol.*, **134**, 13–23.
 33. Leicht, M., Briest, W., Holz, A. and Zimmer, H.G. (2001) Serum depletion induces cell loss of rat cardiac fibroblasts and increased expression of extracellular matrix proteins in surviving cells. *Cardiovasc. Res.*, **52**, 429–437.
 34. Li, G., Yang, Q., Krishnan, S., Alexander, E.A., Borkan, S.C. and Schwartz, J.H. (2006) A novel cellular survival factor—the B2 subunit of vacuolar H⁺-ATPase inhibits apoptosis. *Cell Death Differ.*, **13**, 2109–2117.
 35. Cooper, S.J., Trinklein, N.D., Anton, E.D., Nguyen, L. and Myers, R.M. (2006) Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res.*, **16**, 1–10.
 36. Khambata-Ford, S., Liu, Y., Gleason, C., Dickson, M., Altman, R.B., Batzoglou, S. and Myers, R.M. (2003) Identification of promoter regions in the human genome by using a retroviral plasmid library-based functional reporter gene assay. *Genome Res.*, **13**, 1765–1774.
 37. Yamashita, R., Wakaguri, H., Sugano, S., Suzuki, Y. and Nakai, K. (2010) DBTSS provides a tissue specific dynamic view of Transcription Start Sites. *Nucleic Acids Res.*, **38**, D98–D104.
 38. Kim, T.H., Barrera, L.O., Qu, C., Van Calcar, S., Trinklein, N.D., Cooper, S.J., Luna, R.M., Glass, C.K., Rosenfeld, M.G., Myers, R.M. *et al.* (2005) Direct isolation and identification of promoters in the human genome. *Genome Res.*, **15**, 830–839.
 39. Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D. and Ren, B. (2005) A high-resolution map of active promoters in the human genome. *Nature*, **436**, 876–880.
 40. Koch, F., Jourquin, F., Ferrier, P. and Andrau, J.C. (2008) Genome-wide RNA polymerase II: not genes only! *Trends Biochem. Sci.*, **33**, 265–273.
 41. Wu, J.Q. and Snyder, M. (2008) RNA polymerase II stalling: loading at the start prepares genes for a sprint. *Genome Biol.*, **9**, 220.
 42. Hendrix, D.A., Hong, J.W., Zeitlinger, J., Rokhsar, D.S. and Levine, M.S. (2008) Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. *Proc. Natl Acad. Sci. USA*, **105**, 7762–7767.
 43. Core, L.J., Waterfall, J.J. and Lis, J.T. (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, **322**, 1845–1848.
 44. Taddei, A., Roche, D., Bickmore, W.A. and Almouzni, G. (2005) The effects of histone deacetylase inhibitors on heterochromatin: implications for anticancer therapy? *EMBO Rep.*, **6**, 520–524.
 45. Smith, C.L. and Hager, G.L. (1997) Transcriptional regulation of mammalian genes in vivo. A tale of two templates. *J. Biol. Chem.*, **272**, 27493–27496.
 46. Huang, W., Zhao, S., Ammanamanchi, S., Brattain, M., Venkatasubbarao, K. and Freeman, J.W. (2005) Trichostatin A induces transforming growth factor beta type II receptor promoter activity and acetylation of Sp1 by recruitment of PCAF/p300 to a Sp1.NF-Y complex. *J. Biol. Chem.*, **280**, 10047–10054.
 47. Yokota, T., Matsuzaki, Y., Miyazawa, K., Zindy, F., Roussel, M.F. and Sakai, T. (2004) Histone deacetylase inhibitors activate INK4 gene through Sp1 site in its promoter. *Oncogene*, **23**, 5340–5349.
 48. Januchowski, R., Dabrowski, M., Ofori, H. and Jagodzinski, P.P. (2007) Trichostatin A down-regulate DNA methyltransferase 1 in Jurkat T cells. *Cancer Lett.*, **246**, 313–317.
 49. D'Alessio, A.C., Weaver, I.C. and Szyf, M. (2007) Acetylation-induced transcription is required for active DNA demethylation in methylation-silenced genes. *Mol. Cell. Biol.*, **27**, 7462–7474.
 50. Klan, N., Seuter, S., Schnur, N., Jung, M. and Steinhilber, D. (2003) Trichostatin A and structurally related histone deacetylase inhibitors induce 5-lipoxygenase promoter activity. *Biol. Chem.*, **384**, 777–785.
 51. Lande-Diner, L., Zhang, J., Ben-Porath, I., Amariglio, N., Keshet, I., Hecht, M., Azuara, V., Fisher, A.G., Rechavi, G. and Cedar, H. (2007) Role of DNA methylation in stable gene repression. *J. Biol. Chem.*, **282**, 12194–12200.
 52. Cooper, S.J., Trinklein, N.D., Nguyen, L. and Myers, R.M. (2007) Serum response factor binding sites differ in three human cell types. *Genome Res.*, **17**, 136–144.
 53. Leicht, M., Marx, G., Karbach, D., Gekle, M., Kohler, T. and Zimmer, H.G. (2003) Mechanism of cell death of rat cardiac fibroblasts induced by serum depletion. *Mol. Cell. Biochem.*, **251**, 119–126.
 54. Fujimori, K., Aritake, K. and Urade, Y. (2008) Enhancement of prostaglandin D(2) production through cyclooxygenase-2 and lipocalin-type prostaglandin D synthase by upstream stimulatory factor 1 in human brain-derived TE671 cells under serum starvation. *Gene*, **426**, 72–80.
 55. Petrovic, N., Bhagwat, S.V., Ratzan, W.J., Ostrowski, M.C. and Shapiro, L.H. (2003) CD13/APN transcription is induced by RAS/MAPK-mediated phosphorylation of Ets-2 in activated endothelial cells. *J. Biol. Chem.*, **278**, 49358–49368.
 56. Kalra, N. and Kumar, V. (2004) c-Fos is a mediator of the c-myc-induced apoptotic signaling in serum-deprived hepatoma cells via the p38 mitogen-activated protein kinase pathway. *J. Biol. Chem.*, **279**, 25313–25319.
 57. Motohashi, H., Katsuoka, F., Miyoshi, C., Uchimura, Y., Saitoh, H., Francastel, C., Engel, J.D. and Yamamoto, M. (2006) MafG sumoylation is required for active transcriptional repression. *Mol. Cell. Biol.*, **26**, 4652–4663.
 58. Hahold, C., Poehlmann, A., Bajbouj, K., Hartig, R., Korkmaz, K.S., Roessner, A. and Schneider-Stock, R. (2008) Trichostatin A causes p53 to switch oxidative-damaged colorectal cancer cells from cell cycle arrest into apoptosis. *J. Cell Mol. Med.*, **12**, 607–621.
 59. Roy, S., Packman, K., Jeffrey, R. and Tenniswood, M. (2005) Histone deacetylase inhibitors differentially stabilize acetylated p53

- and induce cell cycle arrest or apoptosis in prostate cancer cells. *Cell Death Differ.*, **12**, 482–491.
60. Hirose,T., Sowa,Y., Takahashi,S., Saito,S., Yasuda,C., Shindo,N., Furuichi,K. and Sakai,T. (2003) p53-independent induction of Gadd45 by histone deacetylase inhibitor: coordinate regulation by transcription factors Oct-1 and NF-Y. *Oncogene*, **22**, 7762–7773.
61. Sakata,R., Minami,S., Sowa,Y., Yoshida,M. and Tamaki,T. (2004) Trichostatin A activates the osteopontin gene promoter through AP1 site. *Biochem. Biophys. Res. Commun.*, **315**, 959–963.
62. Negi,S., Singh,S.K., Pati,N., Handa,V., Chauhan,R. and Pati,U. (2004) A proximal tissue-specific module and a distal negative regulatory module control apolipoprotein(a) gene transcription. *Biochem. J.*, **379**, 151–159.
63. Tamura,K., Hua,B., Adachi,S., Guney,I., Kawauchi,J., Morioka,M., Tamamori-Adachi,M., Tanaka,Y., Nakabeppu,Y., Sunamori,M. *et al.* (2005) Stress response gene ATF3 is a target of c-myc in serum-induced cell proliferation. *EMBO J.*, **24**, 2590–2601.