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

Hans-Jörg Warnatz1,*, Robert Querfurth1, Anna Guerasimova1, Xi Cheng1, Stefan A. Haas2, Andrew L. Hufton1,2, Thomas Manke2, Dominique Vanhecke3, Wilfried Nietfeld1, Martin Vingron2, Michal Janitz4, Hans Lehrach1 and Marie-Laure Yaspo1,*

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

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 \( /C24 \) 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.

© The Author(s) 2010. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
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 reporter assays testing several hundred putative human promoters. 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 pHeRed1-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 4 M 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 HeRed-expressing cells, ZsGreen-expressing cells and co-transfected cells (HeRed- and ZsGreen-positive) was determined. The average transfection efficiency for HeRed 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 II A in promoter regions.

We retrieved associations of expressed sequence tag (EST) identifiers to UniGene cluster identifiers in 45 tissues generated for 5799931 human ESTs clustered into 116190 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 32450 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 II A 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 HeRed. 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 ± 0.2 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 ($P = 1.1 \times 10^{-3}$).

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 A-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 ($P = 2.9 \times 10^{-17}$).

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...
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 II A 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 II A-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...
endogenously expressed genes and Pol II A 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 NFXb and CREB (32–34). Here, after the cells were deprived of serum for 24h, 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 II A 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 II A 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, NRIP1 and PCBP3). Apart from these cases, the majority of cloned HSA21 promoters recapitulated their function in living cells.
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). 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. (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 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.
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 binding matrices (TFs) in the analyzed promoter fragments as markers of serum responses or transcription factors essential to lift reporter gene expression. For the TSA responses and associated histone acetyl transferase (HAT) activities in 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.

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.

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

| Promoter response category | Enriched TF binding matrix | Combined P-value | Promoters with motif | Transcription factor | References to stimulus |
|---------------------------|----------------------------|-----------------|----------------------|---------------------|-----------------------|
| Activation by serum depletion | VSUSF_C | 0.000104 | 48% | USF1 | (54) |
| (n = 40; full 2.5 kb regions) | VSOSF2_Q6 | 0.000155 | 80% | RUNX2 | |
| | VSNFKAPPAB_01 | 0.00279 | 85% | NFkB | (32) |
| | VSMYC_Q2 | 0.00355 | 48% | MYC | (56,63) |
| Loss of activation by serum depletion after truncation to 500 bp | VSNFKAPPAB_01 | 0.00287 | 89% | NFkB | (32) |
| (n = 9; 2 kb distal regions) | VSNERF_Q2 | 0.007 | 78% | ELF2 | |
| Activation by Trichostatin A | VSetsi_B | 0.00938 | 100% | ETS1 | (55) |
| (n = 28; full 2.5 kb regions) | VSMAF_Q6 | 0.00979 | 89% | MAF | |
| | VSPOUF3F2_01 | 0.00127 | 86% | OCT7 | |
| | VSTCF11MAFG_01 | 0.00355 | 93% | MAFG | (57) |
| | VSAPI_Q2 | 0.00372 | 71% | API1 | (61) |
| | VSMEF2_01 | 0.00432 | 86% | MYEF2 | |
| Loss of activation by Trichostatin A after truncation to 500 bp | VSP53_01 | 0.015 | 50% | p53 | (58,59) |
| (n = 6; 2 kb distal regions) | VSOCT1_01 | 0.0258 | 67% | OCT1 | (60) |
| | VSMAF_Q6 | 0.0345 | 83% | MAF | |
| | VSTEF1_Q6 | 0.0396 | 83% | TEAD1 | |

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.

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.
Influences of chromatin states and CpG methylation on promoter reporter activity

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.

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, NFXB, 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). NFXB 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 API) 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 API 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 API 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 late gene expression driven by the endogenous promoters, showed that cloned promoter fragments largely recapitulate gene expression. Under standard conditions, with integration of endogenous signaling pathways into late gene expression driven by the endogenous promoters, showed that cloned promoter fragments largely recapitulate 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

1. Asturias,F.J. (2004) Another piece in the transcription initiation puzzle. Nat. Struct. Mol. Biol., 11, 1031–1033.
2. 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.
3. Juven-Gershon,T., Hu,S.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.
4. Novina,C.D. and Roy,A.L. (1996) Core promoters and transcriptional control. Trends Genet., 12, 351–355.
5. Miller,W., Makova,K.D., Nekrutenko,A. and Hardison,R.C. (2004) Comparative genomics. Annu. Rev. Genomics Human. Genet., 5, 15–56.
6. Trinklein,N.D., Karaouz,E., Wu,J., Hales,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.
7. 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.
8. Sandelin,A., Carninci,P., Lenhard,B., Porajovic,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.
9. Cheng,X., Guerasimova,A., Manke,T., Rosenstiel,P., Haas,S., Warnatz,H.J., Querfurth,R., Niefeld,W., Vanhecke,D., Lehrauf,H. et al. (2009) Screening of human gene promoter activities using transfected-cell arrays. Gene, 450, 48–54.
10. Buggdoyan,S., Roupior,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.
11. Erle,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.
12. Fiebitz,A., Nyarsik,L., Haendler,B., Hu,Y.H., Wagner,F., Thamm,S., Lehrauf,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.
13. Hu,Y.H., Warnatz,H.J., Vanhecke,D., Wagner,F., Fiebitz,A., Thamm,S., Kahlen,P., Lehrauf,H., Yaso,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.
14. Vanhecke,D. and Janitz,M. (2005) Functional genomics using high-throughput RNA interference. Drug Discov. Today, 10, 205–212.
15. Ziauddin,J. and Sabatini,D.M. (2001) Microarrays of cells expressing defined cDNAs. Nature, 411, 107–110.
16. 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.
17. 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.
18. Thaller,M., Querfurth,R., Warnatz,H.J., Lehrauf,H., Yaso,M.L. and Kröbetsch,S. (2006) An efficient and economic enhancer mix for PCR. Biochem. Biophys. Res. Commun., 347, 747–751.
19. Jin,V.X., Singer,G.A., Agosto-Perez,F.J., Liyanarachchi,S. and Duvvuri,R.V. (2006) Genome-wide analysis of core promoter elements from conserved human and mouse orthologous pairs. BMC Bioinformatics, 7, 114.
20. Wingender,E., Chen,X., Frick,E., Gefers,R., Hehl,R., Liebich,I., Krull,M., Matys,V., Michael,H., Ohnhauser,R. et al. (2001) The
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., Wagakuri,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. Gershenson,N.I. and Ioshkhes,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. de Azevedo,J.A., Bauer,M.K., Baeuerle,P.A. and Schulze-Osthoff,K. (1996) Bel-2 down-regulates the activity of transcription factor NF-kappaB induced upon apoptosis. J. Cell. Biol., 134, 13–23.

33. Leicht,M., Briest,W., Holzl,A. and Zimmer,H.G. (2001) Serum depletion induces cell loss of rat cardiac fibroblasts induced by serum depletion. J. Mol. Cell. Cardiol., 33, 127–136.

34. Li,G., Yang,Q., Krishnan,S., Alexander,E.A., Borkan,S.C. and Leicht,M., Briest,W., Holzl,A. and Zimmer,H.G. (2003) Mechanism of cell death of rat cardiac fibroblasts induced by serum depletion. Mol. Cell. Biochem., 251, 265–273.

35. 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.