Screening of human gene promoter activities using transfected-cell arrays

Xi Cheng [a,c], Anna Guerasimova [a,1], Thomas Manke [b], Philip Rosenstiel [d], Stefan Haas [b], Hans-Joerg Warnatz [a], Robert Querfurth [a], Wilfried Nietfeld [a], Dominique Vanhecke [a,2], Hans Lehrach [a], Marie-Laure Yaspo [a], Michal Janitz [a,*]

[a] Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 73, 14195 Berlin, Germany
[b] Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Ihnestr. 73, 14195 Berlin, Germany
[c] Free University Berlin, Department of Biology, Chemistry and Pharmacy, Takustr. 3, 14195 Berlin, Germany
[d] Institut for Clinical Molecular Biology, Christian-Albrechts-University, Schittenhelmstr. 12, 24105 Kiel, Germany

ARTICLE INFO

ABSTRACT

Promoters are the best characterized transcriptional regulatory sequences in complex genomes because of their predictable location immediately upstream of transcription start sites. Despite a substantial body of literature describing transcriptional promoters, the identification of true start sites for all human transcripts is far from complete. The same is true of the key structural and functional elements responsible for promoter action in different cell types. In order to identify elements responsible for promoter activity, we applied transfected-cell array technology to functionally evaluate promoters for genes involved in inflammatory bowel disease. Seventy-four promoters were examined by reverse transfection of a promoter-fluorescent reporter constructs into a human embryonic kidney cell line (HEK293T). Sixteen (21.6%) promoters were found to be active in HEK293 T cells. Correlations between promoter activity and endogenous transcript level were calculated, and 75% of active promoters were found to be associated with transcriptional activity of their gene counterparts. These results provide experimental evidence of promoter activity, which may aid in understanding the regulation of gene expression. Moreover, this is the first large-scale functional study of regulatory sequences to use a high-throughput transfected-cell array technique.

1. Introduction

Gene expression in eukaryotes is known to be an extraordinarily complex process involving a variety of steps, including transcription, RNA processing, mRNA stabilization, translation and protein modification. Regulation of transcriptional initiation plays a crucial role in this cascade, and several DNA sequence elements are known to contribute to transcriptional initiation. Promoters, located immediately upstream of transcriptional start sites (TSSs), are likely one of the most important elements in this regulation.

Promoters consist of two parts, including the core promoter and the extended promoter regions. The core promoter is a small contiguous DNA sequence that interacts directly with components of the basal transcription machinery and initiates transcription. This region is usually located within 150 bp of the TSS (Sandelin et al., 2007; Müller et al., 2007). Extended promoter regions, located between hundreds and thousands of bp away from the TSS, contain other *cis*-acting DNA elements, such as the insulator, enhancer and silencer.

The availability of the genome sequences of many eukaryotic organisms has made global characterization of gene regulation one of the major fundamental goals of biomedical research. Intensive efforts have recently been invested in developing sequence-based and hybridization-based approaches for identifying TSSs and promoters on a genome-wide scale (Cooper et al., 2006; ENCODE Project Consortium, 2007). As a result, a large number of novel transcripts and promoters have been revealed. However, the function of these promoters *in vivo* remains unknown without sufficient experimental data. Functional studies are mainly performed using reporter gene assays, which require large numbers of transfection experiments and, thus, limit genome-wide characterization of promoters.

In the present work, we describe the application of high-throughput transfected-cell arrays (Vanhecke and Janitz, 2004) in a functional study of eukaryotic promoters. Seventy-four promoter fragments for genes potentially involved in the pathomechanism of inflammatory bowel

disease were evaluated. Inflammatory bowel disease (IBD) is triggered by still unknown polygenic factors. However, recent technological advances have helped to reveal the genetic etiology and transcriptomic landscape of such complex diseases. Combining murine models with patient-derived data has substantially contributed to understanding the early events of etiopathogenesis. Application of this approach has revealed that a primary deficiency of the innate immunological barrier function exists in the group of diseases that includes inflammatory bowel disease (Rosenstiel and Schreiber, 2008). Here, promoter activity patterns were correlated with the expression profiles of corresponding genes. Furthermore, a comparative analysis between core regulatory elements and promoter activity profiles in human embryonic kidney (HEK) cells was performed.

2. Materials and methods

2.1. Promoter prediction and PCR amplification with attB primers

TSSs were defined using Ensembl 35. In cases without Ensembl gene annotation, CAGE tags (FANTOM3) and EST-sequences (GeneNest database) were used to determine putative TSSs. Putative promoter regions were defined as the sequences directly upstream of TSSs.

The software PRIDE (http://pride.molgen.mpg.de/pride.html) (Haas et al., 2003) was used to design primer sets to amplify 2.5 kb DNA fragments upstream of and flanking the TSS of each putative promoter. First, touchdown PCR of genomic DNA was performed using the protocol described by Ralser et al. (2006). To generate PCR products suitable for use as substrates in Gateway BP recombination reactions, 12 bases of attB site sequences were added to the 5′ end of each primer (attB1 forward: 5′-AAAAAGCAGGCTNN-3′; attB2 reverse: 5′-AGAAAGCTGGGTN-3′). Then, a second PCR was performed to attach the complete attB sequences to the ends of the PCR products using adapter primers (attB1 adapter: 5′-GGGGACAAGTTTGTA-CAAAAAAGCAGGCT-3′; attB2 adapter: 5′-GGGGACCACTTTGTACAA-GAAAGCTGGGT-3′) (Gateway Technology Instruction Manual, Invitrogen).

2.2. Vector construction, promoter cloning and plasmid preparation

pZsGreen1-1 (Clontech) is a promoterless vector encoding the green fluorescent protein ZsGreen1. To enable rapid and highly efficient cloning of the promoter fragment into the vector, pZsGreen1-1 was converted to a Gateway-adapted vector using restriction cloning. To do so, the attP1- and attP2-flanked fragment from pDONR221 (Invitrogen) was inserted into the multiple cloning site of the pZsGreen1-1 vector.

To clone the amplified promoter fragments, purified attB-PCR products were mixed with the attP-compatible pZsGreen1-1 vector and incubated with Gateway BP Clonase II Enzyme Mix (Invitrogen) for over 3 h at 25 °C. Each reaction volume was 10 µl of the total volume, and the reaction component amounts were calculated following the Gateway technology instruction manual. After incubation with Protein-ase K for 10 min at 37 °C, the mixture was transformed into One Shot TOP10 competent cells (Invitrogen). The resulting colonies were screened by PCR, and plasmids were isolated from positive clones using a QIAprep Spin Miniprep Kit (Qiagen). The identity of the inserts was confirmed by sequencing the 5′ and 3′ ends.

pHcRed1-N1, a CMV promoter-driven vector expressing the far-red fluorescent protein HcRed1 (Clontech), was modified and used as a control for spot localization and transfection efficiency. The CMV promoter, with a total length of 589 bp, contains a 407 bp enhancer region and a 58 bp minimum promoter region. In order to make the control plasmid comparable to the putative promoters lacking the enhancer element, 323 bp of the enhancer region was removed by restriction digestion, resulting in a 266 bp-long 3′ section of the original CMV promoter upstream of the HcRed reporter gene.

2.3. Sample preparation and microarray spotting

Samples for array spotting were prepared using the lipid–DNA method (LD-method) described by Ziauddin and Sabatini (2001) with modifications (Fiebitz et al., 2008). Briefly, 1.6 µg of putative promoter plasmids and 400 ng of the reference HcRed reporter plasmid were mixed, and the volume was adjusted with TE buffer to 6.5 µl. EC buffer (10 µl, Effectene Kit, Qiagen) supplemented with 0.2 M sucrose (Invitrogen) was added to the plasmid mixture, which was incubated with 1.5 µl Enhancer (Qiagen) and then with 7.0 µl Effectene reagent (Qiagen) at room temperature for 5 min and 10 min, respectively. Finally, the samples were mixed with 25 µl of 0.2% (w/v) gelatin and incubated at room temperature for a minimum of 1 h before spotting. The final sample mixture contained 32 ng/µl of promoter construct and 7.5 ng/µl of reference plasmid.

Standard $25 \times 75 \times 1.0$ mm slides covered with poly-L-lysine (Sigma) and VECTABOND Reagent (Vector Labs) were used for array printing. The samples were spotted manually with a 2 µl pipette and long tips (SAFESEAL tips PREMIUM XL 10 µl, Biozmy) by tapping the slide with the filled tip. The samples were spotted following a printed 1.5×1.5 mm grid, such that the distance between adjacent spots was approximately 1.5 mm. The diameter of each spot was between 0.8 and 1.0 mm, and the estimated volume was 10–15 nl. All samples were arrayed in triplicate. After spotting, the slides were air-dried for a minimum of 1 h before transfection. For long-term use, slides were stored in dark, low-humidity conditions at 4 °C.

2.4. Cell culture and reverse transfection

HEK 293T/17 (ATCC, CRL-11268), a cultured HEK cell line, was used for reverse transfection. Cells were cultured in high glucose Dulbecco's modified Eagle's medium (DMEM) (Gibco Invitrogen) supplemented with 10% (v/v) fetal calf serum (Biochrom), 2 mM L-glutamine (Gibco Invitrogen) and 50 µg/ml penicillin/streptomycin (Gibco, Invitrogen). The cells were incubated at 37 °C in humid conditions with 6% CO_2.

Twenty-four hours before transfection, 10^7 cells were plated onto a 60 cm^2 cell culture dish in 10 ml complete growth media. On the day of transfection, cells were harvested and seeded at 3.5×10^6 per slide in 7.5 ml complete media in a Quadriperm chamber (Greiner). The transfected arrays were incubated at 37 °C with 6% CO_2 for 48 h. Two slides were used in parallel. To stop the transfection, slides were fixed for 30 min in 3.7% formaldehyde with 4 M sucrose in PBS. After staining with DAPI (4′,6-diamidino-2′-phenylindole dihydrochloride, SIGMA) and mounting with Fluoromount-G (Southern Biotech), slides were covered with 22×64 mm cover glasses and scanned with a microscope.

2.5. Image acquisition and analysis

All images were acquired using an automated fluorescence microscope (IX81, Olympus) with custom image acquisition software Scan^R (Olympus). The microscope was equipped with a high-resolution CCD camera, and a 10×/1.6 (NA) objective was used. For each spotting position, single frame images were acquired using DAPI staining for auto-focusing and red and green channels for fluorescence signal detection. Background reduction was achieved using the rolling circle option for image processing.

Numerical output from the image sets was generated by the Scan^R Analysis software, an integrated image analysis and data visualization tool (Olympus). Scan^R Analysis identified each trans-fected cell by locating the borders of fluorescent objects in the images based on the object-finding algorithm. The watershed option was used for improved object separation. Cells were considered co-transfected with both reporters when the fluorescence intensity

signals from the ZsGreen1 and HcRed1 reporter constructs exceeded the set thresholds in the green and red channels, respectively.

In order to identify the promoter activity at each spotting position, two selection criteria were taken into account, including the number of co-transfected cells (Nco-tr) and the ratio (Ri) between the number of co-transfected cells and the total number of red cells (Nred) (Ri = Nco-tr / Nred). The threshold for the number of co-transfected cells was defined as three standard deviations above the mean number of ZsGreen1-positive cells transfected with negative controls (pZsGreen1-1 vector modified with pDONR sites) spotted in 10 replicates. The Ri parameter served as a transfection efficiency index and was empirically set to 0.1. This value provided the best correlation of promoter activity with the corresponding gene expression level, as determined by transcriptome sequencing. If the Ri parameter was greater than 0.1 in at least four out of six replicates, the promoter was classified as active. A binary promoter activity index, with 0 indicating inactive and 1 indicating active, was generated for each investigated promoter region.

2.6. Gene expression profiling using the RNA-Seq technique

Publicly available data on gene expression levels in HEK293T cells are provided at http://promotion.molgen.mpg.de/cgi-bin/gbrowse/Hs.Solexa. The detailed procedure used for RNA-Seq transcriptome profiling using Illumina Genome Analyzer has previously been described (Sultan et al., 2008).

2.7. Comparison of promoter activities with UniGene expression profiles

The T-STAG database of tissue-specific gene expression was used to compare tissue-specific promoter activities. This database, which is based on UniGene, associates each gene with a set of p-values describing the likelihood of the gene being expressed in the respective tissue (Gupta et al., 2005). UniGene clusters of the genes for the cloned promoters were identified, and the tissue-specific expressed sequence tag (EST) profile was obtained. For each gene, the number of tissues where ESTs were expressed was determined and compared with the fraction of active and silent fragments.

2.8. Computational sequence analysis of promoter regions

The predicted promoter sequences were analyzed, and common promoter elements (TATA box and CpG islands) were detected. The TransFac MATCH tool (Kel et al., 2003) was used to detect TATA box regions using the position weight matrix TATA_01 (Bucher, 1990). The sequences of cloned fragments were compared to the map of CpG islands predicted by Bock et al. (2007), and the recommended combined epigenetic score (greater than 0.5) was used.

3. Results

3.1. Cloning of 74 promoters for selected genes

Potential TSSs were identified, and the 2.5-kb sequence upstream of each predicted TSS was selected as the promoter fragment. Seventy-four putative promoters (listed in Supplementary Table 1) were successfully amplified from human genomic DNA using the previously described PCR protocol designed for amplification of GC-rich promoter regions (Ralser et al., 2006). The putative promoter regions were cloned into the attP-adapted pZsGreen1-1 vector upstream of the fluorescent reporter gene ZsGreen1. The insertions were verified by sequencing from both the 5′ and 3′ ends.

3.2. Identification of 16 functional promoters in HEK293T cells

Using transfected-cell technology, cloned promoter fragments were evaluated for activity in HEK cells, with reporter gene expression used as the measure of activity (Fig. 1). A cell-based quantification approach was used to determine promoter activity. For each scanning position, the number of cells transfected with the test promoter constructs (showing green fluorescence), the control plasmid HcRed1-1 (showing red fluorescence) or co-transfected (simultaneously showing green and red fluorescence) was determined. Promoter activity was assessed based on two thresholds, including the minimal number of co-transfected cells and the minimal transfection ratio index. A promoter fragment was considered active if it exceeded the combined threshold for at least four replicates out of six. Using this approach, 16 (21.6%) functional promoters in HEK293T cells were identified for the genes listed in Table 1.

The transfection efficiency for each spot on the array was calculated as the ratio of the number of transfected cells to the total number of cells in the spot area (Fig. 2). The number of transfected cells was calculated as the sum of cells positive for either green or red fluorescence; cells that expressed both fluorophores were counted only once. The average number of transfected cells was approximately 100 cells per spot. The total number of cells per scan position estimated by DAPI staining was approximately 1000. According to the area ratio, the number of cells that could theoretically be transfected was estimated to be approximately 580. Thus, the transfection efficiency was approximately 17.2%, and this value is in agreement with a previous work utilizing liposome-based reverse transfection, in which the efficiency was slightly over 17% (Baghdoyan et al., 2004).

3.3. Endogenous transcript levels correlated with promoter activity

The activity of cloned promoter fragments was compared with endogenous transcript levels in the same cell line to identify the correlation between activity and expression level. This analysis sought to determine how the ability of a promoter to initiate gene transcription is influenced by distal cis-regulatory elements and post-transcription regulatory mechanisms because neither of these elements could be monitored in the cell array-based functional assay. Comparison of promoter activity with the levels of corresponding mRNA provides the ratio of false positive and negatives.

Endogenous transcript levels were determined previously by performing high-throughput sequencing of the HEK293T transcriptome using Illumina Genome Analyzer (Sultan et al., 2008). Thus, expression levels of the genes corresponding to cloned promoters have been reported. Among the 16 active promoter fragments, 12 (75.0%) of the corresponding genes were expressed. In contrast, of the 58 silent promoter fragments, only 17 (29.3%) of the corresponding genes were expressed (Fig. 3, Table 1 and Supplementary Table 1).

3.4. Active promoters are enriched for genes with ubiquitous expression

HEK293T were used for reverse transfection experiments, so promoters for genes with more tissue-specific activity might not be functional in this cell line. In contrast, genes with more ubiquitous expression would be more likely to be expressed in the HEK cell line. Therefore, the association between transfected promoter activity and the tissue expression profiles of the corresponding genes was evaluated. For this purpose, UniGene EST clusters of genes corresponding to the studied promoter fragments were retrieved, and the number of tissues containing each EST was counted. Subsequently, the tissue distribution of the ESTs was analyzed in relation to the promoter activity of the corresponding gene. Using this approach, tissue expression data was obtained for 67 genes, including 16 active fragments and 51 silent fragments. A strong correlation was found between the promoter activity state in HEK293T cells and the number

of tissues with corresponding ESTs. In 68.8% of active promoters, the corresponding genes were expressed in more than 25 out of 45 tissues. In contrast, only 12.1% of the inactive promoters corresponded to genes with a broad expression profile. Furthermore, 50.0% of the inactive promoters were expressed in less than 10 tissues (Fig. 4).

3.5. Correlation of activity pattern with promoter sequence characteristics

Core promoters vary in their sequence structure; however, several key DNA sequence elements can be distinguished. These motifs are necessary for the binding of general transcription factors and RNA polymerase II in the pre-initiation complex (Sandelin et al., 2007; Müller et al., 2007). CpG islands and TATA boxes are of particular interest due to the epigenetic effects of the former and the necessity of the latter for initiation of the transcription process (Gross and Oelgeschläger, 2006). Therefore, the present promoter set was analyzed for correlations between the presence of these two sequence motifs and promoter activity status.

CpG islands are genomic regions with high GC content and a high frequency of CpG dinucleotides relative to the bulk DNA (Gardiner-Garden and Frommer, 1987). They are associated with the promoters of most housekeeping genes as well as many tissue-specific genes and play a critical role in gene regulation. Aberrant methylation of promoter-related CpG islands is a common mechanism of inactivation of tumor suppressor genes and is associated with epigenetic causes of all types of cancer (Herman and Baylin, 2003). Various approaches have been described and developed for computational prediction and identification of CpG islands (Takai and Jones, 2002; Wang and Leung, 2004; Hackenberg et al., 2006). Most of these approaches are based
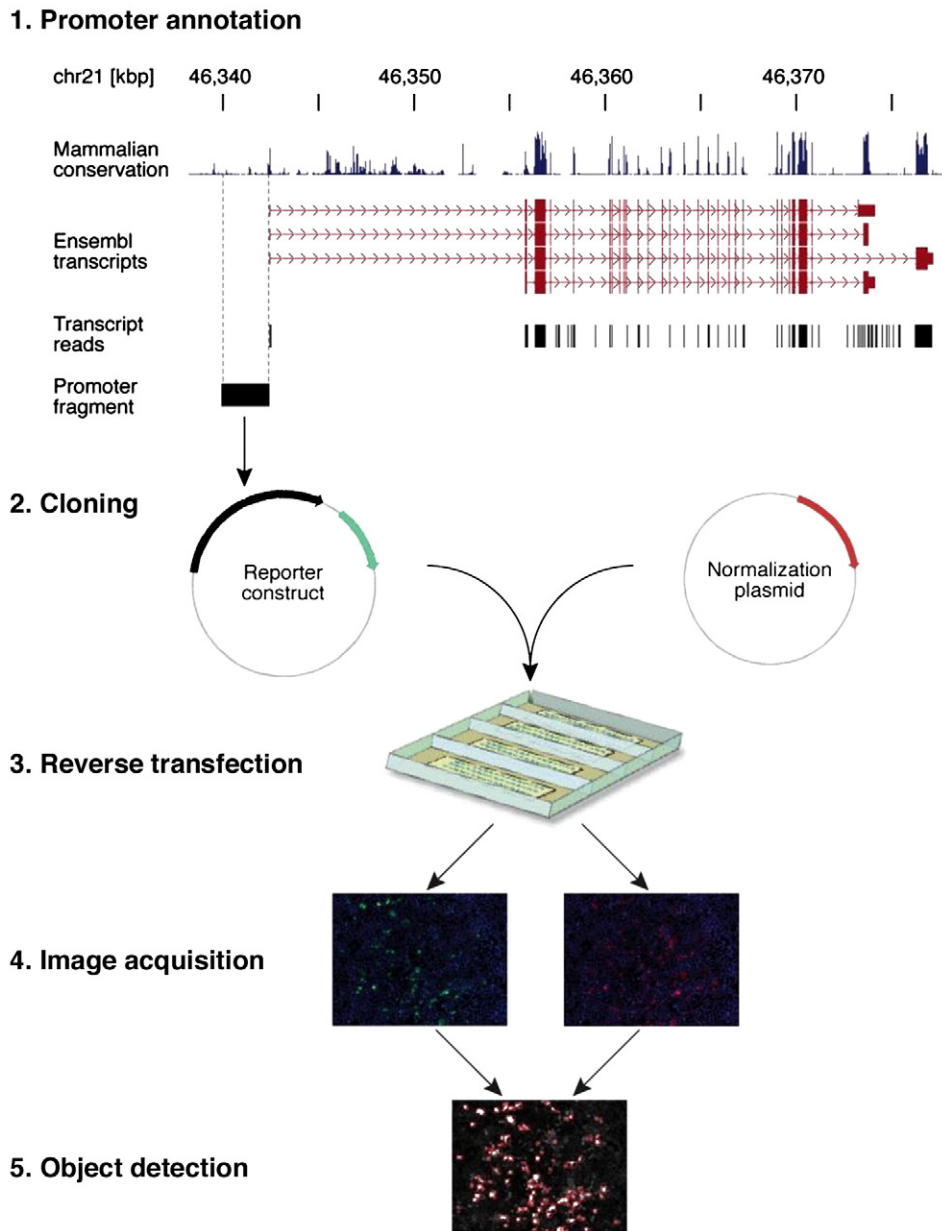

Fig. 1. Workflow for promoter analysis using a transfected-cell array. (1) Promoter regions were selected upstream of transcription start sites from the Ensembl database and amplified by PCR. (2) Amplified promoters were cloned into a reporter plasmid encoding a green fluorescent protein. (3) Reporter constructs were spotted onto glass slides together with a normalization plasmid encoding a red fluorescent protein. Subsequently, the slides were covered with HEK293T cells for reverse transfection. (4) Images were acquired by fluorescence microscopy. (5) Data were analyzed by counting the number of transfected cells using an object-finding algorithm.

Table 1
List of genes of which promoters revealed activity in the transfected-cell array.

No.	Gene name	Gene symbol	Entrez ID	Gene expression
1	Androgen receptor	AR	367	+
2	Activity-regulated cytoskeleton-associated protein	ARC	23237	+
3	Baculoviral IAP repeat-containing 2	BIRC2	329	+
4	Baculoviral IAP repeat-containing 3	BIRC3	330	+
5	Baculoviral IAP repeat-containing 5 (survivin)	BIRC5	332	+
6	Caspase recruitment domain family, member 8	CARD8	22900	+
7	CASP2 and RIPK1 domain containing adaptor with death domain	CRADD	8738	+
8	Dystrobrevin binding protein 1	DTNBP1	84062	+
9	Interferon induced with helicase C domain 1	IFIH1	64135	−
10	NACHT, leucine rich repeat and PYD containing 2	NALP2	55655	−
11	NACHT, leucine rich repeat and PYD containing 4	NALP4	147945	−
12	Receptor-interacting serine-threonine kinase 2	RIPK2	8767	+
13	Receptor-interacting serine-threonine kinase 4	RIPK4	54101	+
14	Toll-like receptor 1	TLR1	7096	−
15	Toll-like receptor 5	TLR5	7100	+
16	TNFRSF1A-associated via death domain	TRADD	8717	+

only on sequence criteria. Here, a high-quality map of so-called bona fide CpG islands in the human genome was used to characterize the CpG islands in the promoter fragments. This map was generated using a CpG island strength prediction method that links the DNA sequence of CpG islands to their epigenetic states, including DNA methylation, histone modification and chromatin accessibility (Bock et al., 2007). CpG islands were detected in 14 (18.9%) out of 74 cloned promoters, including 7 (43.8%) of the 16 active promoters and 7 (12.1%) of the 58

silent promoters (Fig. 5). A 3.6-fold enrichment of CpG islands in active versus silent fragments was observed. These data agree with the description by Cooper et al. (2006), stating that the presence of CpG islands is an important but incomplete indicator of promoter activity.

The TATA box, the best known core promoter element, is located approximately 30 bp upstream of the TSS. The TATA binding protein (TBP) binds to this motif and begins assembly of the pre-initiation complex. TATA boxes are often associated with tissue-specific promoters and are correlated with a clearly defined TSS (Schug et al., 2005; Sandelin et al., 2007). TATA boxes were found in 26 (35.1%) out of 74 cloned promoters. This fraction is higher than the value previously reported by Juven-Gershon et al. (2008), who reported that less than one-third of vertebrate core promoters contained a TATA box. In the present work, 37.5% of the active fragments and 34.5% of silent fragments contained a TATA box (Fig. 5).

4. Discussion

In this study, transfected-cell array (TCA) technology is presented as an efficient and cost-effective tool for functional analysis of predicted promoter regions. Putative promoter regions were located directly upstream of TSSs predicted by Ensembl gene annotations. For functional characterization, a 2.5-kb promoter fragment was selected to include the nearest TSS of the gene. In a previous study by Cooper et al. (2006), the 300-bp sequence upstream of the TSS contributed positively to core promoter activity, while sequences 500 to 1000 bp upstream of the TSS were identified as negative elements in half of the cases. Therefore, the 2.5-kb fragments analyzed here likely contained the full core promoter element as well as more upstream regulatory sequences.

The cloned fragments were tested for promoter activity in the HEK293T cell line using the TCA technique. This method allows parallel analysis of multiple genes and gene products. Furthermore, this method has been used to screen phenotypes via gene silencing (Vanhecke and Janitz, 2005), detect protein intracellular localization
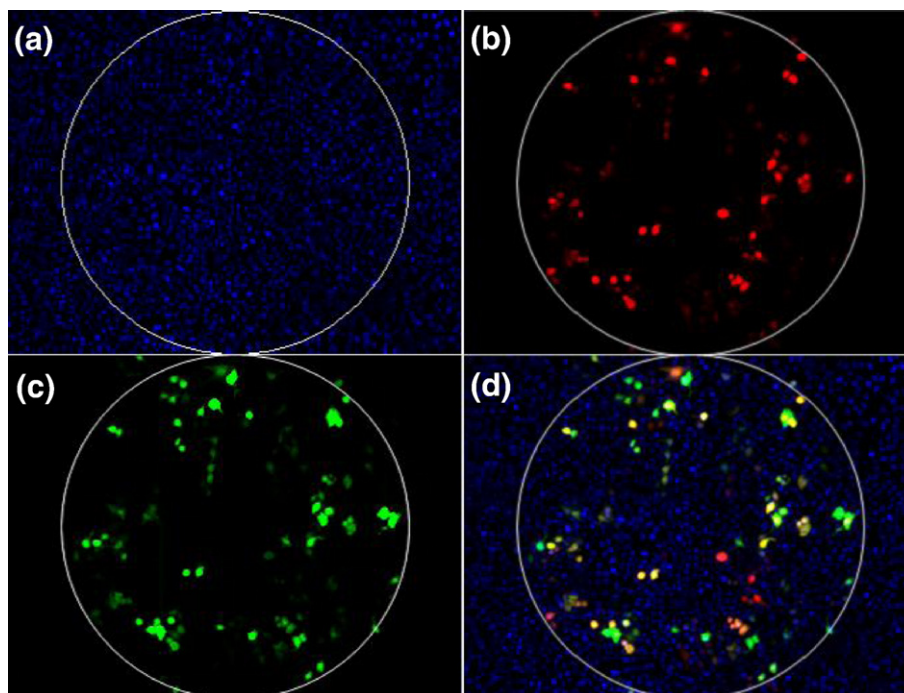

Fig. 2. Cell number counting and transfection efficiency calculation. The transfection efficiency for each spot on the array was calculated as the ratio of the number of transfected cells to the total number of cells in the spot area. (a) The total number of cells per scan position was counted by fluorescent DAPI. The number of cells that could theoretically be transfected (within the spot) was estimated according to the area ratio. The numbers of cells expressing red (b), green (c), or dual fluorescence (d), shown in yellow, were automatically counted. The number of transfected cells was calculated as the sum of cells positive for either green or red fluorescence; cells that expressed both fluorophores were counted only once.

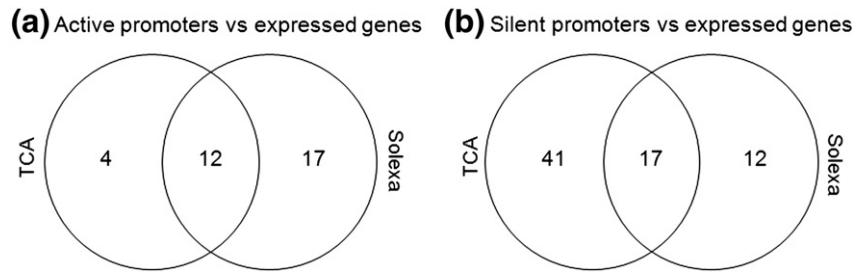**(a)** Active promoters vs expressed genes   **(b)** Silent promoters vs expressed genes

Fig. 3. Correlation of promoter activity with endogenous gene expression in HEK293T cells. The Venn diagrams show the degree of agreement between the promoter reporter activities of 74 tested fragments (indicated as TCA) and the expression data for their associated genes, which was obtained from high-throughput transcriptome sequencing (indicated as Solexa). (a) Active promoter fragments had a 75.0% overlap with corresponding gene expression. (b) Genes associated with silent promoter fragments were less likely to be expressed (29.3%).

(Hu et al., 2006) and investigate protein–protein interactions (Fiebitz et al., 2008). The present study demonstrates another application of the TCA technique, namely, a large-scale functional analysis of transcriptional regulatory elements in living cells.

Overall, 21.6% of the predicted promoters showed activity in the HEK293T cell line with the transient reverse transfection assay. This fraction is lower than but comparable to the 30.8% promoter activity observed in a set of chromosome 21 promoters (unpublished data).

In order to assess the promoter contribution to gene expression in living cells, the promoter activities were compared to endogenous mRNA levels of the corresponding genes in HEK293T cells. Of the promoters that were active in the TCA assay, 75% of the corresponding genes were expressed. No corresponding transcripts were detected for 4 promoters, including *IFIH1*, *NALP2*, *NALP4* and *TLR1*. IFIH1, also known as the melanoma differentiation-associated 5 (MDA-5), is an RNA helicase of which expression is induced during differentiation, cancer reversion, and apoptosis (Kang et al., 2002, 2004). The level of *MDA-5* mRNA remains low in normal tissues, however IFN-β induces its expression in a number of normal and cancer cells (Kang et al., 2004). *NALP2* has been shown to modulate an inflammatory response by inhibiting the NF-κB signaling pathway. *NALP2* mRNA was detected by RT-PCR to be expressed at low levels in HEK293T cells. Its expression is however upregulated by NF-κB activators (Fontalba et al., 2007). According to the UniGene profile, *NALP4* is expressed at low levels in less than 10 tissues. NALP4, also known as NLRP4, has been suggested to play a role in controlling inflammatory responses by suppressing TNFα- and IL-1β-induced NF-κB activation in HEK293 cells (Fiorentino et al., 2002). Moreover, *NALP4* has been shown to

have a very low expression level in human embryos (Zhang et al., 2008). TLR1, in conjunction with TLR2, has been shown to recognize triacylated lipopeptides (Takeuchi et al., 2002; Alexopoulou et al., 2002) and lead to activation of pro-inflammatory cytokine secretion (Sandor et al., 2003). The mRNA levels for *TLR1* remain low, as determined previously using RT-PCR (Kurt-Jones et al., 2004; Farhat et al., 2008), which is consistent with our findings. The low endogenous mRNA levels of the *IFIH1*, *NALP2*, *NALP4* and *TLR1* genes might suggest the presence of distal silencing elements acting on the respective promoters, which were not present in the transfected plasmid promoter constructs. The *NALP4* has transcript variants with alternative TSSs, which may further contribute to the discrepancy between elevated promoter activity and absence of the respective transcript observed in our study. It should be emphasized that the cell array-based assay for high-throughput promoter activity screening should be treated as a pre-selection platform. Promoter candidates with interesting activity patterns can be selected for further studies using, for example, deletional analysis or transfection into non-adherent cells, since these cells cannot be applied to reverse transfection on cell arrays.

For a number of genes that are expressed at relatively high levels in HEK293T cells, the corresponding promoters were inactive in the TCA assay. Several explanations might account for this discrepancy. First, due to the complexity of transcriptional initiation mechanisms and the difficulty of TSS prediction, non-functional or alternative TSSs of tissue-specific transcript variants might have been selected during fragment design. Second, the presence of negative regulatory elements might suppress the activity of some promoters to an undetectable level or totally inactivate the promoter. On the other hand, the positive elements required for initiation of transcription may be located outside of the cloned fragment. Furthermore, experimental noise in promoter activity measurements might lead to an underestimation of activity and thus contribute to the observed differences. Nevertheless, the analytic efficiency of the cell array due to its high-throughput compensates for the relatively high false negative rate.
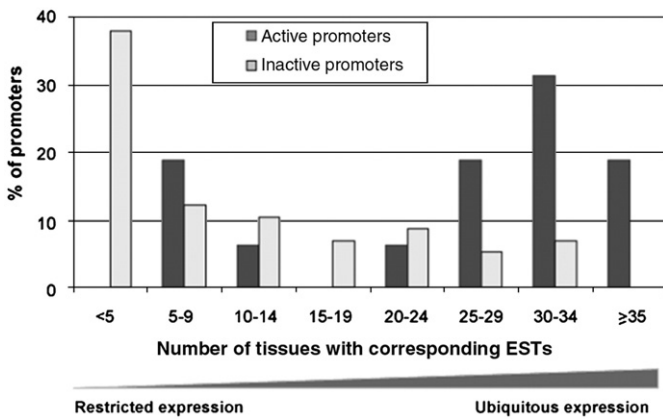

Fig. 4. Tissue expression profiles for active and silent promoter fragments. EST profiles from 45 tissues for genes corresponding to cloned promoters were obtained from the UniGene database. For each gene, the number of tissues in which ESTs have been observed was counted, and bins of these EST tissue numbers were plotted against the fraction of each active and silent promoter. The corresponding genes of 68.8% of the active promoters were expressed in more than 25 out of 45 tissues, while 50.0% of the inactive promoters were expressed in less than 10 tissues.
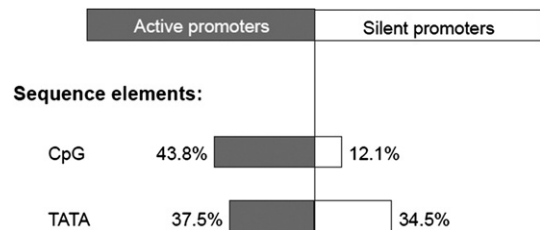

Fig. 5. The presence of core promoter sequence elements in cloned promoters. The promoter sequence motifs TATA box and CpG islands were detected in the cloned promoters. Among the active promoters, 43.8% and 37.5% contained the CpG islands and TATA box, respectively. Among the silent promoters, 12.1% contained CpG islands and 34.5% contained a TATA box.

The presence of promoter sequence motifs was compared with promoter activity, and enrichment of CpG islands was found in active promoters. However, no obvious enrichment of TATA boxes was observed in active promoters. Promoter sequence features were also compared with the tissue-specific expression of the corresponding genes. Of the 32 genes that were expressed in less than 10 tissues, 17 (53.1%) promoters had a TATA box and only 3 (9.4%) had CpG islands. In contrast, among the 9 genes expressed in more than 30 tissues, 3 (33.3%, *DTNBP1*, *CARD8*, and *IFIH1*) and 5 (55.6%, *BIRC2*, *BIRC5*, *RIPK2*, *PTK9L*, and *DTNBP1*) promoters had a TATA box and CpG islands, respectively. These observations confirm the previous description by Schug et al. (2005), reporting that TATA box-containing promoters are commonly associated with tissue-specific genes, while CpG islands are found in more ubiquitously expressed genes.

In summary, this study demonstrates that the TCA technique, together with gene expression and bioinformatics tools, represents a robust technology for large-scale functional studies of regulatory sequences.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2009.10.003.

References

Alexopoulou, L., Thomas, V., Schnare, M., Lobet, Y., Anguita, J., Schoen, R.T., Medzhitov, R., Fikrig, E., Flavell, R.A., 2002. Hyporesponsiveness to vaccination with Borrelia burgdorferi OspA in humans and in TLR1- and TLR2-deficient mice. Nat. Med. 8, 878–884.

Baghdoyan, S., Roupioz, Y., Pitaval, A., Castel, D., Khomyakova, E., Papine, A., Soussaline, F., Gidrol, X., 2004. Quantitative analysis of highly parallel transfection in cell microarrays. Nucleic Acids Res. 32, e77.

Bock, C., Walter, J., Paulsen, M., Lengauer, T., 2007. CpG island mapping by epigenome prediction. PLoS Comput Biol. 3, e110.

Bucher, P., 1990. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J. Mol. Biol. 212, 563–578.

Cooper, S.J., Trinklein, N.D., Anton, E.D., Nguyen, L., Myers, R.M., 2006. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. Genome Res. 16, 1–10.

ENCODE Project Consortium, 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799–816.

Farhat, K., Sauter, K.S., Brcic, M., Frey, J., Ulmer, A.J., Jungi, T.W., 2008. The response of HEK293 cells transfected with bovine TLR2 to established pathogen-associated molecular patterns and to bacteria causing mastitis in cattle. Vet. Immunol. Immunopathol. 125, 326–336.

Fiebitz, A., Nyarsik, L., Haendler, B., Hu, Y.H., Wagner, F., Thamm, S., Lehrach, H., Janitz, M., Vanhecke, D., 2008. High-throughput mammalian two-hybrid screening for protein–protein interactions using transfected-cell arrays. BMC Genomics 9, 68.

Fiorentino, L., Stehlik, C., Oliveira, V., Ariza, M.E., Godzik, A., Reed, J.C., 2002. A novel PAAD-containing protein that modulates NF-kappa B induction by cytokines tumor necrosis factor-alpha and interleukin-1beta. J. Biol. Chem. 277, 35333–35340.

Fontalba, A., Gutierrez, O., Fernandez-Luna, J.L., 2007. NLRP2, an inhibitor of the NF-kappaB pathway, is transcriptionally activated by NF-kappaB and exhibits a nonfunctional allelic variant. J. Immunol. 179, 8519–8524.

Gardiner-Garden, M., Frommer, M., 1987. CpG islands in vertebrate genomes. J. Mol. Biol. 196, 261–282.

Gross, P., Oelgeschläger, T., 2006. Core promoter-selective RNA polymerase II transcription. Biochem. Soc. Symp. 73, 225–236.

Gupta, S., Vingron, M., Haas, S.A., 2005. T-STAG: resource and web-interface for tissue-specific transcripts and genes. Nucleic Acids Res. 33 (Web Server issue), W654–W658.

Haas, S.A., Hild, M., Wright, A.P., Hain, T., Talibi, D., Vingron, M., 2003. Genome-scale design of PCR primers and long oligomers for DNA microarrays. Nucleic Acids Res. 31, 5576–5581.

Hackenberg, M., Previti, C., Luque-Escamilla, P.L., Carpena, P., Martínez-Aroza, J., Oliver, J.L., 2006. CpGcluster: a distance-based algorithm for CpG-island detection. BMC Bioinformatics 7, 446.

Herman, J.G., Baylin, S.B., 2003. Gene silencing in cancer in association with promoter hypermethylation. N. Engl. J. Med. 349, 2042–2054.

Hu, Y.H., Warnatz, Y.J., Vanhecke, D., Wagner, F., Fiebitz, A., Thamm, S., Kahlem, P., Lehrach, H., Yaspo, M.L., Janitz, M., 2006. Cell array-based intracellular localization screening reveals novel functional features of human chromosome 21 proteins. BMC Genomics 7, 155.

Juven-Gershon, T., Hsu, J.Y., Theisen, J.W., Kadonaga, J.T., 2008. The RNA polymerase II core promoter — the gateway to transcription. Curr. Opin. Cell Biol. 20, 253–259.

Kang, D.C., Gopalkrishnan, R.V., Wu, Q., Jankowsky, E., Pyle, A.M., Fisher, P.B., 2002. mda-5: An interferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties. Proc. Natl. Acad. Sci. U. S. A. 99, 637–642.

Kang, D.C., Gopalkrishnan, R.V., Lin, L., Randolph, A., Valerie, K., Pestka, S., Fisher, P.B., 2004. Expression analysis and genomic characterization of human melanoma differentiation associated gene-5, mda-5: a novel type I interferon-responsive apoptosis-inducing gene. Oncogene 23, 1789–1800.

Kel, A.E., Gössling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V., Wingender, E., 2003. MATCH: A tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res. 31, 3576–3579.

Kurt-Jones, E.A., Sandor, F., Ortiz, Y., Bowen, G.N., Counter, S.L., Wang, T.C., Finberg, R.W., 2004. Use of murine embryonic fibroblasts to define Toll-like receptor activation and specificity. J. Endotoxin. Res. 10, 419–424.

Müller, F., Demény, M.A., 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.

Ralser, M., Querfurth, R., Warnatz, H.J., Lehrach, H., Yaspo, M.L., Krobitsch, S., 2006. An efficient and economic enhancer mix for PCR. Biochem. Biophys. Res. Commun. 347, 747–751.

Rosenstiel, P., Schreiber, S., 2008. Expression signatures, barriers and beyond: the role of oxidative stress in murine colitis and human inflammatory bowel disease revisited. Eur. J. Gastroenterol. Hepatol. 20, 496–499.

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., Hume, D.A., 2007. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat. Rev. Genet. 8, 424–436.

Sandor, F., Latz, E., Mandell, L., Repik, G., Golenbock, D.T., Espevik, T., Kurt-Jones, E.A., Finberg, R.W., 2003. Importance of extra- and intracellular domains of TLR1 and TLR2 in NFkappa B signalling. J. Cell Biol. 162, 1099–1110.

Schug, J., Schuller, W.P., Kappen, C., Salbaum, J.M., Bucan, M., Stoeckert Jr, C.J., 2005. Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biol 6, R33.

Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., Schmidt, D., O'Keeffe, S., Haas, S., Vingron, M., Lehrach, H., Yaspo, M.L., 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321, 956–960.

Takai, D., Jones, P.A., 2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc. Natl. Acad. Sci. U S A. 99, 3740–3745.

Takeuchi, O., Sato, S., Horiuchi, T., Hoshino, K., Takeda, K., Dong, Z., Modlin, R.L., Akira, S., 2002. Cutting edge: role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins. J. Immunol. 169, 10–14.

Vanhecke, D., Janitz, M., 2004. High-throughput gene silencing using cell arrays. Oncogene 23, 8353–8358.

Vanhecke, D., Janitz, M., 2005. Functional genomics using high-throughput RNA interference. Drug Discov. Today. 10, 205–212.

Wang, Y., Leung, F.C., 2004. An evaluation of new criteria for CpG islands in the human genome as gene markers. Bioinformatics 20, 1170–1177.

Zhang, P., Dixon, M., Zucchelli, M., Hambiliki, F., Levkov, L., Hovatta, O., Kere, J., 2008. Expression analysis of the NLRP gene family suggests a role in human preimplantation development. PLoS ONE 3, e2755.

Ziauddin, J., Sabatini, D.M., 2001. Microarrays of cells expressing defined cDNAs. Nature 411, 107–110.