MAX-PLANCK-INSTITUT FÜR MOLEKULARE GENETIK

HIGH-THROUGHPUT FUNCTIONAL ANALYSIS OF HUMAN GENE PROMOTERS USING TRANSFECTED CELL ARRAYS

Dissertation
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften
(Dr. rer. nat.)

eingereicht am Fachbereich Biologie-Chemie-Pharmazie
der Freien Universität Berlin

vorgelegt von
Xi Cheng
aus Shaanxi, V. R. China

Berlin
April 2010

1. Gutachter: Prof. Dr. Hans Lehrach

2. Gutachter: Prof. Dr. Gerd Multhaup

Tag der Disputation: September 7, 2010

TABLE OF CONTENTS

1 SUMMARY

Promoters comprise one of most the important classes of eukaryotic transcriptional regulatory elements. Identification and characterization of these elements is vital to the understanding of complex gene regulation networks. Intensive efforts have been undertaken in developing approaches to identify the promoters on a genome-wide scale. However, functional studies, which are mostly performed by using reporter gene assay on a gene-by-gene basis, have become a limitation for genome-wide promoter characterizations. In this study, transfected-cell array (TCA) technology was applied as a high-throughput method for functional evaluation of 255 human promoters for genes from human chromosome 21 (Chr21) and genes involved in inflammatory bowel disease.

The TCA technique offers a robust platform for high-throughput functional studies of genes and proteins in the context of living cells. The TCA technique is a combination of microarray technology with cellular biology methods such as transfection and cell culture. The transfection method used in TCA is named reverse transfection because the order of addition of transfection probes and cells in TCA experiment has been reversed compared with conventional transfection. On cell arrays, large sets of nucleic acid samples are spotted and temporarily immobilized on glass slides. After spotting, adherent cells are added to arrays and thus a monolayer is created on the surface of the slides. Only cells growing on top of the nucleic acid spots became transfected, resulting in localized transfected cell clusters within a lawn of non-transfected cells. Combined with appropriate detection assays, the phenotypic effects of thousands of cell clusters can be monitored in parallel.

In the present study, two hundred and fifty-five promoters were examined by reverse transfection on cell arrays. Promoter fragments with length of 2500 bp were amplified and cloned into a promoter-less vector upstream of a fluorescence reporter gene. For high-throughput functional analysis, the promoter-reporter constructs were introduced into a human embryonic kidney cell line (HEK293T) on transfected cell array. As a result, seventy-one (27.8%) promoters were found to be active in HEK293T cells; in contrast, 184 promoters did not showed transcription activity. High correlations were observed between promoter activities and endogenous transcript levels as well as ubiquitous tissue expression. Over half of the active promoter fragments were found to associate with RNA polymerase IIa (Pol IIa) binding site. Enrichment of CpG islands and downstream promoter elements (DPE) were

observed for active promoters.

Among 184 promoters which showed no activity in TCA assay, eighty-five promoters could be activated by treatment of cells with addition of external stimuli such as PMA, Trichostatin A and arsenite or depletion of serum from the standard growth medium. Therefore, in total 61.2% (156/255) of the cloned promoter fragments could drive reporter gene expression under at least one of the tested circumstances. The corresponding genes of 99 promoters that remained silent in any treatment condition showed less endogenous expression in HEK293T cells and restricted tissue expression.

In addition to the 2500 bp fragments, truncated fragments with length of 500 bp were designed and cloned for a subset of 62 promoters. The promoter activities of shorter fragments were determined and compared with their longer counterparts. The truncation resulted in not only the change of promoter activity, but also different responses to external stimuli, supporting the presence of regulatory elements within distal promoter regions. An enrichment of binding sites for transcription factors that can integrate signals from the administered stimuli into gene expression were found in these regions. The identified promoter response patterns represent a valuable resource for the elucidation of the complex mechanisms governing transcriptional regulation on the level of the promoters on a whole-chromosome scale.

Taken together, the study presented here provides further insights into the diversity of mammalian promoter functions. The experimental data of promoter activities acquired in this study may be helpful in understanding the regulation of gene expression. Moreover, the successful application of TCA in this study demonstrates that the combination of TCA technique with gene expression profiling and bioinformatics tools represents a robust and cost-effective platform for genome-wide functional studies of regulatory sequences.

2 ZUSAMMENFASSUNG

Promotoren stellen eine der wichtigsten Klassen der eukaryotischen, transkriptionellen Regulationselemente dar. Die Identifikation und Charakterisierung dieser Elemente ist von besonderer Bedeutung für das Verständnis komplexer Genregulationsnetzwerke. Bisher gab es intensive Bemühungen, die Transkritionsstartstelle (TSS) und Promotoren genomweit zu erfassen. Jedoch gab es bei funktionalen Studien Grenzen in Bezug auf eine genomweite Charakterisierung der Promotoren, da sie meist durchgeführt werden, indem Reportergenkonstrukte auf einer Gen-für-Gen-Basis benutzt wurden. In dieser Arbeit wurden transfizierte Zellarrays (Transfected-Cell-Array, TCA) als Hochdurchsatzverfahren für die Evaluation der 255 Promotoren der Gene des menschlichen Chromosoms 21 (Chr21) und von Genen, die mit chronisch-entzündlichen Darmerkrankungen (Inflammatory Bowel Desease, IBD) zusammenhängen, verwendet.

TCA bietet eine solide Grundlage für Hochdurchsatz-Funktionsstudien von Genen und Proteinen bei lebenden Zellen. TCA ist eine Technik, die die Microarray-Techniken mit Methoden der Zellbiologie wie Transfektion und Zellkulturen verbindet. Die Transfektionsmethode, die bei TCA zur Anwendung kommt, wird auch „Reverse Transfektion“ genannt, da Transfektionsproben und Zellen im TCA Experiment in umgekehrter Reihenfolge zusammengefügt werden, im Vergleich zu traditionellen Transfektionen. Auf dem Zellarray werden große Mengen von Nukleinsäureproben auf einem Objektträger aufgebracht, danach werden die adhärent wachsenden Zellen dem Array hinzugefügt. Dadurch wird eine Monozellulärschicht auf der Oberfläche des Objektträgers erzeugt. Nur Zellen, die oben auf den Nukleinsäurespots wachsen, werden transfiziert. Dadurch entstehen lokal transfizierte Zellcluster in einem Zellrasen von nicht-transfizierten Zellen. Verbunden mit angemessenen Detektionsproben, können so die phänotypischen Effekte bei Tausenden Zellcluster parallel beobachtet werden.

In der vorliegenden Untersuchung wurden auf dem Zellarray 255 Promotoren mittels reverser Transfektion untersucht. Fragmente von Promotoren in der Länge von 2500 bp wurden vervielfältigt und strangaufwärts eines fluoreszierenden Reportergens in einen promotorenlosen Vektor kloniert. Für eine Hochdurchsatz-Funktionsanalyse wurden die Promotorreporter-Konstrukte in eine menschliche, embryonale Nierenzellenlinie (HEK 293T) auf einem transfizierten Zellenarray untersucht. Als Ergebnis erschienen 71 (27,8%) der

Promotoren in HEK293T aktiv. Im Gegensatz hierzu zeigten 184 Promotoren keine Transkriptionsaktivität. Hier wurden hohe Korrelationen zwischen Promotorenaktivität und endogenen Transkriptionslevels sowie ubiquitäre Expression im Gewebe beobachtet. Über die Hälfte der aktiven Promotorenfragmente wurden dabei beobachtet wie sie an der RNA Polymerase IIa (Pol IIa) Bindestelle andockten.

Unter den 184 Promotoren, die keine Aktivität im TCA-Test zeigten, konnten 85 Promotoren durch die Behandlung der Zellen mit zusätzlichen äußeren Stimuli wie PMA, Trichostatin A und arseniger Säure oder ein abgereichertes Serum aus gewöhnlichem Wachstumsmedium aktiviert werden. Aus diesem Grund konnten unter wenigstens einem der getesteten Umstände bei 61,2% (156/255) der klonierten Promotorenfragmente Reportergen-Expression induziert werden. Die restlichen 99 Promotoren, die unter allen getesteten Umständen inaktiv blieben, zeigten weniger endogene Expression in HEK293T-Zellen und beschränkte Gewebe-Expression.

Zusätzlich zu den Fragmenten mit 2500 bp Länge wurden für ein Subset von 62 Promotoren kürzere Fragmente der Länge von 500 bp hergestellt und kloniert. Die Promotorenaktivität der kürzeren Fragmente wurde bestimmt und mit der der längeren Fragmente verglichen. Das Ergebnis war nicht nur ein Wechsel in der Promotorenaktivität, sondern auch veränderte Reaktionen auf die externen Stimuli. Beides legt die Präsenz von Steuerungselementen in den distalen Promoterabschnitten nahe. In diesen Abschnitten fand sich eine Anreicherung von Bindestellen für Transkriptionsfaktoren, die Signale von den verabreichten Stimuli in die Genexpression integrieren konnten. Die identifizierten Promotorenreaktionen bieten eine wertvolle Quelle für die Aufklärung der komplexen Mechanismen, die die Transkriptionssteuerung auf der Ebene der Promotoren bei ganzen Chromosomen bestimmen.

Zusammenfassend kann festgehalten werden, dass diese Untersuchungen weitere Einblicke in die Komplexität der Promotorenfunktionen von Säugetieren gegeben hat. Die hier erhobenen Daten der Experimente können dazu beitragen, die Regulation der Genexpression besser zu verstehen. Darüber hinaus kann die erfolgreiche Anwendung von TCA in diesen Untersuchungen zeigen, dass die Kombination von TCA-Techniken mit Profilierung von Genexpression und Bioinformatik-Werkzeugen eine solide Basis für eine kostengünstige, genomweite Funktionsanalyse von Regulationssequenzen bietet.

3 INTRODUCTION

The availability of the genome sequences of many eukaryotic organisms has made the global characterization of gene regulation one of the major challenges and fundamental goals for biomedical research. In eukaryotes, gene expression is known to be an extraordinarily complex process involving a variety of steps including transcription, RNA processing, mRNA stabilization, translation and protein modification. Of these steps, the regulation of transcriptional initiation plays a crucial role. Several DNA sequence elements are known to contribute to transcriptional initiation. Promoters, located immediately upstream of transcriptional start sites (TSSs), likely harbor the most important elements in this regulation. Intensive efforts have been recently undertaken in developing approaches, either hybridization-based or sequencing-based, to identify the TSSs and promoters on a genome-wide scale. Thereby a large number of novel transcripts and promoters have been revealed. However, the functionality of the promoters remains unknown without reference further experimental data. The conventional functional studies are mostly performed using reporter gene assay in the microwell plate format, which requires large amount of transfection experiments and has became a limitation for genome-wide promoter characterizations. In this project, we described the application of the transfected-cell array technology as a high-throughput method in functional studies of human promoters.

3.1 The eukaryotic promoter

3.1.1 A promoter is necessary for proper control of transcription initiation

Transcription initiation in eukaryotes includes integration of multiple regulatory steps and requires close collaboration of the enzyme RNA polymerase and numerous of interacting proteins. Three classes of eukaryotic RNA polymerases exist for transcribing different types of RNAs: RNA polymerase I (Pol I) transcribes ribosomal RNA (rRNA), RNA polymerase II (Pol II) transcribes message RNA (mRNA) and RNA polymerase III (Pol III) transcribes transfer RNA (tRNA) and other small RNAs. RNA polymerase II, which catalyzes the transcription of all protein-coding genes and the synthesis of precursor mRNAs (Kornberg, 1999), is the best characterized RNA polymerase.

The first step in transcription initiation is the determination of the transcription start site (TSS) and the recruitment of RNA polymerase to the position. In eukaryotes, the RNA polymerase

itself does not directly recognize the TSS; instead, a battery of proteins called transcription factors (TFs) are principally responsible for recognizing regulatory DNA sequences and positioning the RNA polymerase at the correct DNA region to initiate the transcription. Transcription factors refer to any protein other than RNA polymerase that is required for transcription, and they usually bind to the *cis*-acting DNA sequences adjacent to the genes that they regulate.

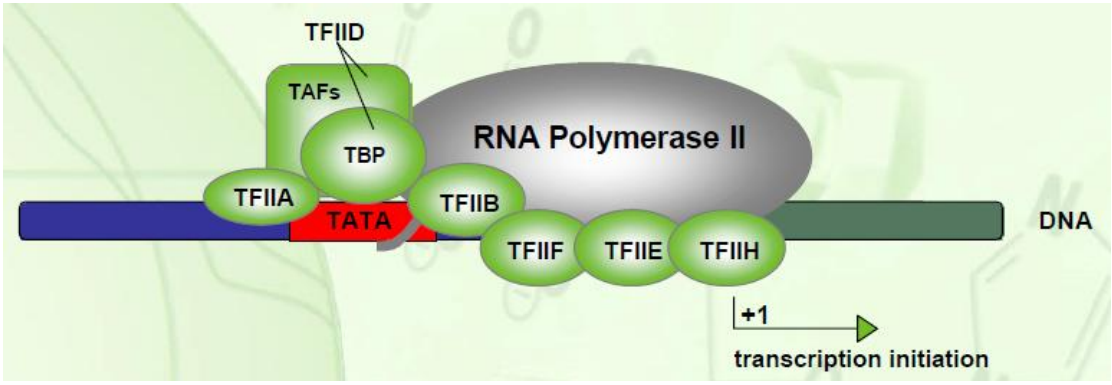


Figure 1. Schematic diagram of the Pol II basal transcription apparatus in eukaryotes. The initiation of transcription by RNA polymerase II requires general transcription factors including TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH to recognize regulatory DNA sequences and position the RNA polymerase at the correct DNA region to initiate the transcription (Jena Bioscience).

An important group of transcription factors for Pol II is the general transcription factors (GTFs), which are ubiquitous in all protein coding genes (Orphanides et al., 1996). The most common GFTs include TFIIA (transcription factors for Pol II, A), TFIIB, TFIID, TFIIE, TFIIF and TFIIH (Thomas and Chiang, 2006). Many of the GTFs don't interact directly with DNA; instead, they together with RNA polymerase II form a transcription pre-initiation complex (PIC) and binds to DNA (Figure 1). The binding of PIC to DNA creates a structure to provide the target that is recognized by Pol II, which helps recruit RNA polymerase over the proper transcription start sites and thus allows the initiation of gene transcription (Lee and Young, 2000). RNA polymerase II and the group of GTFs that directly interact with it are called the basal transcription apparatus.

Besides the general transcription factors, other factors that interact with the proteins of the basal transcription apparatus directly or through a coactivator are also important for transcription. These factors are collectively called upstream transcription factors for that they usually bind to regulatory DNA elements (REs) locating somewhere upstream of the TSS

(Figure 2). Upstream factors vary considerably according to distinct recognition sequences that are present in the proximity of the gene. Depending on the factor, the transcription of the adjacent gene can be either up- or down-regulated. For example, the recruitment of an activator in adjacent of the regulated gene may assists in assembling of basal transcription apparatus, stabilizing the binding of RNA polymerase to DNA or recruiting coactivator proteins to the transcription factor DNA complex, and thus stimulates the transcription. In contrast, the recruitment of a repressor may decrease the transcription by blocking the formation of transcription apparatus or the binding of RNA polymerase to DNA, or by recruiting corepressor factors to the transcription complex.

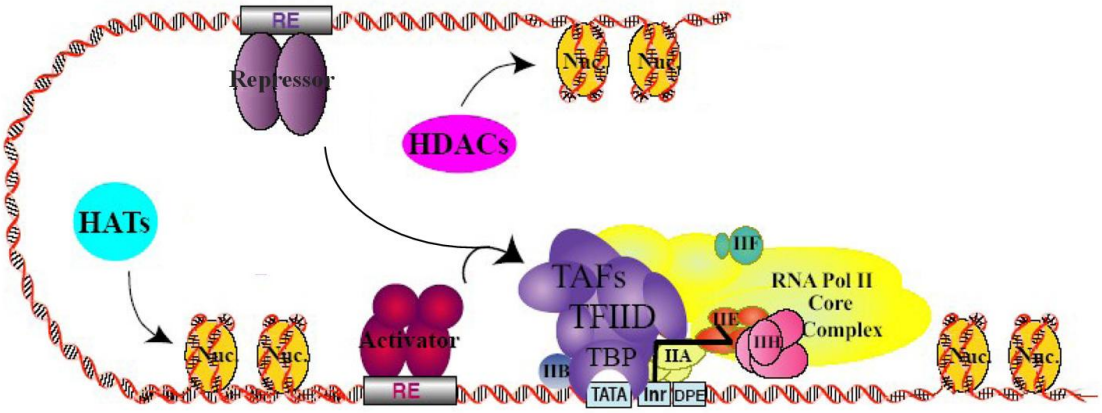


Figure 2. Schematic diagram of complex regulation of transcription by RNA polymerase II in eukaryotes. Transcription by Pol II is dependent on the combinatorial efforts of numerous proteins including RNA polymerase, general and upstream transcription factors and other additional proteins. The recruitment of GTFs and Pol II at the transcription startpoint and the assembly of basal transcription apparatus determine the proper TSS and position the RNA polymerase at the correct DNA region to initiate the transcription. The binding of upstream TFs such as activators or repressors recognize upstream regulatory DNA elements and up- or down-regulated the gene transcription by interacting with the proteins of the basal transcription apparatus directly or through a coactivator. Additional proteins such as HATs and HDACs can also regulate the gene transcription through catalyzing the acetylation or deacetylation of histone (Freiman, 2009).

Moreover, a number of additional proteins such as histone-acetyltransferases (HATs) and histone-deacetylases (HDACs), which catalyze the acetylation or deacetylation of histone proteins, respectively, are also known to play crucial roles in the control of gene expression (Figure 2). Histone acetylation enhances transcription by reducing the affinity between histones and DNA and thereby making it easier for RNA polymerase and TFs to access the promoter region, while deacetylation has effects in an opposite way to repress transcription.

The balance of histone acetylation is maintained by HATs and HDACs (Marks et al., 2003; Roth et al., 2001). Interestingly, some of the HATs and HDACs and the proteins with which they associated, have been found to function as transcription activators/coactivators and repressors/corepressors (Luo and Dean, 1999; Pazin and Kadonaga, 1997; Sterner and Berger, 2000).

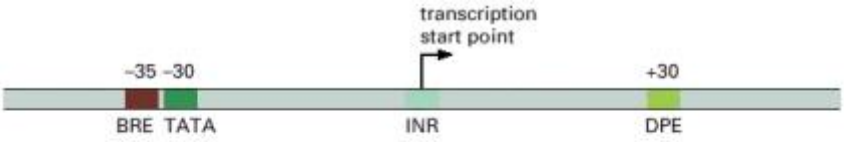
A promoter refers to the region of DNA that is necessary for the transcription of a particular gene and represents the most fundamental *cis*-acting transcriptional regulatory element, typically located in the 5'-flanking region of the gene that it regulates. Concerning the complexity of transcription initiation in eukaryotes that requires the involvement of a large number of proteins, a promoter is defined, in the textbook of Gene VIII (Benjamin Lewin), as *the region containing all these binding sites, that is, which can support transcription at the normal efficiency and with the proper control*. Therefore, a major feature of defining a eukaryotic promoter is the location of binding sites for all transcription factors.

3.1.2 Structure of RNA polymerase II promoters

Based on their functions and distance from the TSS, different regions of the promoter have been termed core promoter, proximal promoter and distal promoter. Sometimes, transcriptional regulatory sequences located upstream of the core promoter are collectively referred to as upstream *cis*-regulatory elements.

3.1.2.1 Core promoter elements

The core promoter refers to the genomic region that surrounds the TSS. It is the minimum stretch of DNA, in many cases approximates 40-50 bp in length, which is required for the recruitment of the transcription apparatus and the initiation of transcription by the RNA polymerase II machinery. Typically, the core promoter encompasses the TSS and extends 35 bp up and downstream of the TSS. Several distinct DNA sequence elements and patterns, such as the TATA box, initiator element (Inr), downstream promoter element (DPE) and TFIIB recognition element (BRE) are associated with core promoters (Figure 3). However, in higher eukaryotic organisms, the structure of core promoters is highly diverse. There are no universal core promoter elements; each element is only found in a subset of genes. Different elements can co-occur in the same promoter, although certain combinations are more likely than others are, and some patterns complement each other (Smale and Kadonaga, 2003).



element	consensus sequence	general transcription factor
BRE	G/C G/C G/A C G C C	TFIIB
TATA	T A T A A/T A A/T	TBP
INR	C/T C/T A N T/A C/T C/T	TFIID
DPE	A/G G A/T C G T G	TFIID

Figure 3. Consensus sequences in the RNA polymerase II core promoter. The core promoter is associated with several distinct DNA sequence motifs such as the TATA box (TATA), Initiator sequences (INR), downstream promoter elements (DPE) and TFIIB recognition element (BRE) (Albert et al., Molecular Biology of the Cell).

TATA box

The TATA box, also called the Goldberg-Hogness box after is discoverers, is the first eukaryotic core promoter element to be identified (Breathnach and Chambon, 1981). It is located 28-34 bp (Sandelin et al., 2007) upstream of the TSS. The consensus sequences for the TATA box, TATAA (Sandelin et al., 2007), is recognized and bound by TATA-box binding protein (TBP). Binding of TBP to TATA box forces the pre-initiation complex (Dennis et al.) to select a TSS in a limited region. The prevalence of TATA boxes is lower than initially thought. Only 10-20% mammalian promoters are found to contain a functional TATA box (Cooper et al., 2006; Gershenzon and Ioshikhes, 2005). In mammals, TATA-containing promoters are commonly associated with tissue-specific genes (Schug et al., 2005).

Inr

The Inr element encompassed the TSS and is identified in a variety of eukaryotes (Breathnach and Chambon, 1981; Corden et al., 1980; Hultmark et al., 1986; Struhl, 1987). Based on functional assays, the consensus for the mammalian Inr was determined as YYANWYY, where the A nucleotide is often at the position +1. The Inr is probably the most common element occurring in focused-core promoters (FitzGerald et al., 2006; Gershenzon et al., 2006; Ohler et al., 2002). A number of proteins have been found to bind to Inr, and the binding of TFIID appears to be particularly important because the sequence specificity of TFIID binding to the Inr region of the core promoter is identical to the Inr consensus sequence (Purnell et al.,

1994). The Inr is functionally independent of the TATA box and has been found in both TATA-containing and TATA-less core promoters. When TATA and Inr exist in one promoter, the spacing between these two elements is critical for their function (O'Shea-Greenfield and Smale, 1992).

DPE

The DPE was identified as a downstream core promoter binding site for purified *Drosophila* TFIID (Burke and Kadonaga, 1996). The DPE is conserved from *Drosophila* to humans and is located precisely at +28 to +32 relative to the A+1 position of the Inr element. The DPE consensus sequence is RGWYVT in *Drosophila* (Kutach and Kadonaga, 2000), while its consensus in human has yet to be determined. However, mammalian core promoters containing sequences that conform to the *Drosophila* consensus have been found to possess DPE activity. Generally, DPE occurs together and functions cooperatively with the Inr element. The spacing between the Inr and DPE is critical for optimal transcription.

BRE

The BRE was originally identified as a TFIIB binding site that is located immediately upstream of some TATA boxes (Lagrange et al., 1998). The BRE consensus is SSRCGCC (Lagrange et al., 1998). Depending on the promoter context, the BRE can either increase or decrease transcription in eukaryotes (Deng and Roberts, 2005, 2007; Lagrange et al., 1998) by an unknown mechanism.

3.1.2.2 CpG islands

CpG islands are genomic regions that contain a high GC content and a high frequency of CpG dinucleotides relative to the bulk of genomic DNA (Gardiner-Garden and Frommer, 1987). The size of CpG islands ranges from 0.5 to 2 kb. It has been found that approximately 40% of mammalian promoters (Fatemi et al., 2005) and 70% of human promoters (Saxonov et al., 2006) contain CpG islands. CpG islands are most often associated with housekeeping or ubiquitous genes (Schug et al., 2005), although there are many exceptions, such as brain-specific genes (Gustincich et al., 2006). Promoters containing CpG islands typically lack TATA or DPE core promoter elements. Only a fraction of CpG-associated promoters have TATA-like elements.

3.1.2.3 Upstream promoter elements

In addition to the core promoter, extended promoter regions, including other *cis*-acting DNA elements such as, gene specific transcription factor binding sites, enhancer, silencer and insulator (Blackwood and Kadonaga, 1998; Bulger and Groudine, 1999; West et al., 2002), can be located from hundreds of base pairs to many kilobases from the TSS. These elements contain recognition sites for numerous sequence-specific DNA-binding proteins that are involved in the regulation of RNA polymerase II transcription. The proximal promoter is the region spreading roughly to 250 bp upstream of the TSS, which contains the binding sites for some specific transcription factors. Enhancers and silencers can be located many kilobases away from the TSS or even on a different chromosome and act either to activate or to repress transcription (Blackwood and Kadonaga, 1998; Bulger and Groudine, 1999). Insulators appear to prevent inappropriate interactions between adjacent chromatin domains; they can either establish domains that separate enhancers and promoters to block their interaction or create a barrier against the spread of heterochromatin (West et al., 2002).

3.1.3 Sharp and broad promoters

In early studies, the core promoter used to be considered a single motif that serves as a universal TSS; however, emerging data indicates the presence of considerable diversity in the structure and function of the core promoter. It was assumed that each promoter has a TSS located at one specific genomic position. The transcription start site (TSS) of a gene is the nucleotide that is the first to be transcribed into the corresponding mRNA. However, a cap analysis of gene expression (CAGE)-based study indicated that most human and mouse promoters lack the distinct TSS; instead, the typical promoter contains a number of closely located TSSs that spread over a broad region of around 50-100 bp (Carninci et al., 2006). With respect to the distribution of the TSSs, mammalian promoters can be separated into two classes of sharp and broad promoters (Sandelin et al., 2007). The sharp promoters have only a single dominant TSS, and the transcription always initiates at a specific nucleotide (Figure 4). In contrast, the broad promoters have multiple, weakly defined TSSs (Figure 4). The transcription may initiate at either different TSSs or one specific nucleotide position (Carninci et al., 2006).

The sharp promoters are mostly associated with the presence of a TATA box. The prevalence of distinct TSS in TATA-containing promoters has been confirmed by the analysis of smaller

promoter sets (Suzuki et al., 2001). However, not all sharp promoters have a TATA box. Sharp promoters are primarily used for tissue-specific expression, which correlates with the feature of TATA-containing promoters. In contrast, CpG islands are found to be overrepresented in broad promoters. The broad promoters are generally associated with ubiquitously expressed genes, and this is also true for promoters with CpG islands.

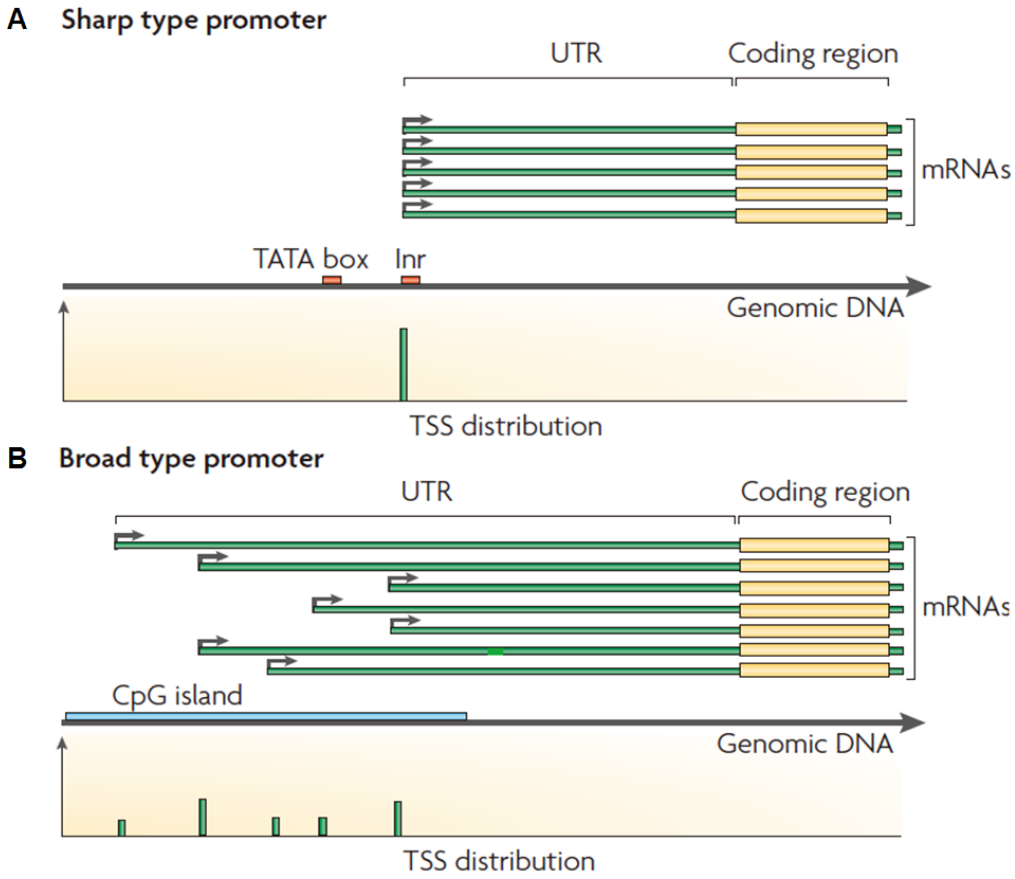


Figure 4. Classification of mammalian promoters by TSS distribution. Mammalian promoters can be classified with respect to the distribution of the TSSs. (A) Promoters termed as the sharp type initiate transcription at one or a few consecutive nucleotides, resulting in a single-peak TSS distribution and the synthesis of same mRNAs. The sharp type promoters are often associated with TATA box and INR elements. (B) Promoters termed as the broad type can initiate transcription over a 50-100 bp region, resulting in a TSSs distribution and a number of mRNAs with different lengths but usually the same protein coding sequences. The broad-type promoters are often TATA-less and CpG-enriched (Sandelin et al., 2007).

3.1.4 Alternative promoters

The completion of the Human Genome Project has revealed that the human genome contains

only 20,000-25,000 protein-coding genes (Consortium., 2004; Lander et al., 2001; Waterston et al., 2002). This number is unexpectedly small compared to the total gene numbers in the genome of fruit fly and nematode worm, which are estimated to be 14,000 and 19,000, respectively (Adams et al., 2000; Consortium., 1998). The multifaceted use of the genes, in addition to the total gene numbers, are generally believed to play a crucial role in generating the functional diversity and phenotypic complexity of human genes. The use of alternative promoters (APs) is one of the key mechanisms for the production of multiple mRNA transcripts and subsequently different proteins or protein isoforms from a single gene (Landry et al., 2003). Previous studies suggested that 14-58% of human genes may have alternative promoters APs (Baek et al., 2007; Cooper et al., 2006).

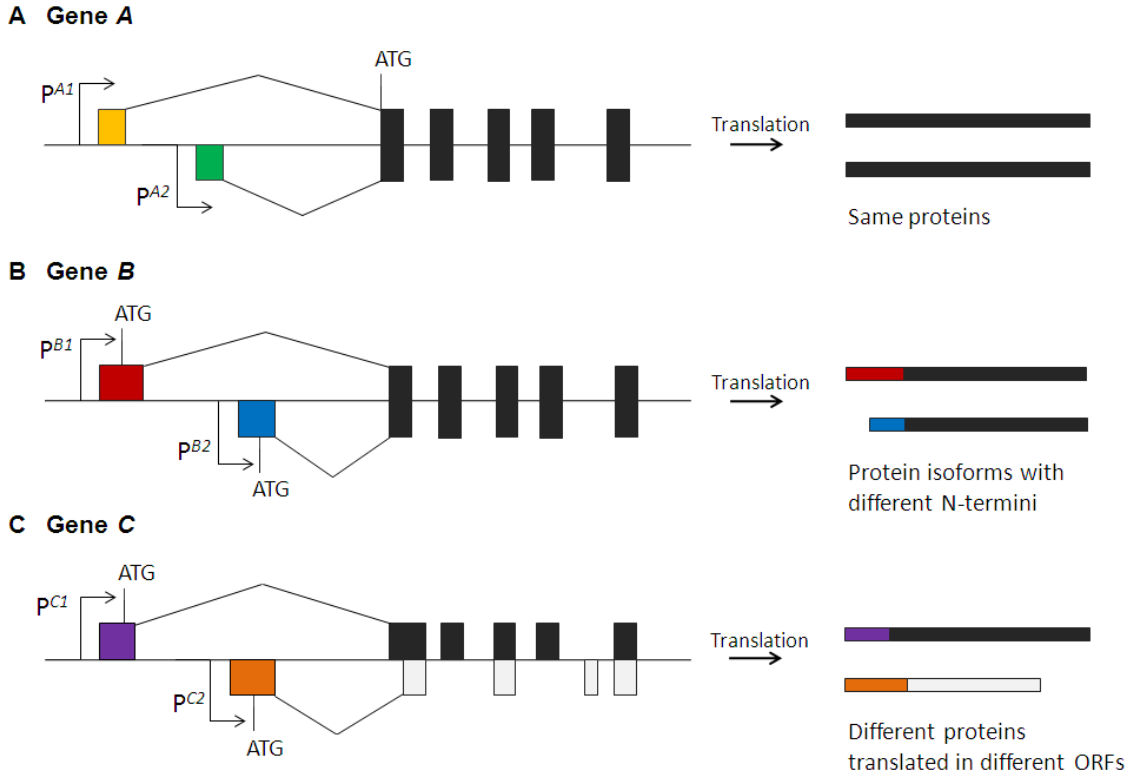


Figure 5. The types and consequences of alternative promoters. (A) The use of alternative promoters (p^{A1} and p^{A2}, represented by arrows) produces the same protein. The translation initiation site (ATG) is included in the common coding exons (black boxes), and therefore the variant 5' initial exons (colored boxes) do not affect the mRNA translation. (B) The use of alternative promoters (p^{B1} and p^{B2}) results in protein isoforms differing in their N-termini. The variant 5' initial exon contains the translation initiation site and contributes to the protein production. (C) The use of alternative promoters (p^{C1} and p^{C2}) creates transcripts with different open reading frames (ORFs, represented by black and white boxes) and produces different proteins (figure modified from (Landry et al., 2003).

Landry and coauthors summarized the genes with APs in recently published studies and classified the APs in to three types in terms of the consequent proteins produced by the promoter (Landry et al., 2003) (Figure 5). First, transcription with alternative promoters produces the same proteins. The mRNAs produced by APs consist of variant first exons and common follow-up exons. The translation initiation site (known as ATG) is included in the common exons. Therefore, the variance of first exon does not affect the mRNA translation (Figure 5A). Second, the use of alternative promoters results in protein isoforms. The first exon of mRNA produced with this type of alternative promoter contains an ATG site and contributes to the mRNA translation, resulting in the production of protein isoforms differing in their N-termini (Figure 5B). Third, the usage of alternative promoters creates transcripts that code for different proteins because they are translated in different open reading frames (ORFs) (Figure 5C).

3.2 Functional studies of promoter

3.2.1 Importance of promoter studies on a genome-wide scale

In order to gain insight into the mechanisms of gene regulation, it is of major importance to identify and characterize regions functionally relevant to transcription initiation. Accurate promoter identification in the human genome has been and remains a challenge. Promoters have defined as an array of *cis*-acting DNA sequence elements required for efficient initiation and control of transcription. Different transcripts of a gene maybe regulated by distinct promoters and result in altered proteins and altered regulatory properties of the respective 5’ UTRs. In addition, differences in cis-regulatory modules of these promoters may also affect the expression patterns of the transcripts.

Promoter regions have been investigated mainly on a gene-by-gene basis. From these studies it became apparent that *cis*-acting regulatory elements of promoters are conserved among orthologous genes (Miller et al., 2004). Mammalian promoters, however, exhibit a high degree of diversity, regarding for example the lengths of core promoter regions and the presence of CpG islands, TATA boxes and other functional elements. Recently, the availability of the genome sequences of many eukaryotes has enabled the computational analyses of *cis*-acting DNA sequence elements on a genome-wide scale. The widely applied approaches are prediction-based and rely on sequence-alignment annotation and genome-wide mapping of binding sites for basic transcription machinery elements towards the identification of

associated TSSs (Cooper et al., 2006; Kim et al., 2005a). Different libraries, such as full-length cDNAs and CAGE tags, have been used for the prediction of TSSs and promoters (Carninci et al., 2006; Suzuki et al., 2002). However, maps of promoter elements obtained by computational analysis are not sufficient to understand promoter function; there is growing demand for systematic experimental approaches to promoter activity evaluation on a genome-wide level.

3.2.2 Conventional techniques for single-gene and large-scale studies

The functional studies of promoters on a gene-by-gene basis are performed using different reporter genes, like luciferases and fluorescent proteins, representing two of the most important and commonly used systems. Here, the promoter to be tested is cloned into a promoter-less vector upstream of a reporter gene, so that the expression of the reporter gene will be under control of the target promoter. Then the constructed plasmid is added to cells for transfection. Thereby, the promoter activity is determined by qualitative or quantitative measurements of reporter gene activities.

For large-scale functional studies of promoters these approaches are conventionally transferred to a microwell plate-based format. In this approach, cells are first prepared in each well, and then different promoter-reporters constructs are added into different microwells for transfection. This approach allows investigating hundreds of promoters in parallel. However, only one transfection reaction can be performed in a microwell and only one promoter is analyzed, which means a huge amount of transfection work and high reagent consumption are required in order to analyze the promoter activity on a large scale.

3.3 Transfected cell array (TCA) technique

3.3.1 Principle of TCA

The success of DNA and protein microarrays motivated researchers to miniaturize the cell-based functional study in a high-density arraying format. Developed by Ziauddin and Sabatini, the Transfected Cell Array (TCA) represents a robust alternative for high-throughput analysis of gene function in mammalian cells (Ziauddin and Sabatini, 2001). The principle of the TCA technique is to combine the microarray technology with cellular biology methods such as transfection and cell culture. The basic idea is to transfect nucleic acid molecules that are

immobilized on a solid surface into mammalian cells and then detect the physiological effects caused by the introduction of the foreign nucleic acid on these cells. The nucleic acid to be transfected, along with a lipid transfection reagent, is printed at a high density on a glass slide. The nucleic acid molecules can be either DNA or RNA. Recently, DNA plasmids containing the sequence coding for short hairpin RNAs (shRNAs) or chemically synthesized small interfering RNAs (siRNAs) have also been successfully arrayed for high-throughput transfection and gene knock-down in mammalian cells (Kumar et al., 2003; Mousses et al., 2003). When the microarray is covered with a monolayer of adherent cells, only the cells growing on the printed areas take up the nucleotides and become transfected, resulting in spots of localized transfection within a lawn of non-transfected cells.

3.3.2 TCA workflow

The process of TCA can be divided briefly into three steps: microarray preparation, reverse transfection and detection assay (Vanhecke and Janitz, 2004) (Figure 6).

(1) Microarray preparation. To prepare the microarrays for TCA, the DNA/RNA molecules have to be printed on glass slides. In contrast to the classical microarrays for gene expression profiling studies in which printed probes are used for hybridization, the nucleic acids in cell microarrays cannot be permanently immobilized on the solid surface because they must enter cells growing on top of the spot in the following transfection step. Therefore, the samples to be arrayed are first dissolved in a gelatin solution, which facilitates the adherence of nucleic acid on the glass slide and the release of the nuclei acid into the cells during transfection. The design and generation of cell arrays does not differ from the classical microarrays. Any robotic arrayer for traditional DNA microarrays can be used in preparation of the cell arrays.

(2) Reverse transfection. Ziauddin and Sabatini named this method “reverse transfection” because the order of addition of nucleic acids and adherent cells in TCA experiment has been reversed compared with conventional transfection. In conventional transfection, DNA/RNA is added to the adherent cells that have been previously cultured on a solid surface, so transfection events occur throughout the entire cell layer (Figure 7A). In contrast, in reverse transfection, DNA/RNA is printed on a glass first before adding cells on top of the spotted array, resulting in localized transfection on spatially separated group of cells (Figure 7B).

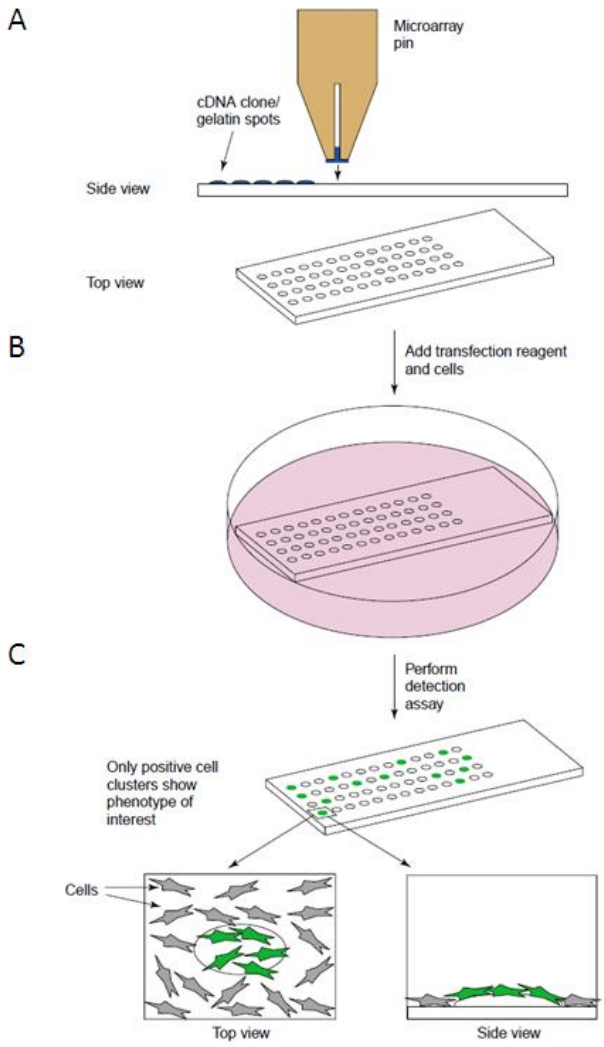


Figure 6. An overview of TCA. The process of TCA consists of three steps: (A) Microarray preparation, (B) reverse transfection and (C) detection assay (Bailey et al., 2002).

The nucleic acids have to be incubated with transfection reagents before covering the array with a monolayer of adherent cells, so that the nucleic acid can be taken up by the overlying cells. Two different methods of adding transfection reagent have been developed for reverse transfection, which are called the gelatin method and lipid method, respectively. In the gelatin method, the mixture of nucleic acids in gelatin solution is printed on glass slides, and the transfection reagents are added by an additional incubation step before transfection. In the lipid method, the transfection reagents are pre-mixed with DNA/RNA before array printing. Similar to normal transfection, the ratio of nucleic acids to transfection reagent and the incubation time are critical for the successful transfection of cell arrays and have to be optimized for each cell line.

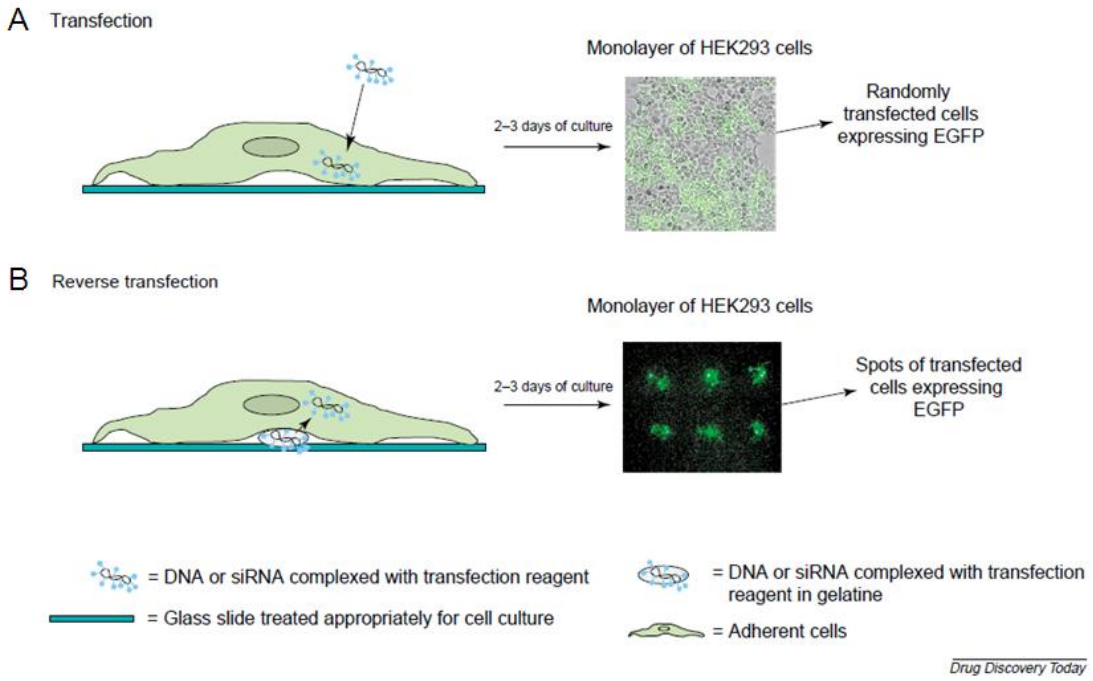


Figure 7. Principle of reverse transfection. (A) In the course of normal transfection, the cells are randomly transfected with a nucleic acid entering the cell from the surrounding solution. (B) During the reverse transfection process, nucleic acid enters the recipient cell from the solid support. Thus, spatial separation of the transfection events can be achieved in the same experiment (Vanhecke and Janitz, 2005).

(3) Functional detection. Depending on the nature of the molecular event to be monitored, signal detection can be performed on either living or fixed cells. The functional assays that are generally carried out on adherent cells can also be applied to TCA. For instance, a common detection method is based on the signal emitted from the fluorescent dyes that are coupled to monoclonal antibodies (e.g. Cy3, Cy5 or Alexa) or biological stains (e.g. DAPI or Hoechst). Alternatively, autofluorescent reporter proteins such as green fluorescent proteins (GFP) and its derivatives allow for direct detection. For signal acquisition, a microarray scanner can be used for the detection of fluorescent signal from the whole slide. Alternatively, a fluorescent microscope allows for the collection of signal intensities from single cells and the application of statistical analyses of signal to background ratios. Standard software for DNA arrays can be used for image analysis of cell arrays, including the identification and quantification of the signal acquired by the scanner or microscope. Recent progress in high-throughput automated fluorescence microscopy and accompanying image process software allow for the large-scale array screening, image processing and data analysis (Baghdoyan et al., 2004; Liebel et al., 2003; Price et al., 2002; Wolf, 2003; Zemanova et al., 2003).

3.3.3 Applications of TCA

The TCA technique has been applied to a variety of research purposes. One of the most common usages is to overexpress genes in order to study protein localization, interaction and function. TCA has been used for the screening of protein intercellular localization and cell phenotype on the chromosome scale (Hu et al., 2006). TCA has also been applied for the identification of protein-protein interaction by the combination with human two-hybrid systems (Fiebitz et al., 2008). The TCA approach can be used as an alternative to protein microarrays for the identification of drug targets and as an expression cloning system for the discovery of gene products that alter cellular physiology. Alternatively, TCA can also be used to investigate the effects of gene silencing. For example, the RNAi technology is combined with the TCA technique to perform large scale gene loss-of-function studies (Erfle et al., 2004; Kumar et al., 2003; Mousses et al., 2003; Silva et al., 2004; Vanhecke and Janitz, 2004; Yoshikawa et al., 2004).

4 OBJECTIVES

Identification and characterization of transcriptional promoters is vital to the understanding of complex networks of eukaryotic gene regulation. With the availability of the genome sequences of human, mouse and other eukaryotic organisms and the development of approaches for the identification of TSSs and promoters on the genome-wide scale, a large number of novel putative promoters have been revealed. However, given the inherent limitations of *in silico* studies relying solely on DNA sequence analysis, the functional characterization of mammalian promoters and associated elements requires experimental confirmation. Conventional functional studies that are mostly performed at the single gene level can hardly meet the needs of such large-scale analysis and have become a limitation for genome-wide promoter characterizations. Therefore, the development of systematic and high-throughput approaches for the investigation of promoter function in mammalian cells has become a growing demand.

Transfected cell array, the combination of microarray technology with transfection in mammalian cells, offers a robust platform for large-scale functional analyses of genes and proteins in the context of living cells. In the study presented here, the transfected-cell array technology was applied as a high-throughput method in the functional evaluation of human promoters. This study was part of a project aimed at large-scale identification and functional characterization of mammalian promoters and associated elements. In particular, the following topics of this study are addressed:

1. The application of TCA technique for high-throughput detection of human promoter activities in reporter gene assay;

2. Computational and functional analyses of promoters including correlation of promoter activity with endogenous transcript levels, RNA Pol IIa binding sites, endogenous gene expression patterns and promoter sequence characteristic.

3. The stimulation of promoter activities with external chemicals and the prediction of upstream regulatory elements by promoter fragment truncation analysis.

5 MATERIALS

5.1 Chemicals

2-Propanol for analysis	Merck, Germany
4',6-diamidino-2'-phenylindole dihydrochloride (DAPI)	Sigma, Germany
Acetone	Merck, Germany
Acryl AquaClean	WAK-Chemie Medical, Germany
Bovine Serum Albumin (BSA)	NEB, Germany
Dimethyl sulfoxide (DMSO)	Sigma, Germany
Dulbecco's modified Eagle's medium (DMEM)	Gibco, Invitrogen, Germany
Ethanol (p.a.)	Merck, Germany
Ethanol absolut zur Analyse	Merck, Germany
Ethidium bromide (1%)	Merck, Germany
Ethylendiamine tetraacetic acid (EDTA)	Merck, Germany
Fetal calf serum (FCS)	Biochrom, Germany
Fluoromount-G	Southern Biotech
Formaldehyde pro analysi min. 37%	Merck, Germany
Gelatin from bovine skin	Sigma, Germany
L-glutamine	Invitrogen, Germany
$MgCl_2$ 50 mM	Invitrogen, Germany
Penicillin/streptomycin	Invitrogen, Germany
Phorbol 12-myristate 13-acetate (PMA)	Sigma, Germany
Poly-L-lysine solution, 0.1 % (w/v) in H_2O (preservative added)	Sigma, Germany
Quadriperm chamber	Greiner, Germany
Rediload	Invitrogen, Germany
Silico Gel Orange	Roth, Germany
Sodium arsenite ($NaAsO_2$)	Merck, Germany
Sodium chloride (NaCl)	Merck, Germany
Sodium hydroxide (NaOH)	Merck, Germany
Trichostatin A	Sigma, Germany

Trypan Blue 0.4%	Sigma, Germany
UltraPure™ Agarose	Invitrogen, Germany
UltraPure™ Sucrose	Invitrogen, Germany
VECTABOND Reagent for Tissue Section Adhesion	Vector Laboratories, USA

5.2 Kits

Effectene Transfection Reagent	Qiagen, Germany
EndoFree Plasmid Maxi Kit	Qiagen, Germany
Gateway BP Clonase II Enzyme Mix	Invitrogen, Germany
MinElute Gel Extraction Kit	Qiagen, Germany
One Shot TOP10 competent cells	Invitrogen, Germany
QIAprep Spin Miniprep Kit	Qiagen, Germany
QIAquick PCR purification Kit	Qiagen, Germany
RNeasy mini Kit	Qiagen, Germany

5.3 Buffers and solutions

Dulbecco's Phosphate Buffered Saline (D-PBS) (+$CaCl_2$/+$MgCl_2$)	Invitrogen, Germany
Dulbecco's Phosphate Buffered Saline (D-PBS) (-$CaCl_2$/-$MgCl_2$)	Invitrogen, Germany
Gel loading buffer (6×)	Bromophenol blue 0.25% (w/v) Xylenecyanol FF 0.25% (w/v) Glycerol 30% (v/v)
NEB Buffer	NEB, Germany
PCR Rxn buffer (10×) (-$MgCl_2$)	Invitrogen, Germany
Phosphate Buffered Saline (PBS) (10×) (pH7.4)	NaCl 137 mM KCl 2.7 mM Na_2HPO_4 100 mM KH_2PO_4 2 mM
Phosphate Buffered Saline (PBS) (1×)	NaCl 13.7 mM KCl 0.27 mM

	Na_2HPO_4 10 mM
	KH_2PO_4 0.2 mM
TAE (1×) (pH 8.0)	Tris-Acetate 40 mM
	EDTA 1 mM
Taq DNA polymerase Reaction buffer	Invitrogen, USA

5.4 Enzymes

Accutase	PAA, Germany
BSA	NEB, Germany
Restriction endonuclease	NEB, Germany
T4 DNA ligase	NEB, Germany
Taq DNA polymerase	Roche, Germany

5.5 Primers

5.5.1 Primer pair for amplification of promoter fragments

*att*B1 forward: 5'-AA AAA GCA GGC TNN - promoter-specific sequences*-3'

*att*B2 reverse: 5'-A GAA AGC TGG GTN - promoter-specific sequences*-3'

(*Primer sequences for specific promoter fragments, see Table 8 and Table 9 in Appendix section 9.2)

5.5.2 Primer pair for installation of the adapter sequences

*att*B1 adapter: 5'-G GGG ACA AGT TTG TAC AAA AAA GCA GGC T-3‘

*att*B2 adapter: 5'-GGG GAC CAC TTT GTA CAA GAA AGC TGG GT-3'

5.5.3 Primer pair for sequencing

forward (pZseq): 5'-CTC AGA TCT CGA GCT CAA GC-3'

reverse (BHseq): 5'-CTG GCC ATG GTG GCG ACC-3'

5.6 Clones and vectors

pDONR221	Invitrogen, Germany
pHcRed1-N1	Clontech, USA

pZsGreen1-1	Clontech, USA

5.7 Molecular weight markers

λ DNA-BstEII Digest	NEB, Germany

5.8 Cell line

HEK 293T/17	ATCC, USA

5.9 Laboratory equipments and disposables

Cover glasses (22 ×64 mm)	BDH, Gremany
1G Illumina Genome Analyzer	Solexa,
BioPhotometer	Eppendorf, Germany
Cellulose acetate membrane (0.45µm)	Falcon® Becton Dickinson
Centrifuge 5810R	Eppendorf, Germany
Corning Centrifuge tube (15 ml, 50 ml)	Corning, USA
Costar Stripettes (5 ml, 10 ml, 25 ml)	Corning, USA
Cryovials	GREINERbio-one, Germany
Diamond Precision Tips (D10, D200)	Gilson, France
Disposable cuvettes (UVette®)	Eppendorf, Germany
DPU-414 Thermal Printer	SII, Germany
Fluorescence microscope IX81	Olympus, Europe
GeneGenius2 Image Analyser	Syngene, UK
Hemocytometer Cell Count Calculator	Neubauer counting slide
HERAcell 240 CO2 Incubator	Kendro, Germany
HERAsafe	Kendro, Germany
IEEE 1394 Digital CCD camera	Hamamatsu, Japan
InstrumentONE	http://www.m2-automation.de M2 Automation, Germany
Lauda E100 Ecoline Immersion Circulators	Lauda, Germany
Leica DM IL Inverted Microscope	Leica, Germany
Liquid Nitrogen Container	CHART, USA
Mastercycler Personal	Eppendorf, Germany

MC1 LC 2200S Balance Calibration	Sartorius, Germany
MicroAmp Tubes and Caps	Applied Biosystems, USA
Microscope slide (76 × 26 mm)	Menzel, Germany
Microscope slides (75 × 25 × 1.0 mm)	Menzel, Germany
Microwave	AEG, Germany
MiniCentrifuge Model GMC-060	LMS, Japan
MiniSpin plus	Eppendorf, Germany
MR 3001	Heidolph, Germany
Parafilm M	Pechiney Plastic Packaging, USA
PCR thermocycler (PTC-100)	MJ Research, USA
PIPETBOY acu	IBS, Switzerland
Pipettes (Pipetman™)	Gilson, France
Precision balance	Mettler, Germany
Reaction tubes and caps (MicroAmp™)	Applied Biosystems, USA
Rotilabo® Drying cabinets	Roth, Germany
SafeSeal Tips PreCision (10 µl, 20 µl, 200 µl, 1 ml)	Biozmy, Germany
SafeSeal tips Premium XL (10 µl)	Biozmy, Germany
Safety comfort gloves	Latech, Germany
Schleicher & Schuell FP 030/2 Disposable Filter Holder (0.45µm)	Whatman, Germany
Table centrifuge (5424)	Eppendorf, Germany
Thermal cycler	Eppendorf, Germany
Thermomixer 5436	Eppendorf, Germany
Thermomixer comfort (1.5 ml)	Eppendorf, Germany
Tissue Culture Dishes (100 × 20 mm, 150 × 20 mm)	TPP, Switzerland
Tissue Culture Flasks (25 cm^2, 75 cm^2, 150 cm^2)	TPP, Switzerland
Biochem-VacuuCenter BVC 21	VACUUBRAND, Germany
Vortex Genie 2	Scientific Industries, USA
Water purification system (Milli-Q™)	Millipore, USA

5.10 Software, databases, and bioinformatic tools

Pride	http://pride.molgen.mpg.de/pride.html
Primer3	http://primer3.sourceforge.net
Scan^R	Olympus, Europe
Scan^R Analysis	Olympus, Europe
TransFac MATCH	BIOBASE, Germany

6 METHODS

6.1 Promoter cloning and TCA analysis

6.1.1 Promoter prediction and amplification

The human gene annotations of Ensembl database v30 was used as basis for the large-scale analysis of promoter sequences. Putative promoter regions were defined relative to the most 5' TSS of all annotated transcripts. The promoter fragments were prepared after two-step PCR using two sets of primers including promoter-specific primers and Gateway adapter primers.

In the first PCR, promoter-specific primers were designed using the software PRIDE (Haas et al., 2003) for the amplification of the region 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 region. To the 5' end of each primer, 12 bases of *att*B site sequences were added for recombination cloning after two-step PCR amplification of the fragments (Gateway technology, Invitrogen). Promoter fragments were amplified following a touchdown PCR protocol optimized for amplification of GC-rich promoter regions (Table 1) (Ralser et al., 2006), using genomic DNA and available genomic BAC and fosmid clones as templates. Then, a secondary PCR was performed with Gateway adapter primers (Invitrogen) to install the complete *att*B adapter sequences to the ends of each promoter fragment.

Table 1. Touchdown polymerase chain reaction program.

Step	Temperature	Time	Cycle number
Initialdenaturation	94 ℃	2 min	1×
Denaturation	94 ℃	30 sec	
Annealing	65 ℃ >> 55 ℃*	30 sec	20×
Extension	72 ℃	30 sec**	
Denaturation	94 ℃	30 sec	
Annealing	55 ℃	30 sec	30×
Extension	72 ℃	30 sec**	
Final extension	72 ℃	10 min	1×

*Annealing temperature was reduced by 0.5 ℃ after each cycle

For the amplification of truncated promoter fragments, the 3' primers were kept the same as for the larger fragments whereas the 5' primers were re-designed to include 500 bp sequences

upstream of the TSS. The protocols for promoter amplification and adapter sequences installation were the same as used for 2.5 kb fragments.

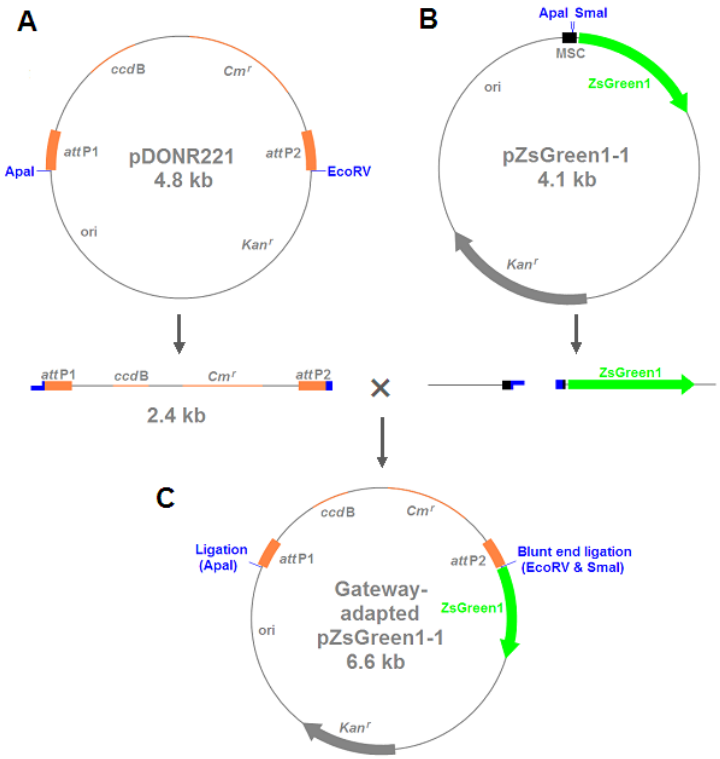


Figure 8. Construction of Gateway-adapted vector. (A) The pDONR221 vector (Invitrogen) was digested with restriction endonuclease ApaI and EcoRV. Subsequently a 2.4 kb *att*P1-and *att*P2-including fragment was obtained and purified. (B) The pZsGreen1-1 vector was digested with restriction endonuclease SmaI and ApaI in turn. (C) The attP sites including fragment and the linearized pZsGreen vector were ligated, resulting in the construction of Gateway-adapted pZsGreen1-1 vector.

6.1.2 Plasmid construction

6.1.2.1 Gateway pZsGreen vector construction

pZsGreen1-1 (Clontech) is a promoterless vector encoding an green fluorescent protein ZsGreen1. ZsGreen is a human codon optimized variant of wild-type *Zoanthus sp.* green fluorescent protein with an excitation maximum at 493 nm and an emission maximum at 505 nm. pZsGreen1-1 can be used to monitor transcription from different promoters or promoter/enhancer combinations inserted into the multiple cloning site (MCS) located upstream of the ZsGreen1 coding sequence.

To enable rapid and highly efficient cloning of the promoter fragment into the vector, pZsGreen1-1 was converted to a Gateway-adapted vector by the insertion of the *att*P1-and *att*P2-flanked fragment from pDONR221 (Invitrogen) into the multiple cloning site of the pZsGreen1-1 vector via restriction cloning (Figure 8).

6.1.2.2 Promoter cloning

To clone the amplified promoter fragments, the PCR products were purified by PEG precipitation. Then the purified attB-PCR product were mixed with the attP- adapted pZsGreen vector and incubated with Gateway BP Clonase II Enzyme Mix (Invitrogen) for over 3 hours at 25℃. Each reaction was 10 µl of the total volume, and the reaction component amounts were calculated following Gateway technology instruction manual (Invitrogen) (Table 2). After incubation with Proteinase K for 10 minutes at 37℃, the mixture was transformed into One Shot TOP10 competent cells (Invitrogen) (see the cloning protocol below). The resulting colonies were screened by PCR, and plasmids were isolated from positive clones using a QIAprep Spin Miniprep Kit (Qiagen). Identity of the inserts was confirmed by 5' and 3' end sequencing.

Table 2. Gateway BP recombination reaction.

Components	Sample	Negative Control	Positive Control
attB-promoter fragments after PCR (40~150 ng/µl)	1.0 - 4.0 µl	-	-
attP-pZsGreen1-1 vector (250 ng/µl)	0.5 µl	0.5 µl	0.5 µl
pEXP7-tet positive control (50 ng/µl)	-	-	0.5 µl
TE Buffer, pH 8.0	to 8.0 µl	7.5 µl	7.0 µl

Cloning Protocol:

- ✧ Step I: PCR purification.
  - ➢ Add 1/3 volume of PEG 30% / 30 mM $MgCl_2$, mix well.
  - ➢ Centrifuge at max speed for 15 min at room temperature.
  - ➢ Discard supernatant carefully, dissolve the pellet in 20 µl TE buffer.

- ✧ Step II: Gateway BP recombination.
  - ➢ Add the components (table 2) to 1.5 ml microcentrifuge tubes at room temperature and mix.

- Remove the BP Clonase enzyme mix from -80 ℃ and thaw on ice (~2 min).
- Vortex the BP Clonase enzyme mix briefly 2 × 2 sec.
- To each sample above, add 2.0 µl of BP Clonase enzyme mix. Mix well by vortexing briefly 2 × 2 sec.
- Incubate reaction at 25 ℃ overnight.
- Add 1 µl Proteinase K solution to each reaction. Vortex briefly. Incubate for 10 min at 37 ℃.

✧ Step III: Transformation.

- Transform 1 µl of the reaction into 20 µl One Shot (Top10) competent cells.
- Incubate on ice for 30 min.
- Heat shock cells on 42 ℃ for 30 sec.
- Add 80 µl of pre-warmed SOC medium and incubate in the shaker at 37 ℃ for 45 min.
- Plate 10-20 µl of the culture onto KAN (100 mg/ml) LB agar plates. Incubate at 37 ℃ overnight.

6.1.2.3 Normalization plasmid HcRed construction

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. HcRed1 was generated by mutagenesis of a non-fluorescent chromoprotein from the reef coral *Heteractis crispa* (Gurskaya et al., 2001) and its coding sequence has been human codon-optimized for higher expression in mammalian cells (Haas et al., 1996). The excitation and emission maximum of HcRed1 occur at 588 nm and 618 nm ±4 nm, respectively. pHcRed1-N1 vector contains a human cytomegalovirus immediate-early promoter (CMV-IE) for strong constitutive expression of the HcRed1 reporter gene. The CMV-IE, with the total length of 589 bp, includes 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 part was removed by restriction digestion, resulting in a 266 bp-long 3’ section of the original CMV promoter upstream of the HcRed reporter gene.

6.1.3 Microarray preparation

6.1.3.1 Sample preparation

Samples for array spotting were prepared using lipid-DNA-method (LD-method) described by Ziauddin and Sabatini (Ziauddin and Sabatini, 2001) with modifications.

Gelatin powder (SIGMA, G-9391) was dissolved in MilliQ water by heating at 60 ℃ for 15 min and sterile-filtered through a 0.45µm cellulose acetate membrane (Falcon®, Becton Dickinson). Cooled solutions can be stored at 4 ℃ for over a year.

The putative promoter plasmids and reference HcRed reporter plasmid were mixed, and the volume was adjusted with TE buffer. EC buffer (Effectene Kit, Qiagen) supplemented with 0.2 M sucrose (Invitrogen) was added. The mixture was incubated first with Enhancer (Qiagen) to let Enhancer built complex, and then with Effectene reagent. Finally, the samples were mixed with 0.2% (w/v) gelatine to reach the final concentration of gelatin in the samples 0.1%. The amount of DNA and transfection reagents can be scaled up if necessary, while the resulting samples are kept to contain 32 ng/µl of promoter construct and 7.5 ng/µl of reference plasmid. Prepared samples have to be incubated at room temperature for a minimum of 1 hr before spotting. The samples can also be stored for a couple of days at 4 ℃ until spotting (see the sample preparation protocol below).

Spotting sample preparation:

- Mix 1.6 µg of putative promoter plasmids and 400 ng of the reference HcRed reporter plasmid. Adjust the volume with TE buffer to 6.5 µl.
- Add 10 µl EC buffer (Effectene Kit, Qiagen) supplemented with 0.2 M sucrose (Invitrogen).
- Add 1.5 µl Enhancer (Effectene Kit, Qiagen), mix by vortexing briefly.
- Incubate for 5 min at room temperature
- Add 7.0 µl Effectene reagent (Effectene Kit, Qiagen) , mix by vortexing briefly.
- Incubate for 10 min at room temperature.
- Add 25 µl of 0.2% (w/v) gelatin, mix by pipetting up and down 3 times.
- Incubate the samples at room temperature for a minimum of 1 hr before spotting. Otherwise store the samples at 4 ℃ for a couple of days until spotting.

6.1.3.2 Microarray spotting

Self-made VPL slides were used for array printing. To make VPL slides, standard 25×75×1.0 mm slides were covered with poly-L-lysine (Sigma) and VECTABOND Reagent (Vector Labs) as follows. The slides were stored in dark and low humidity conditions prior to use.

VPL slides preparation:

- Shake slides in cleaning solution (70 ml 1.75 M NaOH + 160 ml ddH_2O + 240 ml 100% ethanol) in a glass container for 2 hrs
- Wash slides in ddH_2O 3 × 5 min
- Rinse slides in acetone for 5 min, let slides dry shortly
- Incubate slides in VECTABOND™ solution (7 ml VECTABOND™ Reagent + 350 ml acetone) for 5 min
- Wash slides in ddH_2O 3 × 30 sec
- Dry at 37 ℃
- Put slides into plastic container with poly-L-lysine solution (20 ml poly-L-lysine + 20 ml PBS + 160 ml ddH_2O)
- Shake at 4 ℃ for 45 min
- Wash in ddH_2O shortly
- Dry at 55 ℃
- Store slides in the dark and under vacuum.

Slides can be printed in two ways: automatic spotting and manual spotting.

Automatic spotting

Automatic spotting was performed with instrumentONE (iO, M2 Automation, http://www.m2-automation.de) (Figure 9), a high-precision liquid handling system for micro-dispensing applications. The dispensing volume ranges from nanoliter up to microliter. InstrumentONE is a high speed dispensing system using touch-less technology, which reduces the risk of contamination among different samples. The system has capacity for 20 slides arranged on 5 MTP adapters. Arrays were printed using a 500 µm outlet port solenoid valve, which generates droplets of approximately 20 nl. The diamerter of each spot was approximately 500 µm. Distance between the centres of neighbour spots was set to1500 µm. The frequency of dispensing is 500 Hz (500 droplets/s).

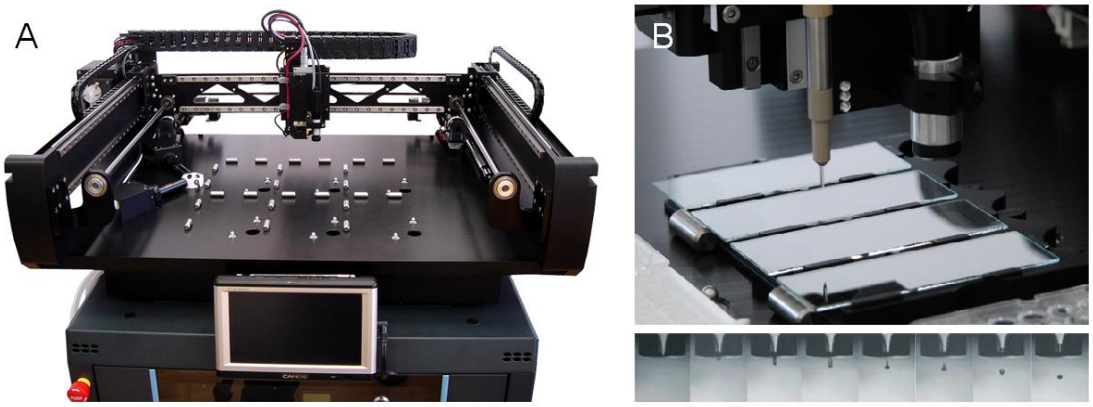


Figure 9. Automatic spotting with InstrumentONE (iO). (A) The iO instrument. used for microarray spotting in this study (B) The micro-dispensing device and a single droplet ejected from the nozzle tip using non-contact printing technique.

Figure 10. Manual spotting. (A) An example of the grid lattice used for manual spotting with the pattern of 1.5 × 1.5 mm square. Therefore the distance between spots was set to 1.5 mm. (B) The procedure of manual spotting. Slides were put on top of the grid lattice, and long tips were used for spotting.

Manual spotting

For small numbers of sample, manual spotting is a useful alternative to automatic spotting systems. For manually prepared cell arrays, a grid with desired size were designed and printed on paper, which in this study is normally 1.5 × 1.5 mm square (Figure 10). Up to 45 × 15 grid lattice can be reached on a standard glass slide. The spotting was carried out in a sterile cell culture hood. The slides were put on top of paper and the spotting was performed following the grid. The samples were spotted with a 2 µl pipette and long tips (SAFESEAL tips PREMIUM XL 10 µl, Biozmy) by quickly and slightly tapping on the slide with the filled tip (Figure 10). For each sample, a batch of slides were spotted using a same tip. The samples were spotted at the center of each square without surpassing the lines, so one spot was separated from its adjacent spots and the distance between the neighboring spots was about

1.5 mm. Each spot had a diameter of about 0.5 to 1.0 mm and was formed by spotting about 10 nl.

For both automatic and manual spotting, all samples were arrayed in triplicates. After spotting, the slides were air-dried for minimum 1 hr before transfection. For long-term use slides were stored in dark and low humidity conditions at 4 ℃ to maintain the effects of transfection reagents. Long-term storage was done in slide boxes placed in a Rotilabo® Drying cabinets (Roth) filled with Silica Gel drier pellets and stored in a 4 ℃ room.

6.1.4 Reverse transfection of cell array

6.1.4.1 Cell culture preparation

HEK293T/17 (ATCC, CRL-11268), a cultured human embryonic kidney cell line, was used for reverse transfection. HEK293 was first established in 1977 as a permanent adherent cell line of human embryonic kidney after transformation with human adenovirus type 5 (Graham et al., 1977). HEK293T is a HEK293 derived cell line that constitutively expresses the SV40 large T antigen, and clone 17 was selected specifically for its high transfectability. Expression of the T-antigen can enhance proliferation of cultured human cells (Bednarz et al., 2000; Kahn et al., 1993). Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) containing 4.5 g/L D-Glucose, sodium pyruvate, 25 mM HEPES and phenol red (Gibco Invitrogen) supplemented with 10% (v/v) fetal calf serum (Biochrom), 2mM L-glutamine (Gibco Invitrogen) and 50 µg/ml Penicillin/streptomycin (Gibco, Invitrogen). Cells were handled on a HERAsafe clean bench (Heraeus) and cultured in Tissue Culture Petri Dishes (TPP, Switzerland) in HERAcell CO_2-incubators (Heraeus) at 37 ℃ in humidity condition with 6% CO_2 concentration.

HEK293T cells were subcultured every 3 to 4 days at a density about 2 x 10^5 cells per tissue culture dish (150 × 20 mm) or 150 cm^2 flask. Cells were rinsed with PBS and detached from culture dishes by incubation with a mixture of proteases Accutase (PAA) for 5 min at 37 ℃. To inhibit Accutase, add same amount of complete growth media to the detached cells and mix by gently pipetting up and down for several times. Then centrifuge the cells for 5 min at 2000 rpm. Supernatant was removed by vacuum and the cell pellet was resuspended in fresh media. To determine cell concentration, dilute 5 µl cell solutions to 1:4 with Trypan Blue 0.4% (Sigma-Aldrich) and count the cells using a Hemocytometer (Neubauer counting slide)

through microscope.

For long-term storage, cells were frozen in liquid nitrogen (-196 ℃). Cells were pelleted and resuspended in freezing solution composed of 90% FCS and 10% dimethyl sulfoxide (DMSO) (SIGMA). Transfer 1.5 ml of cells to 2 ml cryovials (GREINERbio-one). Before putting into liquid nitrogen the vials were pre-cooled at -80 ℃ for 24 hrs. For reviving, remove vials from Liquid Nitrogen and immediately transfer to 37 ℃ water bath. Transfer contents of vial to 15 mL test tube when completely thawed. Slowly add 10 mL pre-warmed complete growth media, centrifuge at 1000g for 5 min. Resuspend pellet in 10 ml fresh media and transfer cells to a 100 × 20 mm dish or 75cm^2 flask.

6.1.4.2 Reverse transfection

10 × 10^6 HEK293T cells were prepared and plated in 10 ml complete growth media in a 60 cm^2 cell culture dish 24 hours before transfection. Alternatively, 5 × 10^6 cells could be cultured for 2 days before transfection. 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 ℃ with 6% CO_2 for 48 hrs. After 24 hrs incubation in the complete growth medium, slides were treated with dfferent conditions by changing the medium to (1) complete medium with 20 nM phorbol 12-myristate 13-acetate (PMA), (2) complete growth medium with 200 nM trichostatin A (SIGMA), (3) fetal calf serum (FCS)-free DMEM medium, (4) complete medium with 20 nM sodium arsenite ($NaAsO_2$, Merck), or (5) FCS-free DMEM medium with 20nM $NaAsO_2$. Two slides were used in parallel for each treatment as well as the control condition with the normal complete growth medium.

To stop the transfection, slides were fixed for 30 min in 3.7% formaldehyde with 4 M sucrose in PBS. After that slides were stained with DAPI (4`,6-diamidino-2`-phenylindole dihydrochloride, SIGMA) and mounted with Fluoromount-G (Southern Biotech). Fluoromount-G is a water-soluble, non-fluorescing compound for slides mounted after a staining procedure having an aqueous final step. Mounting slides with Fluoromount-G reduce fluorochrome quenching during analysis of slides by fluorescence microscopy and provides a semi-permanent seal for long-term storage of slide preparations. Slides were covered with 22

×64 mm cover glasses (BDH) and fixed with nail polish at the edges. Covered slides can then be stored at 4 ℃ for months while keeping fluorescent signal.

6.1.5 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 high-resolution 2/3" CCD camera with 41.6 μm^2 pixel size, and applying 10×/1.6 (NA) objective. 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 image sets was generated by the Scan^R Analysis software, an integrated image analysis and data visualisation tool (Olympus). Scan^R Analysis was used to locate each transfected cell by locating the borders of fluorescent objects in the images based on the object finding algorithm using the watershed option for better object separation. Only the signals from ZsGreen1 or HcRed1 reporter constructs, which fluorescence intensity exceeded the threshold set in the green and red channel, respectively, were used to determine the cells co-transfected with both reporters.

In order to determine the activity for each promoter, different analyses based on either fluorescent intensity or numbers of transfected cells, have been applied. With the intensity-based analysis method, transfected cells were located by an object finding algorithm of red cells (Figure 11). The fluorescent intensities of both green and red reporters (G_i and R_i) and the ratio of green-to-red signal (T_i), were determined for each transfected cell. The promoter activity was quantified as a mean green-to-red ratio ($\overline{T_i}$) of all transfected cells within one spot (Figure 11). The average promoter actibity value was calculated for three spots. A threshold for active promoter was set as three standard deviations above the mean G/R ratio of the negative control (attP-containing pZsGreen1 vector) spotted in 10 replicates. If a promoter was scored as active in two parallel slides, then this promoter was classified as active. Herewith a binary promoter activity index, with 0 for silent and 1 for active promoter, was generated for each promoter region under investigation.

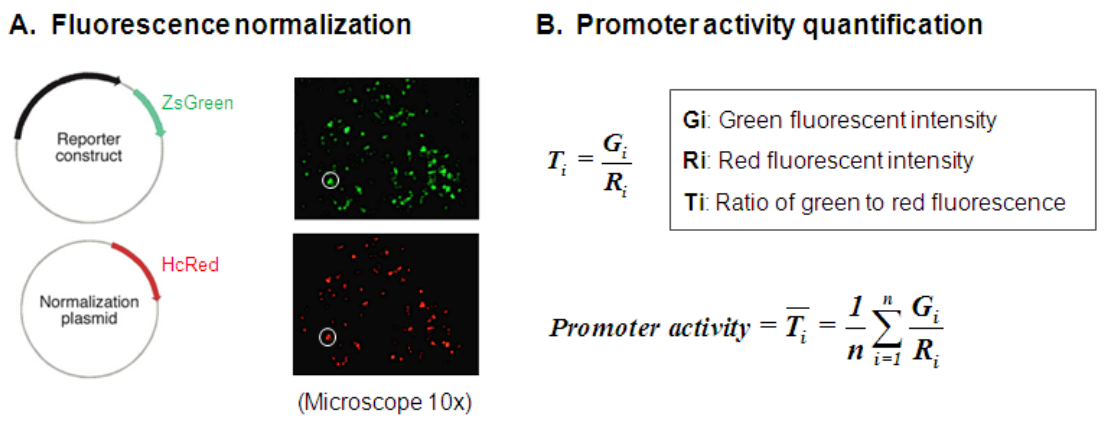


Figure 11. Normalization and quantification of the fluorescent intensity-based analysis. (A) Transfected cells were located by an object finding algorithm of red cells. (B) For each transfected cell, the fluorescent intensities of both green (G_i) and red (R_i) reporters and the ratio of green-to-red signal (T_i) were determined. The promoter activity was quantified as the average T_i of all transfected cells within one spot.

For the transfected cell number-based method, two parameters were taken into account: (1) number of co-transfected cells ($N_{co\text{-}tr}$) and (2) ratio (R_i) between the number of co-transfected cells and the total number of red cells (N_{red}), $R_i = N_{co\text{-}tr}/ N_{red}$. Regarding the former parameter, the threshold for the number of co-transfected cells was defined as three standard deviations (SDs) above the mean number of ZsGreen1-positive cells transfected with negative controls. The R_i parameter served as a transfection efficiency index and was empirically set to a value that provided the best correlation of promoter activity with the corresponding gene expression level, as determined by transcriptome sequencing. If the R_i parameter was greater than the value in at least four out of six replicates, the promoter was classified as active. Herewith a binary promoter activity index, with 0 for inactive and 1 for active promoter, was generated for each promoter region under investigation.

6.2 Computational and functional analyses

6.2.1 Computational sequence analysis of promoter regions

The predicted promoter sequences were analyzed, and common promoter elements (TATA box, INR element, DPE elements and CpG islands) were mapped. The TransFac tool MATCH (Kel et al., 2003) was used for the detection of TATA box, INR and DPE elements using the position weight matrices. The CpG islands was determined by overlapping the sequences of cloned fragments with the map of *bona fide* CpG islands predicted by (Bock et al., 2007), and

the recommended combined epigenetic score (greater than 0.5) was used. The computational sequence analysis for core promoter elements was performed in cooperation with Dr. Robert Querfurth and Dr. Thomas Manke from the MPIMG, Berlin.

6.2.2 Comparison of promoter activities with UniGene expression profiles

UniGene cluster IDs of genes corresponding to cloned promoters were identified by conversion from Entrez GeneIDs using DAVID Gene ID Conversion tool, followed by manual conversion for missing ones. The UniGene EST clusters were retrieved from UniGene FTP site as the file named "Hs.profiles.gz". EST expression profiles for these UniGene clusters were extracted from the "Body Sites" category of the retrieved. For each gene corresponding to the studied promoter fragments, the number of tissues where ESTs were expressed was determined. Subsequently, the tissue distribution of the ESTs was analyzed in relation to the promoter activity of the corresponding gene by comparing with the fraction of active and silent fragments.

6.2.3 Prediction of transcription factor binding site affinities

The promoters that showed specific response patterns in our experiments were examined for common transcription factors binding sites that might explain these responses. Statistical modeling of transcription factor binding affinities was performed as described before (Manke et al., 2008). The matrices describing 610 vertebrate transcription factor binding preferences were provided by the database from TRANSFAC version 12.1 (Wingender et al., 2001). For each binding matrix, the affinity of the matrix for each sequence was calculated and then transformed into p-values as described before (Manke et al., 2008). 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.

6.2.4 Enrichment analysis of gene ontology annotations

Enriched gene ontology (GO) terms were identified using the DAVID functional annotation tool (Dennis et al., 2003). Entrez GeneIDs for promoters activated by different external stimuli were compared to a background set of 126 inactive promoters within the GO category 'biological process'. The gene-term association matrix of the top identified cluster (six enriched categories associated to nine genes) was used for further analysis.

7 RESULTS

7.1 Detection of promoter activities by transfected-cell array

The scheme of working procedures for the detection of promoter activity is illustrated in Figure 10. Briefly, putative promoter regions were defined as the DNA sequences relative to the most upstream TSS of human transcripts based on the Ensembl database (Figure 12A). The designed human promoter fragments were cloned into an attP-site adapted promoter-less vector containing the coding sequence for a fluorescent protein (Figure 12B). Purified promoter-reporter constructs, together with the normalization plasmid encoding a red fluorescent protein, were mixed with transfection reagents and gelatin solution and spotted onto glass slides that were treated with VECTABOND Reagent and poly-L-lysine for better retention of DNAs and cells (Figure 12C). The slides were covered with HEK293T cells and incubated at 37℃ for 48 hours for reverse transfection (Figure 12D). Finally, slides were scanned with an automted fluorescence microscope, and promoter activities were analyzed by the measurement of the fluorescent reporter expression (Figure 12E). The detailed protocols can be found in the Methods.

7.1.1 Amplification and cloning of human promoters

TSSs of human transcripts were determined based on the Ensembl database. Two sets of promoters, one from the genes of chromosome 21 (Chr21) and the other for the genes involved in inflammatory pathways, were designed and amplified using the same methods. The putative promoter fragments were amplified using the optimized touchdown PCR protocol (Ralser et al., 2006). Genomic DNA as well as available genomic BAC and fosmid clones were used as templates. Then the promoter fragments were cloned into the attP-adapted pZsGreen1-1 vector upstream of the fluorescent reporter gene ZsGreen1. The insertions were verified by sequencing from both 5’ and 3’ ends. As a result, 182 promoter fragments from Chr21 genes and 74 from inflammatory genes were confirmed by sequencing. The promoter of gene RIPK4 was amplified in both sets, therefore, in total 255 specific promoter fragments were successfully cloned. The information of those cloned promoters is listed in Table 7.

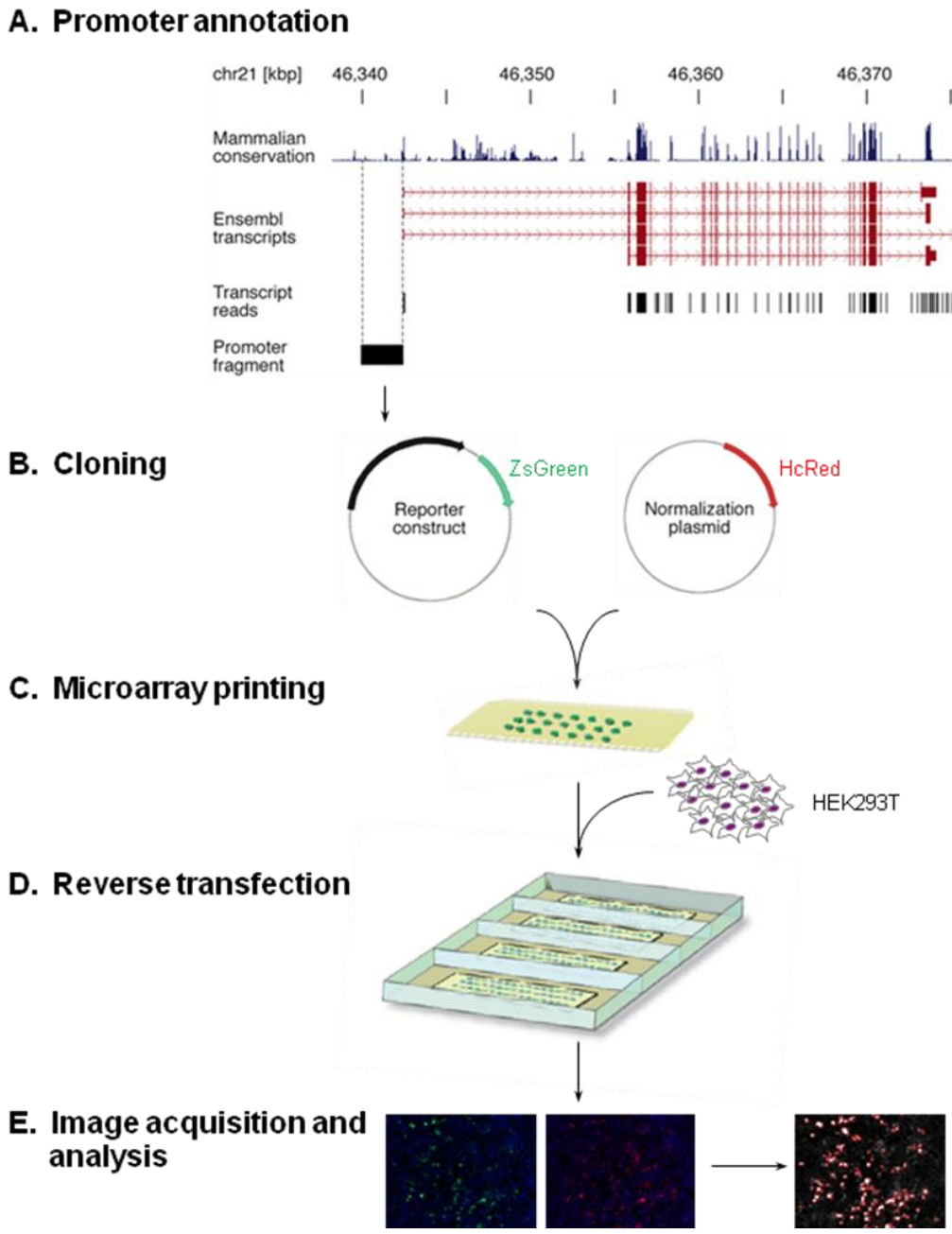


Figure 12. Workflow for promoter analysis using the transfected cell array. (A) Promoter regions were selected upstream of transcription start sites from the Ensembl database and amplified by PCR. (B) Amplified promoters were cloned into a reporter plasmid encoding a green fluorescent protein. (C) Reporter constructs were spotted onto glass slides together with a normalization plasmid encoding a red fluorescent protein. (D) The slide was covered with HEK293T cells for reverse-transfection. (E) Images were acquired and analyzed by fluorescence microscopy.

For a subset of promoters where the determined TSS corresponded to an entry in the database of transcriptional start sites (Wakaguri et al., 2008), truncated fragments were designed in order to compare the activities of longer promoter versus their truncated counterparts and

therefore to understand the functional elements within the extended promoter region. In all, 62 truncated fragments for Chr21 promoters, with the length of 500 bp, have been designed, amplified and cloned into the attP-adapted pZsGreen1-1 vector.

7.1.2 Determination of the transfection efficiency

The transfection efficiency for TCA was calculated as the ratio of the number of transfected cells to the total number of cells in each spot area. To determine the transfection efficiency for the arrays, all cells were marked with a DNA fluorescent dye DAPI, the slides were scanned and image data for total cell counts and green or red fluorescence positive cells were collected using a high throughput fluorescence microscopy system Scan^R (Olympus). The number of DAPI-stained cells in each scanning image frame (672 × 512 μm) (Figure 13A) was counted automatically by the Scan^R analysis software (Olympus). As the diameter of the single DNA spot was measured to be approximate 500 μm, the number of cells per spot could be calculated according to the area ratio. The average number of cells per spots was calculated from all the six spots in two microarray slides.

The transfected cells were considered as the cells positive for either green or red fluorescence. The numbers of cells expressing red (Figure 13B), green (Figure 13C), or dual fluorescence (Figure 13D, shown in yellow) were automatically counted by Scan^R software (Olympus). The number of transfected cells was calculated as the sum of cells positive for green and red fluorescence minus cells expressing dual fluorescence, so that the cells that expressed both fluorophores were counted only once. The average number of transfected cells for each promoter was calculated from all the six spots in two microarray slides. Thus, the transfection efficiency was calculated as the ratio of the average number of transfected cells to the average number of cells in the spot area (Table 3).

The sets of Chr21 and inflammatory promoters were investigated in separate TCA experiments, the processes of the experiments were almost same except the way of array printing. For the bigger set of Chr21 promoters the microarrays were spotted by robot, whereas for the smaller set of inflammatory promoters the arrays were prepared manually. The transfection efficiencies of the two sets were calculated in the same way and compared.

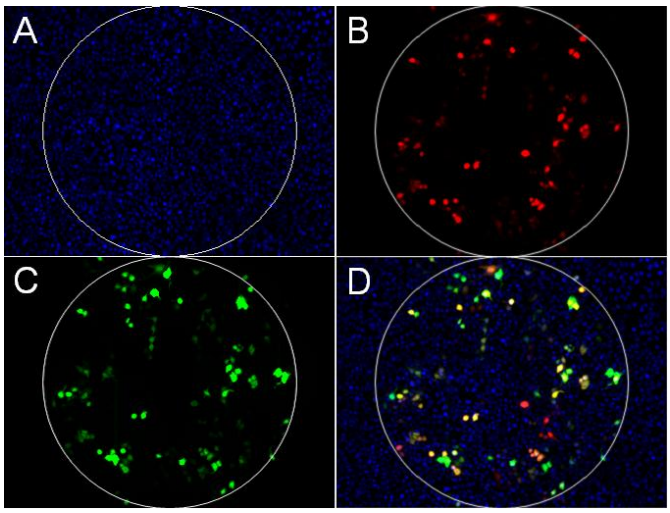


Figure 13. Cell number counting and transfection efficiency calculation. The transfection efficiency for each spot on the array was calculated as the ratio of the number of transfected cells to the total number of cells in the spot area. (A) The total number of cells per scanning frame was counted by fluorescent DAPI. The number of cells that could theoretically be transfected (within the spot, shown in white circle) 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.

Table 3. Transfection efficiency of TCA experiments performed with microarrays spotted with robot and manually.

	Arrays spotted automatically (Chr21 set)	Arrays spotted manually (Inflammatory set)
Average number of cells per image frame	649	1009
Average number of cells per DNA spot	370	576
Average number of transfected cells per spot	68	96
Transfection efficiency	18.4%	16.7%

The transfection efficiencies of TCA performed with arrays spotted automatically and manually were 18.4% and 16.7%, respectively (Table 3), which were both in agreement with previous results for liposome-based reverse transfection, namely 17% efficiency (Baghdoyan et al., 2004). The median number of transfected cells in our robotic spotting was in agreement with the previous results of 30-80 fluorescent cells per spot (Ziauddin and Sabatini, 2001). Because of the similarity of transfection efficiency in robot and manual spotting arrays, the results from the two sets of promoters have been combined for the following analyses.

7.1.3 Quantification of promoter activities on cell arrays

Promoter activities in transfected cells can be quantified either by integration of fluorescence intensities over all cells or by the measurement of intensities in single cells. Cell-based image analysis was described to exhibit advantages compared to whole-well or whole-area analyses with respect to increased sensitivity and superior data extraction for samples with low transfection efficiency (Harada et al., 2005). Therefore, promoter activities were determined using the cell-based quantification approach.

As described in the Methods, the promoter activity could be quantified based on either fluorescence intensity or the number of transfected cells. By using the intensity-based method, the mean G/R ratio for each promoter was determined, and the threshold for active promoters was considered as three standard deviations above the mean ratio of the negative control. With this approach, 61 promoters were identified as functionally active under standard cell culture conditions in HEK293T cells (Table 10). In the transfected cell number-based quantification analysis, two thresholds, a minimal number of co-transfected cells and a minimal transfection ratio index, was utilized. A promoter fragment was considered as functional if its activity exceeded the combined threshold for at least four replicates out of six. By this approach, 71 functional promoters were identified in HEK293T cells (Table 10). The two quantification methods showed correlation coefficient of 0.68 (Table 4).

When the growth conditions were changed by the addition of small chemicals or by the depletion of serum (see Section 6.6), the discrepancy between the results from two quantification analyses increased (Table 4). A dramatic increase of green fluorescence intensity for many promoter-reporter constructs was observed by visual observation when the cells were treated with external stimuli (data not shown). The induced expression of reporter protein suggested that the promoters were activated by the treatment. However, the activation could not be determined by measuring G/R ratio, because not only the green fluorescence intensity of the promoter reporter constructs, but also the red fluorescence intensity of normalized plasmid was increased. As a result, the G/R ratio remained similar or even decreased to some treatment (Table 10), although the fluorescent intensity was increased. In contrast, by using cell number-based approach, most promoters with increased reporter expression were determined as active, which correlated with the visual observation.

Table 4. The quantification analyses based on fluorescence intensity and transfected cell number.

Cell growth condition	Active promoters detected by fluorescence intensity-based quantification	Active promoters detected by transfected cell number-based quantification	Correlation coefficient of two quantifications
Standard medium	61	71	0.68
PMA treatment	49	134	0.34
TSA treatment	50	96	0.47
Serum depletion	59	109	0.54
ARS treatment	64	87	0.65
Serum depletion & ARS treatment	76	67	0.66

7.2 Correlation of promoter activity with further data

7.2.1 Endogenous transcript levels correlate 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 was 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. The comparison of promoter activity with the levels of corresponding mRNA provides the ratio of false positives 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). Among 255 genes corresponding to the cloned promoters, in HEK293T cells 99 genes were expressed, 32 genes potentially expressed and 124 were not expressed. Of the 71 active promoter fragments, 58 of the corresponding genes were expressed (Figure 14A). In contrast, genes associated with silent promoter fragments are less likely to be expressed. Of the 184 silent promoter fragments, only 41 of the corresponding genes were expressed (Figure 14B). The data sets of endogenous transcript levels allowed us to determine true-positive and false-

negative rates for promoter activities measured in transfected cell arrays. Regarding true-positives, 81.7% (58/71) of the active promoters were found to overlap with the corresponding gene expression; whereas regarding false-negative promoter activity, 22.3% (41/184) of the inactive promoter fragments were associated with endogenous transcripts identified by RNA-Seq.

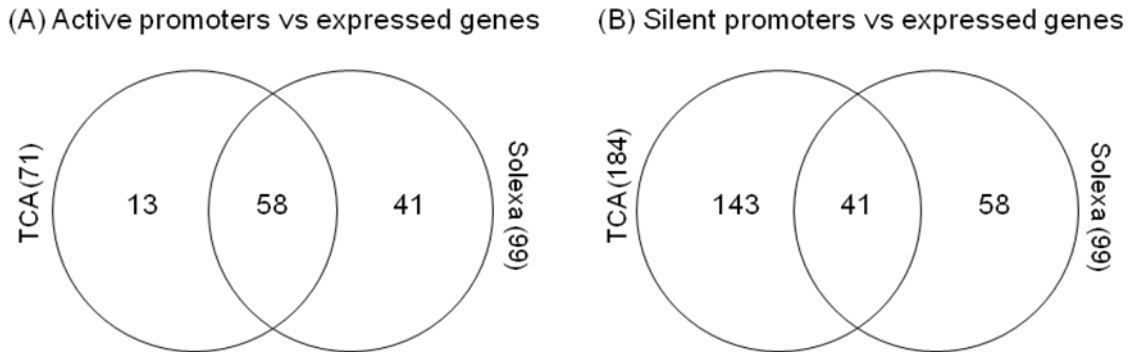


Figure 14. Correlation of promoter activity with endogenous gene expression in HEK293T cells. The Venn diagrams show the degree of agreement between the promoter reporter activities of tested fragments and expression of the associated genes from high-throughput RNA-Seq. (A) Active promoter fragments show 81.7% overlap with corresponding gene expression; (B) genes associated with silent promoter fragments are less likely to be expressed (22.3%).

7.2.2 Co-occurrence of active promoters with RNA Pol IIa binding sites

To assess promoter activity with the presence of functional TSSs, the occupancy of hypophosphorylated RNA polymerase II a (Pol IIa), which has been used as a landmark of transcription initiation (Brodsky et al., 2005), was correlated with the tested promoter fragments. Pol IIa binding sites were previously identified by chromatin immunoprecipitation of Pol IIa from HEK293T cells, followed by high-throughput sequencing (ChIP-Seq) (Sultan et al., 2008).

The overlap between nearby endogenous Pol IIa binding site clusters and promoters was analyzed for all the cloned 2.5 kb fragments. Among 255 cloned promoter fragments, 52 were observed to overlap Pol IIa-binding regions, including 38 active promoter fragments and 14 silent promoter fragments (Table 11). Over half of the active promoter fragments (53.5%, 38 out of 71) were associated with Pol IIa binding sites; in contrast, only 7.6% (14 of 184) of silent promoter fragments were found to overlap Pol IIa islands. The data indicates that active promoter fragments are more likely associated with Pol II binding sites, whereas inactive

promoter fragments are virtually depleted of Pol IIa (Figure 15).

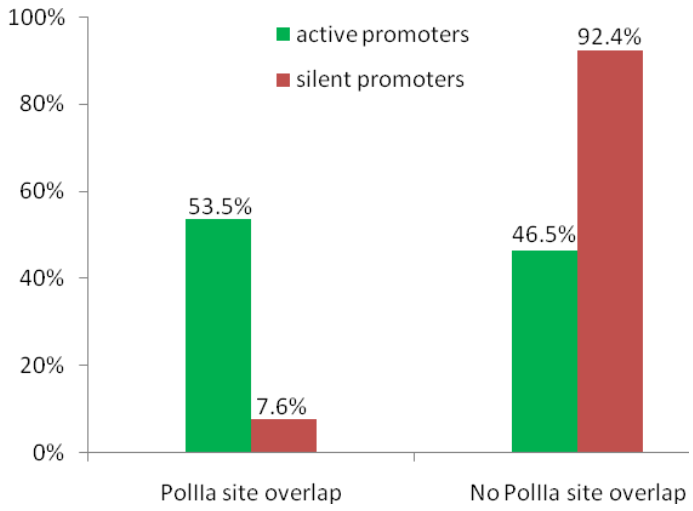


Figure 15. Overlap of active and silent promoter fragments with PolIIa binding sites. The overlap between Pol IIa binding site clusters and tested promoters was analyzed for all 255 cloned fragments. The histogram shows the correlation between promoter activity and Pol II sites. 53.5% of active promoter fragments were associated with overlapping Pol IIa sites. 46.5% of active promoters contained no Pol IIa site. In contrast, only 7.6% of silent promoters were associated with overlapping Pol IIa sites; no Pol IIa site was found in 92.4% of the silent promoters.

7.2.3 Active promoters are enriched for genes with ubiquitous expression

The reverse transfection experiments were performed in the human embryonic kidney cell line HEK293T, which suggests that the promoters for genes with more tissue-specific types of activity might not be functional in this cell line. In contrast, the genes with more ubiquitous character of expression would be more likely expressed in this cell line. Therefore, the correlation between activities of promoters transfected into the HEK293T cells and the tissue expression profiles of the corresponding genes were evaluated. For this purpose, the UniGene EST clusters of genes corresponding to the studied promoter fragments were retrieved. In order to estimate the level of tissue-specificity of genes represented by the ESTs, the numbers of tissues where particular ESTs were identified were counted. Subsequently, the tissue distribution of the ESTs was analyzed in relation to the promoter activity of the corresponding gene. Using this approach, the tissue expression data for 220 corresponding genes were obtained, including 69 active fragments and 151 silent fragments. A strong correlation of the number of tissues with corresponding ESTs to the promoter activity state in HEK293T cells

was observed. The corresponding genes of 69.6% active promoters were expressed in more than 25 out of 45 tissues. In contrast, only 25.2% of the inactive promoters belong to the category of genes with broad expression profile. Furthermore, about half (50.3%) of the inactive promoters were expressed in fewer than ten tissues (Figure 16).

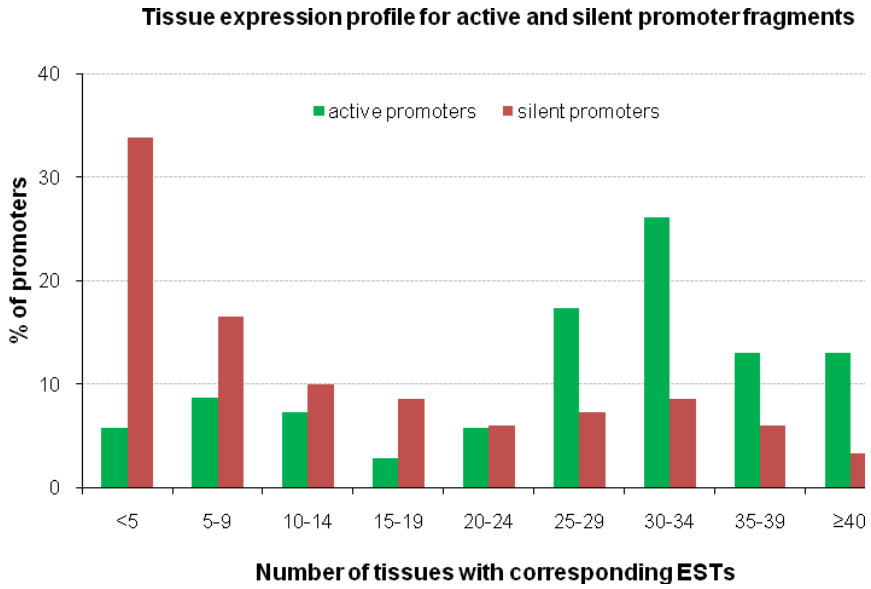


Figure 16. Tissue expression profiles for active and silent promoter fragments. EST profiles from 45 tissues for genes corresponding to 220 out of 255 cloned promoters were obtained from the UniGene database. For each gene, the number of tissues in which ESTs have been observed was counted and bins of these EST tissue numbers were plotted against the fraction of each active and silent promoter. The corresponding genes of 69.6% of active promoters were expressed in more than 25 out of 45 tissues, while 50.3% of the inactive promoters were expressed in fewer than ten tissues.

7.2.4 Sequence characteristics of putative promoter fragments

Promoters vary in their sequence structure; however, they 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 promoter activity, these elements are necessary for binding of general transcription factors and RNA polymerase II into the pre-initiation complex (Muller et al., 2007; Sandelin et al., 2007). To determine whether the observed activities of the cloned promoters are related to the sequence elements, the promoters were analyzed for the occurrence of these four elements within the proximal 500 bps of the TSSs of all cloned fragments.

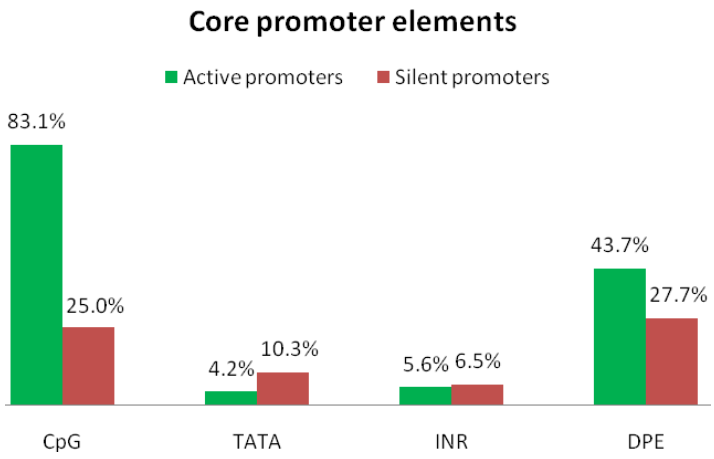


Figure 17. The presence of promoter sequence elements in cloned promoters. The promoter sequence motifs, such as CpG islands, TATA boxes, INR elements and DPE elements, were detected in the cloned promoter fragments. A three-fold enrichment of CpG islands was observed in active promoters (83.1%) compared to silent fragments (25.0%). In contrast, no enrichment of TATA boxes was observed in either active fragments (4.2%) or silent promoters (10.3%). Moreover, no clear-cut correlation between the presence of INR or DPE and promoter activity was observed.

CpG islands are known to play a critical role in gene regulation. They are often associated with the promoters of most housekeeping genes and many tissue-specific genes. Aberrant methylation of promoter-related CpG islands is a common mechanism of inactivation of tumor suppressor genes and the epigenetic causes of all types of cancer (Herman and Baylin, 2003). Different approaches have been described and developed for the computational prediction and identification of CpG islands (Hackenberg et al., 2006; Takai and Jones, 2002; Wang and Leung, 2004), most of which are based only on sequence criteria. Recently, a high-quality map of so-called bona fide CpG islands in the human genome was generated by (Bock et al., 2007). They use a CpG island strength prediction method that links the DNA sequence of CpG islands to their epigenetic states, including DNA methylation, histone modifications and chromatin accessibility, in order to characterize the CpG islands in the promoter fragments (Bock et al., 2007). Based on this map, 105 (41.6%) out of 255 cloned promoters, including 59 (83.1%) of the 71 active promoters and 46 (25.0%) of the 184 silent promoters, were observed to overlap with a CpG island over a sequence length of at least 50 bps (Figure 17). The enrichment of CpG islands in active fragments was over three-fold compared to silent fragments. These data were in line with the description by Cooper et al. (Cooper et al., 2006) that the presence of CpG islands is important as an indicator of an active promoter, however, it is not sufficient for 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 its binding initiates the assembly of the pre-initiation complex. TATA boxes are often associated with tissue-specific promoters and correlated with a clearly defined TSS (Sandelin et al., 2007; Schug et al., 2005). TATA boxes were present in only 22 (8.6%) of the 255 cloned promoters. Regarding reporter activity of the 40 TATA-containing promoters, TATA boxes occurred in 10.3% of silent fragments (19 of 184) and 4.2% of the active fragments (3 of 71) (Figure 17), an enrichment that was calculated to be non-significant. Moreover, the sequences of promoter fragments were analyzed for the initiator (INR) element and downstream promoter elements (DPE). No trend was observed for the INR element, which was present in 5.6% of active fragments (4 of 71) and in 6.5% of silent promoters (12 of 184). A slight enrichment of DPE elements were observed for active promoters, as it was found in 43.7% of the active promoters (31 of 71) and in 27.7% of silent promoters (51 of 184) (Figure 17). However, no clear-cut correlation between the presence of these motifs and promoter activity was observed, which is in agreement with previous work (Cooper et al., 2006; Trinklein et al., 2003). Though TATA box, initiator and downstream promoter elements are functionally important, they are not always required for promoter activity (Muller et al., 2007; Sandelin et al., 2007).

7.3 Activation of silent promoters by further treatments

As described above, only 27.8% (71/255) of cloned fragments were identified as functional promoters in HEK293T cells. It would be interesting to know if the remaining 184 silent promoters could also drive gene expression when activated by any external stimuli, or if the activities of the 71 functional promoters would be affected by the stimuli treatments. For this the HEK293T cells were treated with different stimuli: (1) phorbol 12-myristate 13-acetate (PMA), (2) Trichostatin A (TSA), (3) depletion of serum from the medium, (4) arsenite and (5) serum depletion combined with arsenite treatment. Figure 16A shows an activity matrix of promoters that change their activity upon different treatments.

Phorbol 12-myristate 13-acetate treatment

Phorbol 12-myristate 13-acetate (PMA), also known as 12-O-Tetradecanoyl- phorbol-13-acetate (TPA), is the most common phorbol ester and has been used as biomedical research tool in models of carcinogenesis. The most notable biological property of phorbol esters is the

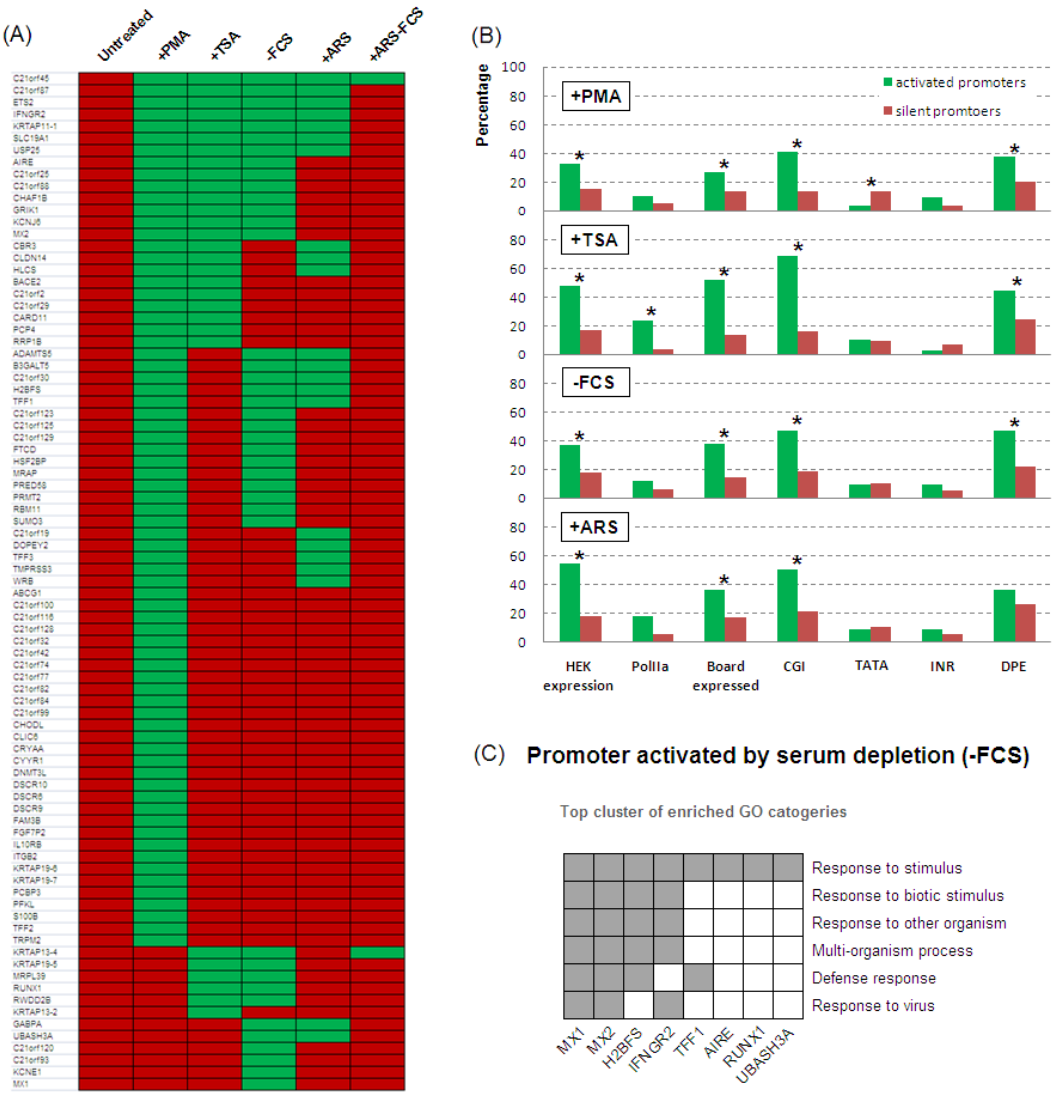


Figure 18. Activation of promoters after treatments. (A) The figure shows activity matrix of 85 promoters that were previous silent but induced to be active after treatments of HEK293T cells with the addition of Phorbol 12-myristate 13-acetate (+PMA), addition of Trichostatin A (+TSA), depletion of fetal calf serum (-FCS), addition of arsenite (+ARS) and the combination of the -FCS and +ARS. Green and red boxes indicate active and silent promoters detected by the TCA assay, respectively. (B) Bar diagram shows the association of endogenous gene expression (HEK expression), Pol IIa binding region (Pol IIa), ubiqutinous tissue expression (Board expression) and promoter sequence elements (CGI, TATA, INR and DPE) with activated and silent promoter fragments in each treatment. Green and red bars represent the fraction of promoters that are activated by the treatment and remain inactive in each condition. Asterisks indicate the enrichments were considered as significant **(**hypergeometric test, $p<0.01$) (C) Six GO categories associated to eight genes are enriched in the set of promoters activated by serum depletion, compared to the background of all 182 cloned promoter fragments.

capacity to act as tumor promoters through activation of protein kinase C (PKC) (Blumberg, 1988). Phorbol esters are functional analogs of the endogenous PKC activator diacylglycerol (Adams et al., 2000). By mimicking DAG, they modulate PKC signaling pathways and diverse cellular responses including gene transcription, cellular growth and differentiation and apoptosis in many cells. PMA has been found to up-regulate the expression of a variety of genes in different cells through the activation of PKC (Batarseh et al., 2008; D'Angelo et al., 1996; Jalagadugula et al., 2008; Keen et al., 2006; Liu et al., 2002; Malakooti et al., 2005; Rowe et al., 2008; Yano et al., 2008). In contrast, PMA was also found to inhibit EDAG gene expression by down-regulating the transcription factor GATA-1 through the PKC/MAPK signaling pathway (Li et al., 2008).

Besides, PMA could stimulate the some gene promoter activity (hNHE3) in a PKC-independent manner (Malakooti et al., 2006). Various DNA elements, including the binding sites of the transcription factors such as AP-1, NF-kB, SP-1, GATA-1, as well as several unidentified DNA sequences, have been described to be responsive to PMA-treatment in the promoters of several genes (Angel et al., 1987; D'Angelo et al., 1996; Keen et al., 2006; Malakooti et al., 2005).

PMA treatment activated 73 of 184 inactive promoters, whereas 10 of 71 previously active promoters were silenced (Figure 18A, Table 11). Significant enrichments of endogenously expressed genes were observed (Figure 18B). Moreover, the enrichment for the core promoter elements of CpG islands and DPE were found to present in 41% and 38% of PMA-activated promoters respectively, whereas TATA box present in 14% of promoters that remain inactive after PMA treatment. No enrichment for INR elements was detected (Figure 18B). However, the pattern change of promoter activity by PMA was not in an entirely reproducible manner.

Trichostatin A treatment

Trichostatin A (TSA) is a specific histone deacetylase (HDAC) inhibitor that induces histone hyperacetylation and thereby enhances transcription by reducing the affinity between histones and DNA and increasing the accessibility of promoter region for RNA polymerase and transcription factors. Although transiently transfected plasmids are not subject to the same regulatory mechanisms as native chromatin (Smith and Hager, 1997), it has been shown that repressive chromatin structures can be formed on a transiently transfected promoter construct

and released by TSA treatment in an SP1-factor dependent manner (Luciakova et al., 2000). Similarly, an SP1-dependent mechanism of gene expression that can be reversed through activation by TSA was shown for other genes (Huang et al., 2005; Yokota et al., 2004). It would be interesting to see if the transcription activity of our cloned promoter fragments would be influenced by TSA treatment.

TSA treatment activated 29 of the 184 inactive promoters; in contrast, only 4 of the 71 previously active promoters were silenced (Figure 18A). A large fraction of the TSA-activated genes (52%, 15 of 29) exhibited a broad expression profile (ESTs in over 25 tissues), in contrast to those which remained silent (14%, 21 of 155). Significant enrichments of endogenously expressed genes and Pol IIa binding regions were observed. Moreover, the core promoter elements of CpG islands and DPE were found to present in 69% and 45% of TSA-activated promoters respectively, whereas no enrichment for either TATA or INR elements was observed (Figure 18B).

Serum depletion

Serum depletion is a model of cellular stress that is associated with depletion of growth factors and nutrients and may elicit cell death through oxidative stress and subsequent apoptosis (Le Gall et al., 2000; Li et al., 2006; O'Brien et al., 1996), through, for example, the activation of NFkappaB (Grimm et al., 1996; Leicht et al., 2001) or the transcription factor CREB after serum removal from cardiac fibroblasts.

In this experiment, two promoters were silenced after depleting serum from the TCA assay for 24 hours; in contrast, 40 promoters were activated (Figure 18A). Significant enrichments of CpG islands and DPE elements were observed, both of which present in 48% of promoters activated by serum depletion (19/40), but only in approximate 20% of promoters that remained silent (Figure 18B).

An enrichment of genes associated with cellular responses to the environment was observed when looking at the functional annotations of promoters activated by serum depletion. Twenty percent (8 of 40) of the promoters that were activated by serum depletion correspond to genes associated with cellular responses to the environment (Figure 18C). In contrast, only 8% of the promoters remaining inactive belong to this category. Although it was unexpected to find significant enrichments in this small data set, our results were in line with the previous

observations, that stress conditions activate genes that are involved in cellular response to external stimuli (Murray et al., 2004).

Arsenite treatment and serum depletion combined with arsenite treatment

Generally, the effects of arsenite in cell culture are multiple and dependent on the dose applied. At higher doses, arsenite treatment in human skin fibroblast cell culture was shown to enhance several signaling-related, cell cycle control, stress-responsive genes and several transcription factors, for example, HSF2, ZFP36, MYCL1, WT1, ATF2, SP3, MAD and GTF6L1 (Yih et al., 2002). At the same time, the activities of certain genes, also those of some transcription factors such as JUNB, PML EGR1 and ELK1, were reduced. At lower concentrations, as in the present study, arsenite leads to DNA repair and anti-oxidation activation, i.e., the induction of cellular protective effects. Also, exposure to lower doses has a stimulating effect on cell proliferation (He et al., 2007). In our TCA assay, 22 promoters were activated by arsenite treatment, and 6 previously active promoters became silent. When treating the transfected-cell arrays with a combination of serum depletion and arsenite, 38 promoters that were activated through serum depletion were silenced again, 14 promoters were active under both treatments, and 8 promoters were activated after arsenite addition, but not by serum depletion alone (Figure 18A).

Altogether, treatment HEK293T cells with external stimuli activated the promoter fragments that were previous silent in the untreated condition. Eighty-five previous inactive promoters were observed to drive gene expression under at least one of the tested circumstances. In contrast, 99 fragments could not be activated by any treatment and remained inactive (Table 11). The promoters activated by each treatment except the combination of serum depletion and arsenite treatment, showed enrichment of gene with endogenous gene expression in HEK293T cells, broad expression patterns and CpG islands (Figure 18B). TSA-activated promoters were enriched for genes marked by RNA Pol IIA occupancy. Interestingly, significant enrichment of core promoter element DPE was observed for the treatment of PMA, TSA and serum depletion.

7.4 Presence of regulatory elements within extended promoter region

To further understand the functional elements within the extended promoter region, promoter-reporter constructs with truncated promoter fragments were generated. For a subset of 62

promoters, 500 bp promoter fragments were cloned, thus removing the distal ~2,000 bases (Figure 19A). The reporter activities of 500 bp fragments were analyzed (Table 12) and compared with their 2.5 kb counterparts. Under standard conditions, 29 (of the 62) shorter

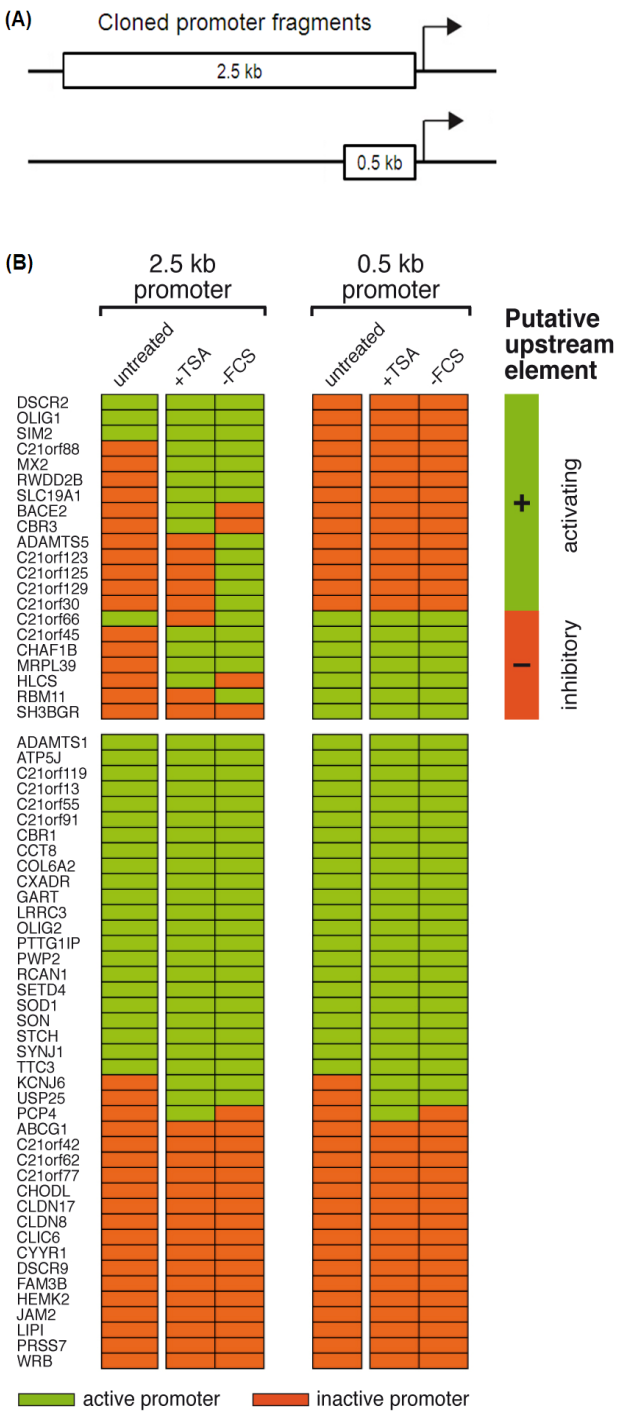


Figure 19. Truncation of promoter fragments can result in loss of responsiveness to external stimuli. (A) For a subset of promoters, truncated promoter fragments were cloned, and the reporter activities were compared with their longer counterparts. (B) The panels show activity matrix of 62 promoters with their long (2.5 kb) and truncated fragments (500 bp) assayed under standard growth conditions (untreated), after addition of 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 (-2,500 to -500 bp) can be inferred from the observed results of truncation, namely activating upstream elements (14 promoters) or inhibitory elements (7 promoters).

fragments could drive transcription in reporter assays (Figure 19B). 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 ~2,000 bases.

Regarding the response of short fragments to treatments with TSA and depletion of serum, 40 short promoters were found to recapitulate, under all conditions, the activity patterns observed for the long fragments (Figure 19B, lower part), while the other 21 behaved differently (Figure 19B, 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, 7 truncated promoters could be activated under more conditions than their longer counterparts, suggesting that they have lost inhibitory elements located in the distal ~2,000 bp sequences.

Table 5. Top-enriched transcription factor binding matrices among promoters that loss respond to external stimuli activation (Warnatz et al., 2010, paper submitted).

Promoter response category (no. of promoters; analyzed regions)	Enriched TF binding matrix	Combined p-value*	Promoters with motif	Transcription factor
Activation by Trichostatin A (n=28; full 2.5 kb regions)	V$POU3F2_01	0.00127	86%	*OCT7*
	V$TCF11MAFG_01	0.00355	93%	*MAFG*
	V$AP1_Q2	0.00372	71%	*AP1*
	V$MEF2_01	0.00432	86%	*MYEF2*
Loss of activation by Trichostatin A after truncation to 500 bp (n=6; 2 kb distal regions)	V$P53_01	0.015	50%	*p53*
	V$OCT1_01	0.0258	67%	*OCT1*
	V$MAF_Q6	0.0345	83%	*MAF*
	V$TEF1_Q6	0.0396	83%	*TEAD1*
Activation by serum depletion (n=40; full 2.5 kb regions)	V$USF_C	0.000104	48%	*USF1*
	V$OSF2_Q6	0.000155	80%	*RUNX2*
	V$NFKAPPAB_01	0.00279	85%	*NFkB*
	V$MYC_Q2	0.00355	48%	*MYC*
Loss of activation by serum depletion after truncation to 500 bp (n=9; 2 kb distal regions)	V$NFKAPPAB_01	0.00287	89%	*NFkB*
	V$NERF_Q2	0.007	78%	*ELF2*
	V$ETS1_B	0.00938	100%	*ETS1*
	V$MAF_Q6	0.00979	89%	*MAF*

*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 potential regulatory elements that contribute to the response patterns observed after treatment with external stimuli were identified. Affinities for 610 known vertebrate transcription factor (Murray et al.) binding matrices (Wingender et al., 2001) were ranked for the 2.5 kb sequence of promoters activated by serum depletion or TSA treatment, and for the

distal 2,000 bp region of promoters which had lost their capacity to be activated by either of those stimuli upon truncation. Of the TF binding matrices enriched in the sequences from each class of promoter, only TFs expressed in HEK293T cells were retained. The top-enriched matrices of the identified TFs are listed in Table 5. Results are in accordance with previous reports that demonstrated the importance of MAFG, AP1 (FOS/JUN), p53 and OCT1 in TSA responses and associated histone deacetylase inhibition (Habold et al., 2008; Hirose et al., 2003; Motohashi et al., 2006; Roy et al., 2005; Sakata et al., 2004) and USF1, NFκB, MYC and ETS1 activities in serum response (Fujimori et al., 2008; Grimm et al., 1996; Kalra and Kumar, 2004; Petrovic et al., 2003; Tamura et al., 2005). 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.

8 DISCUSSION

8.1 Promoter probes preparation

8.1.1 Promoter design

Accurate prediction of TSSs and promoters is one of the most important problems and the biggest challenges in DNA sequence analysis, mainly due to the complexity of TSS distributions and our insufficient understanding of the transcription initiation process. Recent studies have revealed that the previous concept of a gene having one TSS was not true; instead, for a number of genes TSSs can be found spread over the gene loci, not only at the 5' end, but also downstream, sometimes in exons and introns (Carninci et al., 2005; Carninci et al., 2006). Moreover, previous studies suggested a number of human genes may have alternative promoters (APs) (Baek et al., 2007; Cooper et al., 2006). Besides, promoter regions are frequently shared or overlap each other, such as in sense/antisense genes and in bidirectionally promoted genes (Katayama et al., 2005). All these make the prediction of TSS and promoter a complicate and difficult task.

In this study, putative promoter regions were defined relative to the most upstream TSS for each gene, therefore only one promoter was designed for each gene. By using the most upstream TSS for promoter prediction, other TSSs, if exist, were ignored, which might lead to the loss of the dominant core promoters normally used for transcription initiation of the respective gene. However, TCA experiment in untreated HEK293T cells and cells with treatments of different external stimuli or depletion of serum showed that 61.2% (156/255) of the cloned Chr21 promoters drive reporter gene expression under at least one of the tested circumstances, indicating that our promoter selection method identified a large number of promoters that drive expression *in vivo*. Furthermore, fifteen promoters were found to have APs according to some pervious predictions (Baek et al., 2007; Kim et al., 2005b) or available databases (Eukaryotic Promoter Database), most of the which were also determined as active promoters in the TCA assay, suggesting that the our selection method was likely to be successful to identify the most functional promoters (Table 6). Three promoters with APs were determined as silent promoters in the TCA assay. For example, the cloned *WRB* promoter was not active in untreated HEK293 cells, but the *WRB* gene showed endogenous expression in HEK293T cells and board EST expression in 35 tissues, indicating the cloned fragment might not be the strongest one and the dominant core promoter was likely to be

missed. The *WRB* promoter showed activity after treatments of HEK293T cell with external stimuli, suggesting that the cloned fragments might be an alternative promoter of *WRB* gene. To check whether our cloned promoters represent the most dominant or strongest promoter, further annotation and analysis of the cloned promoter set could be perform on the basis of the CAGE and EST data.

Table 6. The list of genes that are known or predicted to have alternative promoters.

HGNC Symbol	Standard growth condition	Treated growth condition	HEK expression	References
ADAMTS1	1	1	1	(Kim et al., 2005b)
C21orf91	1	1	1	
DONSON	1	1	1	
DSCR2	1	1	1	
NDUFV3	1	1	1	
RCAN1 (DSCR1)	1	1	1	
SETD4	1	1	1	
SON	1	1	1	
STCH	1	1	1	
ATP5J	1	1	1	(Baek et al., 2007)
DTNBP1	1	1	1	
RCAN1	1	1	1	
TIAM1	1	1	1	
ERG	0	0	n/a	
CIITA	0	0	n/a	
WRB	0	1	1	Eukaryotic Promoter Database (EPD)

For functional characterization, promoter fragments with the length of 2500 bp were cloned. The fragments encompass the respective TSS of the gene and are located upstream of the TSS. Compared to previous studies where the length of promoters was restricted to 500-1000 bp (Cooper et al., 2006; Trinklein et al., 2007), the design of 2500 bp fragments aimed at more comprehensive coverage of the upstream regulatory sequences so that more potentially relevant regulatory elements could be covered by our analysis. There is no absolute definition for the length of a core promoter; however, CAGE tag clusters analyses suggested that most core promoters are smaller than 150 bp (Carninci et al., 2006; Sandelin et al., 2007). As highlighted previously (Cooper et al., 2006), the 300 bp sequence upstream of the TSS contributed positively to core promoter activity, and therefore seems to be a preferential localization of negative regulatory elements within the upstream 1000 to 500 bp region relative to the TSS of a promoter. This suggestion prompted us to design shorter promoter fragments and compare activities of truncated promoter fragments with their 2,500 bp

counterparts.

8.1.2 Promoter cloning

Cloning of larger number of promoters is a very time consuming process. The most critical steps in this process include the amplification the promoter fragments and the construction of promoter-reporter plasmids. In the work presented here, the promoter fragments were amplified from human genomic DNA. To decrease the rate of the primer-template mismatching that caused by the complex structure of genomic DNA, BAC and fosmid clones, if available, were used as the prior selection of templates for the amplification of promoter fragments. Moreover, touchdown PCR technique was applied to increase the specific binding of primers to template and avoid nonspecific amplification. For promoters which the expected amplicons were not obtained from PCR, the primers were re-designed and the PCR reactions were repeated with newly designed primers. In this study, primers for over 300 promoter fragments were designed, while eventually fragments of 255 promoters were successfully amplified.

For the construction of promoter-reporter plasmids, the Gateway recombinatorial cloning system (Invitrogen) was utilized and provided estimated success rates. The cloning efficiency of the Gateway system has been reported to associate with the size of fragments to be cloned into the vector, which, for the length of our cloned promoter fragments, was estimated to be about 50% (Marsischky and LaBaer, 2004).

8.2 Promoter activity analysis by TCA assay

For functional characterization of the putative promoters for the ability to induce transcription in human cells, the transfected-cell array (TCA) technique has been adapted. TCA allows analysis of multiple genes and gene products in parallel and has been used to screen phenotypes via gene silencing (Vanhecke and Janitz, 2005), to detect protein intracellular localization (Hu et al., 2006) and to investigate protein-protein interactions (Fiebitz et al., 2008). The present study showed another application of the TCA technique in the large-scale examination of putative promoter regions in living cells, which permitted the examination of promoter activities on a single microscopic slide in a statistically reliable manner.

8.2.1 Reverse transfection within HEK293T cells

The reverse transfection was performed within HEK293T cells. One of the most important advantages for choosing the HEK293T cells as the environment for promoter activity testing is the high reverse transfection rate that can be achieved when using this cell line. A previous study in our group has shown successful reverse transfection of NF-kB in HEK293/293T, Hela, WI-38, PC-3, COS 7 and HepG2 cells with comparable transfection efficiencies and fluorescent intensities in most cell lines (Fiebitz et al., 2008). However, in the present study, preliminary trials of reverse transfection performed within cell lines such as Hela and WI-38 using the same protocol applied for HEK293T showed little reporter fluorescence. One likely explanation for the undetectable level of fluorescent reporter protein in our study was the relatively low activity of tested promoters compared to the CMV promoter used in the previous study. Further optimization of transfection sample preparation protocol like increasing the DNA concentration of promoter-reporter constructs or utilizing other transfecction reagents or methods might be applied to multiply the expression of reporter proteins and thus increase the fluorescence signal.

8.2.2 Two methods for promoter activity evaluation

For promoter activity evaluation, two analysis methods based on either fluorescent intensity or numbers of transfected cells were applied and compared. The cell number parameter, rather than integrated fluorescent intensities, showed results that correlated better with visual observation and was used for future analyses. It was probably because the number of plasmid DNA entering the cells can differ considerably, which results in large variation of signal intensities among cells. It is also difficult to control the number of cells that are in direct contact with DNA-transfection reagent complexes. Nevertheless, the estimation of transfection efficiency in our assay was in line with previous reports involving the TCA technique (Baghdoyan et al., 2004) and proved to be sufficient for the chosen method of promoter activity quantification.

8.2.3 Promoter activities and modulation by treatment with external stimuli

The promoter activities could be modulated by treatment with small chemicals or by applying stress conditions. The expected effect of a pan-HDAC inhibitor (Khan et al., 2008), such as TSA, would be activation of transcription from the repressed regions of the chromosome. In

fact, a considerable number of promoters were observed to be activated upon treatment with PMA and TSA. Histone acetylation represents only one example of the mechanisms leading to gene repression (Lande-Diner and Cedar, 2005). Therefore, TSA treatment might result only in partial gene activation. On the other hand, CpG-enrichment in the sequence of TSA-up-regulated promoters coincides with the ubiquitous expression pattern of their genes. This observation might support previous findings that TSA-driven activation of transcription is partially dependent on DNA methylation (Januchowski et al., 2007; Klan et al., 2003; Lande-Diner et al., 2007).

Altogether, 61.2% (156/255) of the tested promoter fragments were active in either cell culture condition, i.e., without treatment, stimulated by chemicals or cultured in depleted media. The percentage was in line with the previous description that 60% of putative promoters were functional in at least one of the 16 cell types (Cooper et al., 2006). Within the 99 promoters that were silent in any condition, 64 of the corresponding genes were determined by transcriptome sequencing as not expressed in HEK293T cells. Moreover, Pol IIa binding sites were missing in most of these promoters; only four promoter fragments were found to contain the Pol IIa binding sites. Furthermore, 58.4% (45/77) of the corresponding genes were found to exhibit tissue-specific transcription with the expression in fewer than ten tissues. In all, the corresponding genes of promoters remained silent in any treatment condition showed less endogenous expression in HEK293T cells and *in vivo*.

8.3 Further functional analysis of promoters

8.3.1 Promoter activity versus endogenous transcript levels

Data showed that promoter reporter activities recapitulated endogenous gene expression to a great extent (87.1% concordance), which is in agreement with previous observations (Cooper et al., 2006). Nevertheless, transcripts for 13 genes were not detected, whose promoters were active on the cell arrays, including *AR, BIRC3, C21orf115, C21orf13, DSCR4, DSCR8, IFIH1, KRTAP21-2, NLRP2, NLRP4, RSPH1, TLR1* and *TLR5*. Seven of these genes (*AR, C21orf13, DSCR4, DSCR8, NLRP4, RSPH1* and *TLR5*) are expressed in only a few tissues according to EST data (Table 11). Messenger RNAs for genes *DSCR4, DSCR8, IFIH1, NLRP2* and *TLR1* are found at very low levels (Dunn et al., 2006), which might explain the patterns observed in our study.

The low endogenous mRNA levels of these genes might indicate that the presence of distal elements repressing the expression of those genes in HEK293T cells 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. Another possible reason for the discrepancy between active promoters and the low level of transcripts is mRNA with low abundance and high turnover. 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, deletion analysis or transfection into non-adherent cells because these cells cannot be applied to reverse transfection on cell arrays.

Conversely, among 184 genes that are found to be expressed in HEK293T cells, 41 of the corresponding promoters remained silent in the TCA assay. In fact, 28 of these promoters were activated by external stimuli treatments, suggesting that the corresponding cloned fragments contained a functional core promoter but lacked enhancers or binding sites for factors modulating the chromatin structure. The remaining 13 inactive promoters might be missing the core promoter required for transcription, either due to inadequate TSS annotation or use of an alternative TSS in HEK293T cells. Besides the reasons mentioned above, experimental noise in promoter activity measurements and mRNA sequencing might lead to an underestimation of results and thus contribute to the observed differences.

8.3.2 Mapping of Pol IIa sites

Mapping of Pol IIa sites indicate sites of active transcription or pre-initiation complex formation, with binding sites localizing near the transcription start sites. As expected, the presence of Pol IIa binding sites was significantly associated with the active promoter fragments (Figure 15; Table 11). Of the fragments that were found to contain Pol IIa binding sites, 73.1% (38 of 52) were determined as active promoters in the TCA assay. The strong overlap between the positive results of promoter activity and Pol IIa binding sites indicated that both approaches could independently identify many of the same functional promoters. The promoter fragments that were bound by Pol IIa but were not functionally active in TCA assay could represent sites where the preinitiation complex was assembled but paused and not transcriptionally active (Krumm et al., 1995; Krumm et al., 1992).

8.3.3 Promoter sequence elements and tissue expression of corresponding genes

The correlation of promoter activities with various sequence features involved in promoter initiation complex assembly, such as CpG islands, TATA box, INR and DPE elements, was examined. The finding of enrichment in CpG and DPE but not TATA or INR elements in active promoters confirms previous observations (Cooper et al., 2006; Kim et al., 2005b). 83.1% of the active promoters in HEK293T cells contain GpC islands, which is in agreement with the previous observation that CpG islands are present within more than 80% of the promoters active in other cell types, such as primary fibroblasts (Kim et al., 2005b).

Moreover, promoter sequence motifs were compared to the tissue-specific expression of the corresponding genes. Among the 83 genes that were expressed in more than 25 tissues, 68 promoters were found to contain CpG islands. In contrast, of the 86 genes that were expressed in fewer than 10 tissues, only 15 had CpG islands. These observations confirm the previous description that CpG islands are commonly associated with ubiquitously expressed genes (Schug et al., 2005).

8.3.4 Cis-regulatory elements within distal promoter regions

The activities of 62 truncated promoter fragments were tested and compared with their 2.5 kb counterparts. The comparison 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 (Cooper et al., 2007; Leicht et al., 2001; Leicht et al., 2003). 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 (Fujimori et al., 2008). NFκB has been found potently activated upon serum starvation in HEK293 cells, leading to apoptosis (Grimm et al., 1996). Ets domain-containing transcription factors, such as ETS1 identified here, are implicated in the response to serum in endothelial cells (Petrovic et al., 2003). Finally, MYC has been previously implied in responses to growth factor-deprived conditions, where it is involved in induction of apoptosis (Kalra and Kumar, 2004).

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 AP1) 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 (Motohashi et al., 2006), leading to gene activation. TSA can induce p53-mediated cell cycle arrest or apoptosis, depending on the cell type (Habold et al., 2008; Roy et al., 2005). TSA can also induce gene expression via OCT1, independently of p53 (Hirose et al., 2003). In addition, TSA can promote the binding of AP1 to a recognition site and activate the expression of the osteoporin gene in a mouse mesenchymal cell line (Sakata et al., 2004). 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.

8.4 Advantages and limitations of TCA

This study demonstrates that the transfected cell array technique, together with gene expression and bioinformatics tools, represents a robust and powerful technology for large-scale functional studies of regulatory sequences.

8.4.1 Advantages of TCA

The TCA technique offers a robust platform for high-throughput functional studies of genes and proteins. Different samples are printed at a high density on a glass slide. A density of hundreds to thousands spots per standard microscope slides can be achieved depending on the diameter of the DNA spots and the distance between the neighboring spots. Theoretically, up to about 7000 spots with the diameter and distance of 150 μm can be printed on one slide by using a robotic arrayer. Combined with appropriate detection assays, the phenotypic effects of

thousands of cell clusters can be monitored in parallel. One of the most distinct advantages of TCA is to provide an opportunity to study gene function in the context of living cells on a large scale. Gene expression in mammalian cells undergoes a series of complicated processes and requires interaction with networks, in vitro investigation of human genes in prokaryotes, yeast and other model organisms may not show the properties of the genes or their products. The use of human cells provides intact *in vivo* processing networks for functional study of human genes. Owing to the advances in automated microscopy, TCAs can be used to monitor live cells in culture, allowing the discovery of transient or time-dependant phenotypes, such as programmed cell death and the intracellular movement of a GFP-tagged protein.

TCA has several advantages over the conventional functional assays performed on mammalian cells in microwell-plate format. A prominent feature of the cell array technique is that less material is required for transfection. Because the samples are printed at high density on the slide, the consumption of DNA/RNA as well as transfection reagents can be substantially reduced for each gene tested in the TCA approach. This makes TCA, at the moment, the most cost-effective tool for genome-wide gene functional analyses (Janitz et al., 2006). The TCA technique also requires fewer cells. This is of particular importance for some cell lines that are difficult to culture. One leading advantage of cell arrays is the increased comparability among different genes. Because the transfection and analysis processes for all the genes are carried out in the same way, the signals from different samples share the same background, which facilitates the identification of signal from each sample and comparison among different samples. Once the cell arrays have been prepared, they can be applied to different types of cells.

8.4.2 Limitations of TCA

However, several important limitations of cell arrays are intrinsic to the process of transfection.

(1) The application of TCA is limited to a certain fraction of cell types. Only adhesive cells, which can form a monolayer on the surface of the cell array, can be used for current cell array technology, whereas the suspended cells are not compatible with this technique. To overcome this limit, the study of Kato et al. seems to be a promising approach (Kato et al.). The authors designed a biocompatible anchor for cell membrane components. The slides fabricated with

such anchors were amenable to the attachment of non-adherent human erythroleukemic K562 cells. If this could also be applied and optimised for TCA protocol, potentially the suspended cell lines could be used for reverse transfection.

(2) Moreover, not all the adhesive cell lines but only cells with high transfection efficiency are compatible with the current lipid-based transfection methods. To perform TCA assay in the cells with lower transfection efficiency, increasing the DNA concentration of promoter-reporter constructs during the preparation of transfection sample might multiply the amount of reporter proteins that enter cells and thus increase the fluorescence signal. However, too much DNA might be toxic to cells. In the previous study of Fiebitz et al., it was found that reverse transfection with total DNA concentration of approximately 50 ng/µl was found to give the best fluorescent signals; however, the signals dropped off sharply when the DNA concentration was over 60 ng/µl, indicating that further increasing the reporter concentration did not improve the quality of the fluorescent signal.

Besides, cells that are difficult to be transfected with lipid-based transfection methods, such as most primary cell lines, were barely applicable to cell arrays. Hopefully, this technical limitation can be conquered by the development of new biological strategies. For example, Sabatini's group has described a new type of cell array, with the combination of lentivirus with TCA technology, referred to as lentivirus-infected cell microarray (LICM), for high-throughput screening in a broad range of mammalian cells including immortalized and primary cells (Bailey et al., 2006). This approach was important in expanding the application of cell arrays to cell types that were thought to be untransfectable in the past.

(3) Cell arrays are not applicable for screening of cellular phenotypes with a long period over 72 h. The best results for reverse transfection experiments with HEK293T cells were found with time period of 48-65 hrs. The fluorescent signals were still very weak 24 hrs after seeding, while incubation for 3 days resulted in the death of many cells and thus in the destruction of the monolayer on the slide. Decreasing the number of seeding cells and refreshing the media every day could prolong the incubation period of cell arrays; however, it could not prevent cell death and monolayer detachment. Therefore, the detection of cellular phenotypes during long periods remains challenging with TCA technology.

(4) Transfection of promoter-reporter constructs in certain cell lines might not always

represent the full activity of the tested promoter in vivo, especially for the genes with tissue-restricted expression patterns. This is the limitation not only for TCA but also for normal transfection assays. The cell lines might lack some crucial factors required for the transcription initiation, thus the transcription complex could not be formed at the promoter region, and transcription could not initiate. There is also another possibility that alternative transcription factors exist in the experimental cells, resulting in the activation or suppression of promoters and the discrepancy between promoter activity and the corresponding gene expression.

To evaluate the promoter activity of tissue-restricted expression genes, specific cell lines could be applied to the TCA approach. For that purpose, the transfection protocol has to be optimized for each new cell line. In case either no appropriate cells are available or the specific cell lines are not applicable to cell array technology, co-transfection of promoter constructs with plasmids expressing a certain transcription factor could be a potential alternative. The co-transfection could be used to determine the stimulation or suppression of promoters by any transcription factor.

8.5 Outlook

TCA technology has been shown with a variety of applications, including the study of protein localization or protein-protein interaction by overexpressing genes (Fiebitz et al., 2008; Hu et al., 2006), the gene loss-of-function study by knocking down genes in combination with RNAi technology (Erfle et al., 2004; Kumar et al., 2003; Mousses et al., 2003; Silva et al., 2004; Vanhecke and Janitz, 2004; Yoshikawa et al., 2004), and the functional study of transcription promoters by expressing a reporter gene constructed in a plasmid downstream of the promoter sequence. For functional studies, it is also possible, instead of transfected cells with only one kind of DNA or RNA, to co-transfect different DNAs, or DNA with RNA into cells. For instance, the promoter construct could be co-transfected with plasmid expressing certain transcription factor, as mentioned above, to determine the regulatory effect of the transcription factor on the promoter. Besides, although current implementations of cell arrays use mostly printed DNA or RNA, it might be possible to create other classes of cell microarrays by printing substances such as small chemical compounds, oligonucleotides, or proteins. Small chemicals can also be co-transfected with nucleic acids into cells, which would be a useful approach to study the protein-drug interaction in large-scale manner.

Furthermore, cell microarrays are not limited to microscope slides; arrays can be formed on any solid surface that supports the attachment of mammalian cells.

Taken together, cell arrays are a potentially powerful high-throughput tool for the post-genomic era. In combination with other molecular and cellular technique and efficient functional detection assays, TCAs will find broad utility in biological research.

9 BIBLIOGRAPHY

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., *et al.* (2000). The genome sequence of Drosophila melanogaster. Science (New York, NY *287*, 2185-2195.

Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R.J., Rahmsdorf, H.J., Jonat, C., Herrlich, P., and Karin, M. (1987). Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor. Cell *49*, 729-739.

Baek, D., Davis, C., Ewing, B., Gordon, D., and Green, P. (2007). Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. Genome research *17*, 145-155.

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 research *32*, e77.

Bailey, S.N., Ali, S.M., Carpenter, A.E., Higgins, C.O., and Sabatini, D.M. (2006). Microarrays of lentiviruses for gene function screens in immortalized and primary cells. Nature methods *3*, 117-122.

Bailey, S.N., Wu, R.Z., and Sabatini, D.M. (2002). Applications of transfected cell microarrays in high-throughput drug discovery. Drug discovery today *7*, S113-118.

Bednarz, J., Teifel, M., Friedl, P., and Engelmann, K. (2000). Immortalization of human corneal endothelial cells using electroporation protocol optimized for human corneal endothelial and human retinal pigment epithelial cells. Acta ophthalmologica Scandinavica *78*, 130-136.

Blackwood, E.M., and Kadonaga, J.T. (1998). Going the distance: a current view of enhancer action. Science (New York, NY *281*, 60-63.

Blumberg, P.M. (1988). Protein kinase C as the receptor for the phorbol ester tumor

promoters: sixth Rhoads memorial award lecture. Cancer research *48*, 1-8.

Bock, C., Walter, J., Paulsen, M., and Lengauer, T. (2007). CpG island mapping by epigenome prediction. PLoS computational biology *3*, e110.

Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Annual review of biochemistry *50*, 349-383.

Brodsky, A.S., Meyer, C.A., Swinburne, I.A., Hall, G., Keenan, B.J., Liu, X.S., Fox, E.A., and Silver, P.A. (2005). Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells. Genome biology *6*, R64.

Bulger, M., and Groudine, M. (1999). Looping versus linking: toward a model for long-distance gene activation. Genes & development *13*, 2465-2477.

Burke, T.W., and Kadonaga, J.T. (1996). Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes & development *10*, 711-724.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., *et al.* (2005). The transcriptional landscape of the mammalian genome. Science (New York, NY *309*, 1559-1563.

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. Nature genetics *38*, 626-635.

Consortium., C.e.S. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. Science (New York, NY *282*, 2012-2018.

Consortium., I.H.G.S. (2004). Finishing the euchromatic sequence of the human genome. Nature *431*, 931-945.

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 research *16*, 1-10.

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 research *17*, 136-144.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., and Chambon, P. (1980). Promoter sequences of eukaryotic protein-coding genes. Science (New York, NY *209*, 1406-1414.

D'Angelo, D.D., Oliver, B.G., Davis, M.G., McCluskey, T.S., and Dorn, G.W. (1996). Novel role for Sp1 in phorbol ester enhancement of human platelet thromboxane receptor gene expression. The Journal of biological chemistry *271*, 19696-19704.

Deng, W., and Roberts, S.G. (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB. Genes & development *19*, 2418-2423.

Deng, W., and Roberts, S.G. (2007). TFIIB and the regulation of transcription by RNA polymerase II. Chromosoma *116*, 417-429.

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 biology *4*, P3.

Dunn, C.A., Romanish, M.T., Gutierrez, L.E., van de Lagemaat, L.N., and Mager, D.L. (2006). Transcription of two human genes from a bidirectional endogenous retrovirus promoter. Gene *366*, 335-342.

Erfle, H., Simpson, J.C., Bastiaens, P.I., and Pepperkok, R. (2004). siRNA cell arrays for high-content screening microscopy. BioTechniques *37*, 454-458, 460, 462.

Fatemi, M., Pao, M.M., Jeong, S., Gal-Yam, E.N., Egger, G., Weisenberger, D.J., and Jones, P.A. (2005). Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. Nucleic acids research *33*, e176.

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.

FitzGerald, P.C., Sturgill, D., Shyakhtenko, A., Oliver, B., and Vinson, C. (2006). Comparative genomics of Drosophila and human core promoters. Genome biology *7*, R53.

Freiman, R.N. (2009). Specific variants of general transcription factors regulate germ cell development in diverse organisms. Biochimica et biophysica acta *1789*, 161-166.

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.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. Journal of molecular biology *196*, 261-282.

Gershenzon, N.I., and Ioshikhes, I.P. (2005). Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. Bioinformatics (Oxford, England) *21*, 1295-1300.

Gershenzon, N.I., Trifonov, E.N., and Ioshikhes, I.P. (2006). The features of Drosophila core promoters revealed by statistical analysis. BMC genomics *7*, 161.

Graham, F.L., Smiley, J., Russell, W.C., and Nairn, R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. The Journal of general virology *36*, 59-74.

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. The Journal of cell biology *134*, 13-23.

Gurskaya, N.G., Fradkov, A.F., Terskikh, A., Matz, M.V., Labas, Y.A., Martynov, V.I., Yanushevich, Y.G., Lukyanov, K.A., and Lukyanov, S.A. (2001). GFP-like chromoproteins

as a source of far-red fluorescent proteins. FEBS letters *507*, 16-20.

Gustincich, S., Sandelin, A., Plessy, C., Katayama, S., Simone, R., Lazarevic, D., Hayashizaki, Y., and Carninci, P. (2006). The complexity of the mammalian transcriptome. The Journal of physiology *575*, 321-332.

Haas, J., Park, E.C., and Seed, B. (1996). Codon usage limitation in the expression of HIV-1 envelope glycoprotein. Curr Biol *6*, 315-324.

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 research *31*, 5576-5581.

Habold, 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. Journal of cellular and molecular medicine *12*, 607-621.

Hackenberg, M., Previti, C., Luque-Escamilla, P.L., Carpena, P., Martinez-Aroza, J., and Oliver, J.L. (2006). CpGcluster: a distance-based algorithm for CpG-island detection. BMC bioinformatics *7*, 446.

Harada, J.N., Bower, K.E., Orth, A.P., Callaway, S., Nelson, C.G., Laris, C., Hogenesch, J.B., Vogt, P.K., and Chanda, S.K. (2005). Identification of novel mammalian growth regulatory factors by genome-scale quantitative image analysis. Genome research *15*, 1136-1144.

He, X.Q., Chen, R., Yang, P., Li, A.P., Zhou, J.W., and Liu, Q.Z. (2007). Biphasic effect of arsenite on cell proliferation and apoptosis is associated with the activation of JNK and ERK1/2 in human embryo lung fibroblast cells. Toxicology and applied pharmacology *220*, 18-24.

Herman, J.G., and Baylin, S.B. (2003). Gene silencing in cancer in association with promoter hypermethylation. The New England journal of medicine *349*, 2042-2054.

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.

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.

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. The Journal of biological chemistry *280*, 10047-10054.

Hultmark, D., Klemenz, R., and Gehring, W.J. (1986). Translational and transcriptional control elements in the untranslated leader of the heat-shock gene hsp22. Cell *44*, 429-438.

Janitz, M., Vanhecke, D., and Lehrach, H. (2006). High-throughput RNA interference in functional genomics. Handbook of experimental pharmacology, 97-104.

Januchowski, R., Dabrowski, M., Ofori, H., and Jagodzinski, P.P. (2007). Trichostatin A down-regulate DNA methyltransferase 1 in Jurkat T cells. Cancer letters *246*, 313-317.

Kahn, C.R., Young, E., Lee, I.H., and Rhim, J.S. (1993). Human corneal epithelial primary cultures and cell lines with extended life span: in vitro model for ocular studies. Investigative ophthalmology & visual science *34*, 3429-3441.

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. The Journal of biological chemistry *279*, 25313-25319.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., *et al.* (2005). Antisense transcription in the mammalian transcriptome. Science (New York, NY *309*, 1564-1566.

Kato, K., Umezawa, K., Miyake, M., Miyake, J., and Nagamune, T. (2004). Transfection microarray of nonadherent cells on an oleyl poly(ethylene glycol) ether-modified glass slide. BioTechniques *37*, 444-448, 450, 452.

Keen, J.C., Cianferoni, A., Florio, G., Guo, J., Chen, R., Roman, J., Wills-Karp, M., Casolaro, V., and Georas, S.N. (2006). Characterization of a novel PMA-inducible pathway of interleukin-13 gene expression in T cells. Immunology *117*, 29-37.

Kel, A.E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V., and Wingender, E. (2003). MATCH: A tool for searching transcription factor binding sites in DNA sequences. Nucleic acids research *31*, 3576-3579.

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. The Biochemical journal *409*, 581-589.

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.* (2005a). Direct isolation and identification of promoters in the human genome. Genome research *15*, 830-839.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005b). A high-resolution map of active promoters in the human genome. Nature *436*, 876-880.

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. Biological chemistry *384*, 777-785.

Kornberg, R.D. (1999). Eukaryotic transcriptional control. Trends in cell biology *9*, M46-49.

Krumm, A., Hickey, L.B., and Groudine, M. (1995). Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation. Genes & development *9*, 559-572.

Krumm, A., Meulia, T., Brunvand, M., and Groudine, M. (1992). The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region. Genes & development *6*, 2201-2213.

Kumar, R., Conklin, D.S., and Mittal, V. (2003). High-throughput selection of effective RNAi probes for gene silencing. Genome research *13*, 2333-2340.

Kutach, A.K., and Kadonaga, J.T. (2000). The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Molecular and cellular biology *20*, 4754-4764.

Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes & development *12*, 34-44.

Lande-Diner, L., and Cedar, H. (2005). Silence of the genes--mechanisms of long-term repression. Nat Rev Genet *6*, 648-654.

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. The Journal of biological chemistry *282*, 12194-12200.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Landry, J.R., Mager, D.L., and Wilhelm, B.T. (2003). Complex controls: the role of alternative promoters in mammalian genomes. Trends Genet *19*, 640-648.

Le Gall, M., Chambard, J.C., Breittmayer, J.P., Grall, D., Pouyssegur, J., and Van Obberghen-Schilling, E. (2000). The p42/p44 MAP kinase pathway prevents apoptosis induced by anchorage and serum removal. Molecular biology of the cell *11*, 1103-1112.

Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annu

Rev Genet *34*, 77-137.

Leicht, M., Briest, W., Holzl, 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. Cardiovascular research *52*, 429-437.

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.

Li, C.Y., Fang, F., Xu, W.X., Xu, C.W., Zhan, Y.Q., Wang, Z.D., Ding, Y.L., Li, Y.H., Sun, H.B., and Yang, X.M. (2008). Suppression of EDAG gene expression by phorbol 12-myristate 13-acetate is mediated through down-regulation of GATA-1. Biochimica et biophysica acta *1779*, 606-615.

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 and differentiation *13*, 2109-2117.

Liebel, U., Starkuviene, V., Erfle, H., Simpson, J.C., Poustka, A., Wiemann, S., and Pepperkok, R. (2003). A microscope-based screening platform for large-scale functional protein analysis in intact cells. FEBS letters *554*, 394-398.

Luciakova, K., Hodny, Z., Barath, P., and Nelson, B.D. (2000). In vivo mapping of the human adenine nucleotide translocator-2 (ANT2) promoter provides support for regulation by a pair of proximal Sp1-activating sites and an upstream silencer element. The Biochemical journal *352 Pt 2*, 519-523.

Luo, R.X., and Dean, D.C. (1999). Chromatin remodeling and transcriptional regulation. Journal of the National Cancer Institute *91*, 1288-1294.

Malakooti, J., Sandoval, R., Amin, M.R., Clark, J., Dudeja, P.K., and Ramaswamy, K. (2006). Transcriptional stimulation of the human NHE3 promoter activity by PMA: PKC independence and involvement of the transcription factor EGR-1. The Biochemical journal

396, 327-336.

Malakooti, J., Sandoval, R., Memark, V.C., Dudeja, P.K., and Ramaswamy, K. (2005). Zinc finger transcription factor Egr-1 is involved in stimulation of NHE2 gene expression by phorbol 12-myristate 13-acetate. American journal of physiology *289*, G653-663.

Manke, T., Roider, H.G., and Vingron, M. (2008). Statistical modeling of transcription factor binding affinities predicts regulatory interactions. PLoS computational biology *4*, e1000039.

Marks, P.A., Miller, T., and Richon, V.M. (2003). Histone deacetylases. Curr Opin Pharmacol *3*, 344-351.

Marsischky, G., and LaBaer, J. (2004). Many paths to many clones: a comparative look at high-throughput cloning methods. Genome research *14*, 2020-2028.

Miller, W., Makova, K.D., Nekrutenko, A., and Hardison, R.C. (2004). Comparative genomics. Annual review of genomics and human genetics *5*, 15-56.

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. Molecular and cellular biology *26*, 4652-4663.

Mousses, S., Caplen, N.J., Cornelison, R., Weaver, D., Basik, M., Hautaniemi, S., Elkahloun, A.G., Lotufo, R.A., Choudary, A., Dougherty, E.R., *et al.* (2003). RNAi microarray analysis in cultured mammalian cells. Genome research *13*, 2341-2347.

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. The Journal of biological chemistry *282*, 14685-14689.

Murray, J.I., Whitfield, M.L., Trinklein, N.D., Myers, R.M., Brown, P.O., and Botstein, D. (2004). Diverse and specific gene expression responses to stresses in cultured human cells. Molecular biology of the cell *15*, 2361-2374.

O'Brien, V., Frisch, S.M., and Juliano, R.L. (1996). Expression of the integrin alpha 5 subunit in HT29 colon carcinoma cells suppresses apoptosis triggered by serum deprivation. Experimental cell research *224*, 208-213.

O'Shea-Greenfield, A., and Smale, S.T. (1992). Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. The Journal of biological chemistry *267*, 6450.

Ohler, U., Liao, G.C., Niemann, H., and Rubin, G.M. (2002). Computational analysis of core promoters in the Drosophila genome. Genome biology *3*, RESEARCH0087.

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. Genes & development *10*, 2657-2683.

Pazin, M.J., and Kadonaga, J.T. (1997). What's up and down with histone deacetylation and transcription? Cell *89*, 325-328.

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. The Journal of biological chemistry *278*, 49358-49368.

Price, J.H., Goodacre, A., Hahn, K., Hodgson, L., Hunter, E.A., Krajewski, S., Murphy, R.F., Rabinovich, A., Reed, J.C., and Heynen, S. (2002). Advances in molecular labeling, high throughput imaging and machine intelligence portend powerful functional cellular biochemistry tools. Journal of cellular biochemistry *39*, 194-210.

Purnell, B.A., Emanuel, P.A., and Gilmour, D.S. (1994). TFIID sequence recognition of the initiator and sequences farther downstream in Drosophila class II genes. Genes & development *8*, 830-842.

Ralser, M., Querfurth, R., Warnatz, H.J., Lehrach, H., Yaspo, M.L., and Krobitsch, S. (2006). An efficient and economic enhancer mix for PCR. Biochemical and biophysical research communications *347*, 747-751.

Roth, S.Y., Denu, J.M., and Allis, C.D. (2001). Histone acetyltransferases. Annual review of biochemistry *70*, 81-120.

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 and differentiation *12*, 482-491.

Sakata, R., Minami, S., Sowa, Y., Yoshida, M., and Tamaki, T. (2004). Trichostatin A activates the osteopontin gene promoter through AP1 site. Biochemical and biophysical research communications *315*, 959-963.

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.

Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proceedings of the National Academy of Sciences of the United States of America *103*, 1412-1417.

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

Silva, J.M., Mizuno, H., Brady, A., Lucito, R., and Hannon, G.J. (2004). RNA interference microarrays: high-throughput loss-of-function genetics in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America *101*, 6548-6552.

Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annual review of biochemistry *72*, 449-479.

Smith, C.L., and Hager, G.L. (1997). Transcriptional regulation of mammalian genes in vivo. A tale of two templates. The Journal of biological chemistry *272*, 27493-27496.

Sterner, D.E., and Berger, S.L. (2000). Acetylation of histones and transcription-related factors. Microbiol Mol Biol Rev *64*, 435-459.

Struhl, K. (1987). Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. Cell *49*, 295-297.

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 (New York, NY *321*, 956-960.

Suzuki, Y., Taira, H., Tsunoda, T., Mizushima-Sugano, J., Sese, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Morishita, S., *et al.* (2001). Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites. EMBO reports *2*, 388-393.

Suzuki, Y., Yamashita, R., Nakai, K., and Sugano, S. (2002). DBTSS: DataBase of human Transcriptional Start Sites and full-length cDNAs. Nucleic acids research *30*, 328-331.

Takai, D., and Jones, P.A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proceedings of the National Academy of Sciences of the United States of America *99*, 3740-3745.

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.

Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol *41*, 105-178.

Trinklein, N.D., Aldred, S.J., Saldanha, A.J., and Myers, R.M. (2003). Identification and functional analysis of human transcriptional promoters. Genome research *13*, 308-312.

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 research *17*, 720-731.

Vanhecke, D., and Janitz, M. (2004). High-throughput gene silencing using cell arrays. Oncogene *23*, 8353-8358.

Vanhecke, D., and Janitz, M. (2005). Functional genomics using high-throughput RNA interference. Drug discovery today *10*, 205-212.

Wakaguri, H., Yamashita, R., Suzuki, Y., Sugano, S., and Nakai, K. (2008). DBTSS: database of transcription start sites, progress report 2008. Nucleic acids research *36*, D97-101.

Wang, Y., and Leung, F.C. (2004). An evaluation of new criteria for CpG islands in the human genome as gene markers. Bioinformatics (Oxford, England) *20*, 1170-1177.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., *et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

West, A.G., Gaszner, M., and Felsenfeld, G. (2002). Insulators: many functions, many mechanisms. Genes & development *16*, 271-288.

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 research *29*, 281-283.

Wolf, D.E. (2003). Quantitative digital and video microscopy. Methods in cell biology *72*, 319-336.

Yih, L.H., Peck, K., and Lee, T.C. (2002). Changes in gene expression profiles of human fibroblasts in response to sodium arsenite treatment. Carcinogenesis *23*, 867-876.

Yokota, T., Matsuzaki, Y., Miyazawa, K., Zindy, F., Roussel, M.F., and Sakai, T. (2004). Histone deacetylase inhibitors activate INK4d gene through Sp1 site in its promoter.

Oncogene *23*, 5340-5349.

Yoshikawa, T., Uchimura, E., Kishi, M., Funeriu, D.P., Miyake, M., and Miyake, J. (2004). Transfection microarray of human mesenchymal stem cells and on-chip siRNA gene knockdown. J Control Release *96*, 227-232.

Zemanova, L., Schenk, A., Valler, M.J., Nienhaus, G.U., and Heilker, R. (2003). Confocal optics microscopy for biochemical and cellular high-throughput screening. Drug discovery today *8*, 1085-1093.

Ziauddin, J., and Sabatini, D.M. (2001). Microarrays of cells expressing defined cDNAs. Nature *411*, 107-110.

10 APPENDIX

10.1 List of cloned promoters

Table 7. Information of 255 genes and the coordinated promoters.

HGNC Symbol	Entrez GeneID	Ensembl GeneID	Strand	Coordinates Promoter (hg18)
ABCC13	150000	ENSG00000155288	+	chr21:14565550-14568026
ABCG1	9619	ENSG00000160179	+	chr21:42506845-42509332
ADAMTS1	9510	ENSG00000154734	-	chr21:27139564-27142048
ADAMTS5	11096	ENSG00000154736	-	chr21:27260677-27263142
AIRE	326	ENSG00000160224	+	chr21:44527749-44530171
AR	367	ENSG00000169083	+	chrX:66545580-66548048
ARC	23237	ENSG00000198576	-	chr8:143692792-143695282
ATP5J	522	ENSG00000154723	-	chr21:26029811-26032298
ATP5O	539	ENSG00000159186	-	chr21:34210007-34212471
B3GALT5	10317	ENSG00000183778	+	chr21:39951048-39953597
BACE2	25825	ENSG00000182240	+	chr21:41459151-41461640
BACH1	571	ENSG00000156273	+	chr21:29590654-29593127
BCL10	8915	ENSG00000142867	-	chr1:85454203-85456676
BIRC2	329	ENSG00000110330	+	chr11:101720681-101723170
BIRC3	330	ENSG00000023445	+	chr11:101690983-101693445
BIRC5	332	ENSG00000089685	+	chr17:73719507-73721971
BTG3	10950	ENSG00000154640	-	chr21:17907049-17909482
C21orf100	118421	ENSG00000178446	+	chr21:28831026-28833507
C21orf109	193629	ENSG00000198966	+	chr21:29485857-29488385
C21orf110	378822	ENSG00000175302	-	chr21:13682137-13684629
C21orf114	378826	ENSG00000196736	+	chr21:17997554-18000048
C21orf115	378827	ENSG00000198308	-	chr21:41480091-41482576
C21orf116	378820	ENSG00000197510	+	chr21:15341962-15344429
C21orf119	84996	ENSG00000186837	+	chr21:32684864-32687345
C21orf120	56245	ENSG00000198655	-	chr21:33085192-33087677
C21orf121	150142	ENSG00000182852	+	chr21:42299928-42302404
C21orf123	378832	ENSG00000183535	-	chr21:45669371-45671908
C21orf125	284836	ENSG00000188660	+	chr21:43691912-43694347
C21orf128	150147	ENSG00000184385	-	chr21:42401581-42404068
C21orf129	150135	ENSG00000185481	-	chr21:42008964-42011443
C21orf13	150082	ENSG00000157578	-	chr21:39739488-39741977
C21orf15	54094	ENSG00000198401	-	chr21:14138261-14140738
C21orf19	54092	ENSG00000179408	+	chr21:36421720-36424169
C21orf2	755	ENSG00000160226	-	chr21:44583683-44586134
C21orf25	25966	ENSG00000157617	-	chr21:42247178-42249685
C21orf29	54084	ENSG00000175894	-	chr21:44955928-44958415
C21orf30	54083	ENSG00000183637	+	chr21:44701759-44704224
C21orf32	54081	ENSG00000185225	-	chr21:44418039-44420487
C21orf33	8209	ENSG00000160221	+	chr21:44375499-44377959
C21orf34	388815	ENSG00000174496	+	chr21:16362927-16365422
C21orf41	54073	ENSG00000183653	+	chr21:29888048-29890527
C21orf42	54072	ENSG00000185433	-	chr21:25725452-25727974
C21orf45	54069	ENSG00000159055	-	chr21:32573285-32575737
C21orf55	54943	ENSG00000177692	-	chr21:33785722-33788091

C21orf56	84221	ENSG00000160284	-	chr21:46428691-46431162
C21orf57	54059	ENSG00000182362	+	chr21:46528277-46530768
C21orf58	54058	ENSG00000160298	-	chr21:46568194-46570659
C21orf62	56245	ENSG00000159100	-	chr21:33107818-33110311
C21orf66	94104	ENSG00000159086	-	chr21:33066005-33068482
C21orf67	84536	ENSG00000183250	-	chr21:45184211-45186704
C21orf70	85395	ENSG00000160256	+	chr21:45181880-45184359
C21orf74	54143	ENSG00000184856	+	chr21:22401164-22403690
C21orf77	55264	ENSG00000186842	-	chr21:32870050-32872515
C21orf82	114036	ENSG00000198773	+	chr21:34474400-34476851
C21orf84	114038	ENSG00000185186	-	chr21:43722444-43724935
C21orf86	257103	ENSG00000184836	-	chr21:45538991-45541467
C21orf87	257357	ENSG00000183582	-	chr21:39608216-39610680
C21orf88	114041	ENSG00000184809	-	chr21:39906588-39909123
C21orf9	642976	ENSG00000174680	+	chr21:30040637-30043126
C21orf90	114043	ENSG00000182912	+	chr21:44759035-44761516
C21orf91	54149	ENSG00000154642	-	chr21:18113524-18116045
C21orf93	246704	ENSG00000184274	-	chr21:45549612-45552014
C21orf94	246705	ENSG00000178457	+	chr21:28305075-28307554
C21orf99	149992	ENSG00000155258	+	chr21:13329918-13332393
CARD11	84433	ENSG00000198286	-	chr7:2856741-2859148
CARD14	79092	ENSG00000141527	+	chr17:75773748-75776241
CARD6	84674	ENSG00000132357	+	chr5:40874600-40877064
CARD8	22900	ENSG00000105483	-	chr19:53450920-53453348
CASP1	834	ENSG00000137752	-	chr11:104477407-104479839
CBR1	873	ENSG00000159228	+	chr21:36361659-36364175
CBR3	874	ENSG00000159231	+	chr21:36426826-36429321
CBS	875	ENSG00000160200	-	chr21:43369513-43371989
CCT8	10694	ENSG00000156261	-	chr21:29367838-29370329
CHAF1B	8208	ENSG00000159259	+	chr21:36677130-36679599
CHODL	140578	ENSG00000154645	+	chr21:18536580-18539050
CIITA	4261	ENSG00000179583	+	chr16:10876070-10878543
CLDN14	23562	ENSG00000159261	-	chr21:36870702-36873185
CLDN17	26285	ENSG00000156282	-	chr21:30460811-30463296
CLDN8	9073	ENSG00000156284	-	chr21:30510182-30512658
CLIC6	54102	ENSG00000159212	+	chr21:34961076-34963551
COL6A2	1292	ENSG00000142173	+	chr21:46339940-46342468
CRADD	8738	ENSG00000169372	+	chr12:92571127-92573618
CRYAA	1409	ENSG00000160202	+	chr21:43459831-43462251
CXADR	1525	ENSG00000154639	+	chr21:17804755-17807171
CYYR1	116159	ENSG00000166265	-	chr21:26867490-26869951
DAOA	267012	ENSG00000182346	+	chr13:104920373-104922863
DEFA4	1669	ENSG00000164821	-	chr8:6783151-6785515
DEFA5	1670	ENSG00000164816	-	chr8:6901677-6904157
DEFA6	1671	ENSG00000164822	-	chr8:6771021-6773503
DEFB1	1672	ENSG00000164825	-	chr8:6722884-6725365
DEFB103A	55894	ENSG00000177243	-	chr8:7275234-7277714
DEFB104A	140596	ENSG00000176782	+	chr8:7728914-7731396
DEFB105A	245908	ENSG00000186562	-	chr8:7717605-7720101
DEFB106A	245909	ENSG00000186579	+	chr8:7718772-7721250
DEFB118	117285	ENSG00000131068	+	chr20:29417588-29420060
DEFB119	245932	ENSG00000180483	-	chr20:29442019-29444510
DEFB123	245936	ENSG00000180424	+	chr20:29489705-29492120
DEFB125	245938	ENSG00000178591	+	chr20:13872-16354

DEFB126	81623	ENSG00000125788	+	chr20:68813-71297
DEFB127	140850	ENSG00000088782	+	chr20:83633-86120
DEFB128	245939	ENSG00000185982	-	chr20:118221-120702
DEFB129	140881	ENSG00000125903	+	chr20:153435-155921
DEFB4	1673	ENSG00000171711	+	chr8:7261803-7264294
DIP2A	23181	ENSG00000160305	+	chr21:46700731-46703242
DNMT3L	29947	ENSG00000142182	-	chr21:44506479-44508946
DONSON	29980	ENSG00000159147	-	chr21:33936516-33939013
DOPEY2	9980	ENSG00000142197	+	chr21:36456275-36458717
DSCAM	1826	ENSG00000171587	-	chr21:41002576-41005067
DSCR10	259234	ENSG00000181425	+	chr21:38497679-38500151
DSCR2	8624	ENSG00000183527	-	chr21:39477262-39479751
DSCR4	10281	ENSG00000184029	-	chr21:38415231-38417718
DSCR6	53820	ENSG00000183145	+	chr21:37297882-37300362
DSCR8	84677	ENSG00000198054	+	chr21:38413800-38416281
DSCR9	257203	ENSG00000182342	+	chr21:37500228-37502722
DTNBP1	84062	ENSG00000047579	-	chr6:15771233-15773722
ERG	2078	ENSG00000157554	-	chr21:38955504-38957985
ETS2	2114	ENSG00000157557	+	chr21:39096606-39099081
FAM3B	54097	ENSG00000183844	+	chr21:41608107-41610545
FGF7P2	394217	ENSG00000185390	+	chr21:13640996-13643455
FTCD	10841	ENSG00000160282	-	chr21:46399933-46402407
GABPA	2551	ENSG00000154727	+	chr21:26026308-26028759
GART	2618	ENSG00000159131	-	chr21:33837070-33839561
GRIK1	2897	ENSG00000171189	-	chr21:30233700-30236181
H2BFS	54145	ENSG00000197597	+	chr21:43807052-43809521
HEMK2	29104	ENSG00000156239	-	chr21:29179593-29182057
HLCS	3141	ENSG00000159267	-	chr21:37284363-37286884
HMGN1	3150	ENSG00000157576	-	chr21:39642871-39645357
HSF2BP	11077	ENSG00000160207	-	chr21:43903829-43906245
ICEBERG	59082	ENSG00000118168	-	chr11:104514974-104517467
IFIH1	64135	ENSG00000115267	-	chr2:163000585-163002984
IFNAR2	3455	ENSG00000159110	+	chr21:33521655-33524131
IFNGR2	3460	ENSG00000159128	+	chr21:33695070-33697584
IGSF5	150084	ENSG00000183067	+	chr21:40056834-40059326
IL10RB	3588	ENSG00000159113	+	chr21:33558145-33560559
ITGB2	3689	ENSG00000160255	-	chr21:45173136-45175590
JAM2	58494	ENSG00000154721	+	chr21:25931012-25933503
KCNE1	3753	ENSG00000180509	-	chr21:34806397-34808884
KCNE2	9992	ENSG00000159197	+	chr21:34655744-34658207
KCNJ6	3763	ENSG00000157542	-	chr21:38210610-38213059
KRTAP11-1	337880	ENSG00000182591	-	chr21:31175722-31178213
KRTAP13-1	140258	ENSG00000182816	+	chr21:30687730-30690210
KRTAP13-2	337959	ENSG00000198390	-	chr21:30666415-30668897
KRTAP13-3	337960	ENSG00000186975	-	chr21:30720137-30722599
KRTAP13-4	284827	ENSG00000186971	+	chr21:30721966-30724424
KRTAP15-1	254950	ENSG00000186970	+	chr21:30732023-30734507
KRTAP19-1	337882	ENSG00000184351	-	chr21:30774516-30777004
KRTAP19-2	337969	ENSG00000186977	-	chr21:30781593-30784026
KRTAP19-3	337970	ENSG00000186967	-	chr21:30786097-30788623
KRTAP19-5	337972	ENSG00000186965	-	chr21:30796288-30798778
KRTAP19-6	337973	ENSG00000186964	-	chr21:30836027-30838518
KRTAP19-7	337974	ENSG00000186925	-	chr21:30855484-30857974
KRTAP20-1	337975	ENSG00000186923	+	chr21:30908189-30910642

KRTAP20-2	337976	ENSG00000184032	+	chr21:30926962-30929452
KRTAP21-2	337978	ENSG00000187026	-	chr21:31041363-31043881
KRTAP22-1	337979	ENSG00000186924	+	chr21:30892821-30895301
KRTAP23-1	337963	ENSG00000186980	-	chr21:30642830-30645287
KRTAP6-1	337966	ENSG00000184724	-	chr21:30908072-30910587
KRTAP6-2	337967	ENSG00000186930	-	chr21:30893110-30895563
KRTAP8-1	337879	ENSG00000183640	-	chr21:31107417-31109908
LIPI	149998	ENSG00000188992	-	chr21:14501128-14503616
LRRC3	81543	ENSG00000160233	+	chr21:44697380-44699772
MCM3AP	8888	ENSG00000160294	-	chr21:46529668-46532054
MEFV	4210	ENSG00000103313	-	chr16:3246647-3249125
MRAP	56246	ENSG00000170262	+	chr21:32590613-32593069
MRPL39	54148	ENSG00000154719	-	chr21:25901652-25904183
MRPS6	64968	ENSG00000198743	+	chr21:34416907-34419379
MX1	4599	ENSG00000157601	+	chr21:41723064-41725558
MX2	4600	ENSG00000183486	+	chr21:41653349-41655807
NCAM2	4685	ENSG00000154654	+	chr21:21572318-21574809
NDUFV3	4731	ENSG00000160194	+	chr21:43184006-43186436
NLRC3	197358	ENSG00000167984	-	chr16:3567249-3569705
NLRC4	58484	ENSG00000091106	-	chr2:32402581-32405056
NLRP1	22861	ENSG00000091592	-	chr17:5428005-5430496
NLRP10	338322	ENSG00000182261	-	chr11:7941644-7944106
NLRP11	204801	ENSG00000179873	-	chr19:61035119-61037614
NLRP12	91662	ENSG00000142405	-	chr19:59019363-59021848
NLRP13	126204	ENSG00000173572	-	chr19:61135497-61137983
NLRP14	338323	ENSG00000158077	+	chr11:7013881-7016368
NLRP2	55655	ENSG00000022556	+	chr19:60166081-60168505
NLRP3	114548	ENSG00000162711	+	chr1:243904908-243907384
NLRP4	147945	ENSG00000160505	+	chr19:61052766-61055225
NLRP5	126206	ENSG00000171487	+	chr19:61200413-61202899
NLRP6	171389	ENSG00000174885	+	chr11:266071-268565
NLRP7	199713	ENSG00000167634	-	chr19:60150649-60153113
NLRP8	126205	ENSG00000179709	+	chr19:61148561-61151057
NLRP9	338321	ENSG00000185792	-	chr19:60941587-60944058
NLRX1	79671	ENSG00000160703	+	chr11:118542244-118544638
NRG1	3084	ENSG00000157168	+	chr8:31614544-31617022
NRIP1	8204	ENSG00000180530	-	chr21:15262386-15264883
OLIG1	116448	ENSG00000184221	+	chr21:33361876-33364296
OLIG2	10215	ENSG00000185569	+	chr21:33317619-33320101
PCBP3	54039	ENSG00000183570	+	chr21:46138145-46140616
PCNT	5116	ENSG00000160299	+	chr21:46566034-46568514
PCP4	5121	ENSG00000183036	+	chr21:40158769-40161252
PDXK	8566	ENSG00000160209	+	chr21:43960899-43963424
PFKL	5211	ENSG00000141959	+	chr21:44541836-44544341
PGLYRP1	8993	ENSG00000008438	-	chr19:51218164-51220642
PGLYRP2	114770	ENSG00000161031	-	chr19:15451268-15453758
PGLYRP3	114771	ENSG00000159527	-	chr1:150096272-150098759
PGLYRP4	57115	ENSG00000163218	-	chr1:150134079-150136549
PIGP	51227	ENSG00000185808	-	chr21:37367166-37369690
POFUT2	23275	ENSG00000186866	-	chr21:45532267-45534731
PPIAL3	375113	ENSG00000198618	+	chr21:19149473-19151936
PRDM15	63977	ENSG00000141956	-	chr21:42172481-42175008
PRED15	n.a.	ENSG00000196190	+	chr21:21935609-21938097
PRED58	n.a.	ENSG00000184941	-	chr21:45359281-45361782

PRMT2	3275	ENSG00000160310	+	chr21:46877008-46879482
PRSS7	5651	ENSG00000154646	-	chr21:18697811-18700325
PTTG1IP	754	ENSG00000183255	-	chr21:45118173-45120595
PWP2	5822	ENSG00000160220	+	chr21:44349197-44351677
PYCARD	29108	ENSG00000103490	-	chr16:31121704-31124198
PYDC1	260434	ENSG00000169900	-	chr16:31135937-31138409
RBM11	54033	ENSG00000185272	+	chr21:14507854-14510390
RCAN1	1827	ENSG00000159200	-	chr21:34909272-34911750
RIPK1	8737	ENSG00000137275	+	chr6:3019568-3022048
RIPK2	8767	ENSG00000104312	+	chr8:90836727-90839218
RIPK3	11035	ENSG00000129465	-	chr14:23878994-23881392
RIPK4	54101	ENSG00000183421	-	chr21:42060321-42062773
RRP1	23076	ENSG00000160208	+	chr21:43901496-43903847
RRP1B	8568	ENSG00000160214	+	chr21:44031450-44033903
RSPH1	89765	ENSG00000160188	-	chr21:42789539-42791967
RUNX1	861	ENSG00000159216	-	chr21:35343467-35345958
RWDD2B	10069	ENSG00000156253	-	chr21:29313588-29316036
S100B	6285	ENSG00000160307	-	chr21:46849396-46851848
SAMSN1	64092	ENSG00000155307	-	chr21:14877546-14880043
SETD4	54093	ENSG00000185917	-	chr21:36354560-36357002
SH3BGR	6450	ENSG00000185437	+	chr21:39743158-39745625
SIM2	6493	ENSG00000159263	+	chr21:36990821-36993342
SLC19A1	6573	ENSG00000173638	-	chr21:45786768-45789287
SOD1	6647	ENSG00000142168	+	chr21:31951461-31953923
SON	6651	ENSG00000159140	+	chr21:33834772-33837266
STCH	6782	ENSG00000155304	-	chr21:14677339-14679816
SUMO3	6612	ENSG00000184900	-	chr21:45062530-45064917
SYNJ1	8867	ENSG00000159082	-	chr21:33022115-33024554
TFF1	7031	ENSG00000160182	-	chr21:42659723-42662204
TFF2	7032	ENSG00000160181	-	chr21:42644180-42646649
TFF3	7033	ENSG00000160180	-	chr21:42608540-42611080
TIAM1	7074	ENSG00000156299	-	chr21:31853117-31855601
TLR1	7096	ENSG00000174125	-	chr4:38628984-38631440
TLR2	7097	ENSG00000137462	+	chr4:154977763-154980253
TLR3	7098	ENSG00000164342	+	chr4:187362975-187365447
TLR4	7099	ENSG00000136869	+	chr9:117543696-117546185
TLR5	7100	ENSG00000187554	-	chr1:219688355-219690843
TLR6	10333	ENSG00000174130	-	chr4:38653669-38656159
TLR7	51284	ENSG00000196664	+	chrX:12642386-12644848
TLR8	51311	ENSG00000101916	+	chrX:12681936-12684400
TMPRSS3	64699	ENSG00000160183	-	chr21:42689307-42691767
TPH2	121278	ENSG00000139287	+	chr12:70616401-70618862
TRADD	8717	ENSG00000102871	-	chr16:65751277-65753742
TRPM2	7226	ENSG00000142185	+	chr21:44595091-44597558
TTC3	7267	ENSG00000182670	+	chr21:37364992-37367415
TWF2	11344	ENSG00000173366	-	chr3:52248303-52250669
UBASH3A	53347	ENSG00000160185	+	chr21:42694574-42697114
USP25	29761	ENSG00000155313	+	chr21:16021771-16024254
WRB	7485	ENSG00000182093	+	chr21:39671650-39674093

10.2 List of promoter-specific primer sequences

Table 8. Primers for the amplification of 2.5 kb promoters fragments.

HGNC Symbol	Primer Forward	Primer Reverse
ABCC13	AAATACCACTTGTGCCTATGTTGA	GGCCTGGAGGAGGCAAGG
ABCG1	AGCCCTGGGGGCCTAAGT	CGGAAGGTGAAAGGCATGTC
ADAMTS1	TAGACTGAAATACATTACATATGTCTTT	ATAGCCCCTGGCTCGGAG
ADAMTS5	AAAGCAAGGACCACAAAAGAGAA	CACACTTGCTTGCAGGATTGA
AIRE	TTCCCAGGAGGCCATATCTCA	CGCGCCGCTGTTATACCC
AR	GACTGGAAGACTAGTTTAGGTTTTA	ATCTTCCACCTACTTCCCTTACC
ARC	GTGCTCTCCCACCGGTCTCG	GCAGGTCCGGCTCGGCTG
ATP5J	GAGAAAAGTATATCAAGCATACTCAT	CTCGTTCCGGGGCCTTTTC
ATP5O	GCAGCCATTAAAATGATACTTAGAA	TGGCGTAACCGCTAGGTTC
B3GALT5	GAGGAGATTGGCTGTGAGTC	TTCTGGGCTTGGTTTGTTTACC
BACE2	AAGCCAAGGACCACGATAAAG	CGGGAAACTCGAGTCCCG
BACH1	AAACAAAAAGTGCCACATCTTTCA	CGGCCCGACTGACTGAAG
BCL10	CCTTATATCCACACGGTTCACTC	ACAGAGGGACTCGGGGGTC
BIRC2	GTATAGAAACAATCAAACTCATTAACAA	CACTCTGACGCACGATGAC
BIRC3	AACTAGGTGTTCTCTGTTTTAATTTG	TCGCCTGGCCTTTTCCGT
BIRC5	GCTCCTGGAACTCGGTTTTG	AACGGGTCCCGCGATTCA
BTG3	TTTGGAGGTAGGGAAGTTAAAGAT	CGCGTTGAGAGGACTGGC
C21orf100	ATAACCAGAGCACCCATTTATAATG	TTTTATACAGTCAGTCCCAATGTTTT
C21orf109	TGGTTTGACTTCAACATCTATATTGTTTCC	CCAGAATTGCATCCTTCCTGTGTCATCC
C21orf110	GCTGTGTTTTCATTTGAAGTTAGTT	AGGAGGAGATAGAAATAAAATATATTAGT
C21orf114	AGATGCCAGCATCATGCTTC	ACTTCTTGTATCTTTTTTTTTTATTTCC
C21orf115	TTCCCACCACCATGTTGACTTC	ACGAGGCAGCTGTACTATGG
C21orf116	GGAGTAGCACTTTTAATGTCCTTTA	ATTGGTCTCAATAGCTTATCCTTG
C21orf119	TTAGAATGGCGATCATTAAAAAGTC	CCCAAGACTCTGGATCCCTATC
C21orf120	TAGGCTTGGTCTCTGCAACTA	TGGAGCTGAGCCACTGTAGA
C21orf121	ATGGTCTCTAAGAGGTAATTAAGTTAA	GGAGGCGGGGTGTTTTTAGA
C21orf123	GCAGAGGTGTGAGGTCAAGAGCAG	GCAGGAGGTTTCTCAGATGCCC
C21orf125	GAACCCTGGGCACTCGACC	GGGGCAGAGGGCAATGGAC
C21orf128	ACCAGGTCACCTGCCAAGAC	TGTTGTGGCAGCTGACGAG
C21orf129	GGAGGATGGAAGTGGAGACTAC	TCCTCACCTCTGCGTTCCC
C21orf13	TGATTAAATCATCCTGGATTTAGGG	TGAGGGGGACGACGGCAG
C21orf15	CCACTAGCTTGCTGAGACATTA	CGTCCAGGGCCAGACCAG
C21orf19	AGTAAACCACCAGAACCCACAA	CCGCCTATTTTTATTTTTATTCTTATTTA
C21orf2	CCTGCCTTAACTGATAACATTGTC	CCGCTGACACGTTGGCTC
C21orf25	GGGTGCTGGTTAGTTTACAGTTTG	GGTGGCCGGAGGAGGCTCT
C21orf29	CCTCACAAACTGCTAGGATTTC	TCTGCGTCAGCCTTTCCAG
C21orf30	CTGCTGTCACTTTTGTAGTAGGT	GGACAGGGCCTCCGGTGT
C21orf32	GAGACACCAGGGCCACAAATC	CTTCTTGGCGTGTCTTTAAATGT
C21orf33	GTCTTGTGGGGGCTCAAGAC	GCACCCACCCCTCTACAAAC
C21orf34	ATTCACTTAGAGAAAAAAGGATTTTAC	AGTAGTGGCACCCTAAGAGAG
C21orf41	ACCCCACAGCTTGATGTCTG	AACAGGTTTGCAGTGGATTAAATAA
C21orf42	TGCACTGTAGCAAGGCTGGCATTT	GCATGAAGGAGATGAAGATATCTGTTCTCA
C21orf45	GTCATCTTTTTCAAGCATCTCATTT	AGAAGGGGCAAAAGCCTGAA
C21orf55	TGTTTTCTGCGTGAAGTGTTG	TAGACCATTTGCCACTTTACACT
C21orf56	GCATTCCCACCAACAATGAGTT	AAATCACCTCAGTGTTGTTTCTTG
C21orf57	CCTGCTCCGATGACATGTTAG	CCCGGACGTACCGGTGTC
C21orf58	CCACAAAGTCAACAGACTGGTG	CCAGCAGCGCGAACTTTT
C21orf62	AATAGAGGTTGAGTACGCCAAAC	TCTGCCTGGCCACGGATC

C21orf66	GGATGCTCTTAGCTCCAAAACT	CCCACGCGCTCTCAATACT
C21orf67	GTACATGGGAAAATATCAGACTCAC	AGGTGCGTCGGTGAAAGG
C21orf70	ACCTGATTAACACAGTTCCTTATTG	AAGCTGGTCCTGCGTTCTC
C21orf74	TTCCTGTTGTCCATTTGTATGTTTTCTTTT	TCCCAAACAGCTGAGATTAAAGACATAAGC
C21orf77	CCGTGCTCCCTGCTTCTG	CAAATGGAGAAAGGGGAACAAAG
C21orf82	CACTTCATCACTCTGGTATATTATTAT	TTTATAGACTCCCACCCCCTTT
C21orf84	TTCCGTGGCCGTGTACACA	CACCTTTGCCCTGGCTATGTA
C21orf86	GTCCTAATTTTAAAGCACACACAAC	TGGGGACTGTGAGAATCTGTC
C21orf87	TCACTGCAGCGTCAATCCC	ACAGGAAAACGGGACTGTGG
C21orf88	TCACCTTTCATCTTCAGACACAGG	TCCCGTCGCCTCCCGGAGC
C21orf9	TGTGGGAATAAAATACAGAGGATAAC	AACTCCAGATCGAATTAGGATCA
C21orf90	GGGAGGCCCAGTCATTGGA	AACTGGTATGTGTCCTAGAGAGA
C21orf91	CACAATCTCAGCTCACACAACATCCACCTC	TGCGGCTCGGGTTCCTCCAC
C21orf93	ATTTGCACATGCGTTCATGTTC	GGTGCTTCTCGCTTTTAATTTTAC
C21orf94	TCTAATTTTAGTCCCCATTGTTGG	TGTTCTTTGCTTTCTTCATTGATTT
C21orf99	AACTCATGTGATCTGGGTGGTG	AGCCACCTGCTAGAGAGAG
CARD11	CCCTGTTTCAAAAATAATAAAATGAAATA	AGAGAAAAAAGGAGAGAAAGTATGT
CARD14	CTTTGCCTTTAATCATGACATTTGT	CTTCCAGGGACTATCCCACC
CARD18	ATGTCACATATATAAGCTAATACAAATAT	CTGAATGGATAAAAATTCTTCTCTTTT
CARD6	CATCTCTACAAATAATAAAAAAAATATGG	CTCTGATCCGGGGTCCTTGT
CARD8	TTCCATAATAATTAAGAAAAAGATATTGG	GAGAATCCCCTTAGGTCGGTTT
CASP1	TTTTGTCATACCCTTTCTTCTTAATC	TCTTTTTACCTTCCTTTTATACATTTTG
CBR1	CCCTCTTAATAAATTTAATCCATCCACTCT	CACCCGTGGGTTACACGCCC
CBR3	GTCTGTTCGAAGGCATTGCC	ACCGGGCTTCGGACCACC
CBS	TGATATAACTAAGTGAATAAATAAATGAG	CGTCCTCGAGGACCGACC
CCT8	CCCAAAAGGCCCTACGTAATAC	AGGAGCACGCACAGCCTT
CHAF1B	CCTCCCAGATGCTGCAATTAC	AGAGGCACCGTCACCTCC
CHODL	ATTTTAATTCATAGCCCCTACTTTTTT	TGCGCCCTGGTCCTGGAT
CIITA	CGAACCTCTGGTACAGGGAA	GGAGCCCGCCAAGCTAAG
CLDN14	GCCTCCAGAACGCTAAAACA	CCTGCCTCTCTCAGCTTTTC
CLDN17	GAGGGAGAAGGTGTCAGTACAAT	TTTACTTTGAAGACTGGTTCAATTC
CLDN8	TTCTGTCTCATGCGTTCCTGT	TCTGTGCCACAGAAGAGAACA
CLIC6	ACACCCCCCCCCAATAGCC	AGATCCCTGTCCTTGATCGG
COL6A2	CCAGAATCATCTTTGACCCACGACCT	GAGTAAGCAGCAGGGAGGGAA
CRADD	TTAAGAAGTACCTAATAAACCTTGGTATTTGTAT	GCTCCTCCCGGACGCAAAAG
CRYAA	TAAATCAAAATGAATGCAAAAAATTTGT	GAGGCTGGCAGGAGTCAG
CXADR	ACAAAACTAATATTGGATAATATTTGTCC	AGGTGAACTCGCGTCACTTC
CYYR1	GCCTTTCCATGTCAATTGAGTTT	GGGTGGCCCCGACTCCTC
DAOA	CTTGGATGATTTCTTATTGGCTTT	TGGAGCGATTGTAGGAATTGTC
DEFA4	GTATGTGAGTTGTGGATGGAAATG	ACTGGGGACACAGCTTCAATAT
DEFA5	ATAATACGTGTTGGTGAGGATATAAA	GTGAGGAGTCAGCCTGGATTT
DEFA6	GGAGTGGGGTGGGGTTAACTC	ATGGCTAGGGTCGCTGGA
DEFB1	CAAGGTTGAGGATGTACCTCTG	GAGACGCTGGCTCCTTTGG
DEFB103A	AAGGAAAGTCTGACATGAACAAAG	CAGTTAATTAGGAACCCAATGGTAA
DEFB104A	TGAGAAGGGAGAGCCTTTGAC	GAAGCAGACGCAAGTCGG
DEFB105A	GACTGTTTTGAGGATTACATGACAA	TGGGTCTTGAGAGATCAGCAC
DEFB106A	TGGGTCTTGAGAGATCAGCAC	TTTTAGGGAAGACTCTTCGGAAT
DEFB118	ATAATGGGAATTTAAATCACATAGGGT	TGTGAGTCTGAACAAGAGTTTTATT
DEFB119	GAACCTACCCACGTGTCCATAT	GAGATGAGCTTTGCTTTTATCAG
DEFB123	GAGATGGAGTTTCACCATATCTG	GCTGAGCTAATGGGAGAAGGA
DEFB125	TTTCTTTTTTATGGCTGAATAATACTC	CAGCAAGGAATTTAAGTGGTCAC
DEFB126	ACTACAGAGCAACTAGGTAAAAAAT	AGAAGTCTCTATTCAGTATGACTCT
DEFB127	CTTCCACTAAACAACTAGCGAAT	AGTGTGTCCTTAAGCTTCTGAA
DEFB128	CCTCACTCCAGGAACTAGTGTT	TACAAATAGTAATATCTCTTAGGATGTT

DEFB129	AGTTGAGTTTTTCTGCTACATTCC	GCCTCTTCGTGGCCTTATG
DEFB4	GTAGCAACCTGTGCAATTTCTTC	ACATCAAGCCTTCCACCTTATAAA
DIP2A	CCCTAATCAGTATGACTGGTGTCCTTCTA	CGGCCCTACACGGGAGCG
DNMT3L	ATAGAGGGAGACTCCTCAAAAAAC	TTTGGGGACCTCAGTGGGGA
DONSON	AAAAAAAAAAGGAAAGGAATTTACCAG	GTAGGGGGTGGTGAGAAGGTG
DOPEY2	CAAGGTGCTGGCAGATTCG	ACTGTGGGAGATGAAAGATAAATAG
DSCAM	GCAGGAGCCTGTAACCCCT	AGACAACACAGGGATCCATAGG
DSCR10	TAAGGTCAGAATCTTCTCCCATAAT	CCCAGCACAACTGATCTTCTAT
DSCR2	CAATAGGATTTAATTTGTTGTTAATATGT	GCGCGGGCAATAAGTCCC
DSCR4	GGACTGTAAGGAGAGAGGTCAC	GGTGAGTTAAGAAGCGGAAAGT
DSCR6	ACATATTTTTATTAACTGCATACTATTCT	CGGCTTCTCCGTCCCTCT
DSCR8	TTTAAGGAAGGACAAAAAAAATCATAG	ATGAAGCTAAGGAAACAGGAAATAA
DSCR9	ACAAGATCTCCATATGAAAATAAATGA	GGCTCACCCCAGAGATAAAG
DTNBP1	AATAAAACAATAGTGTCAGCATAAAGA	CCCCGCACTCCCACTACC
ERG	CCAGGATTTTATAACTCATAACACCA	CGGAGCGATCTCTAGGAAC
ETS2	ATTTGTCTTCTCAGCATCTGACT	TGAGGCGCTTCCTGTAGAAA
FAM3B	GAGAAATAAATGTGTGCTGTTTAAGT	GGCCAGTCGGGCAAATGG
FGF7P2	ACCATGTTGGCCAGGTTAGTATT	TGCAAGGAATCACAGAAAGACAT
FTCD	GCTTAAACCACAGAAATTTATTCTATC	GCCGCCAGGGCCTTTATAC
GABPA	AAGTGAGTTTTTTTCTCTCCAAGG	GCTGGGGCAACCCAAGGTC
GART	CTGATAAAAGTCTGAATGGTCAAAAA	CCCCAGCCGCTCAGAAAC
GRIK1	TTCAAACTTTAGCCACTCACTATC	TCTTAATTCATGCCGAGATACAG
H2BFS	TCACCTCATCAAGTACTTTGTACT	CGAGCAGCAGATCGTGAAA
HEMK2	TTGGATGTCACAACTAGTGAATATG	TGCCCTCACCCTCATGCG
HLCS	TTTTGGAATTTCTCTGGAAATTTCC	TGGTCGCCCGCGAATCCCTG
HMGN1	AGCCAGCCCCTCAAATTTTTC	TGGGAGAACCGGATGGAACC
HSF2BP	GAGGCTTTTTGAAGGAAGTGATAC	TTTGTGACGCCATCAGCCC
IFIH1	GGTATTATGTGATGCTAATAGGGT	CGCGGCACTTTGGACTCT
IFNAR2	AAGTTTCTTAACTTTTGCTTTTTTCC	GGCCTACGAGGAGCGATAC
IFNGR2	CAAAGCGCTGGGATTAGAGG	GGGCGTCCCGGTGGAGGG
IGSF5	TCTTCTTTACAGTTTTCTCCTATTCT	AGGTAGCCAATTTTTCAAGTTGTT
IL10RB	TCAGCCTCCCTCGCAAGTA	CCGGGAACCAGCGGAGAT
ITGB2	TGTGCTGGGGTGGGTAGGA	GGATGTGTCTCCAAGAAAAGTGT
JAM2	ACGGTGACCATTTACTTCACAAC	GGGAAACGCCTCCTGTTTTG
KCNE1	TTCTCACTCCGTGACTCACC	AGGCTTATTTCCAGAAAGTACTCT
KCNE2	GTGATTTCTGTTTAAAAGTTTTTTCTTAA	GCACCTTCACCTTACAACCATC
KCNJ6	CCCAAGGGTGGAGACCAAAAA	ACGCGGCTCCGAGATAAAA
KRTAP11-1	GTTCAGAGACTCTGTTTTAGTTGA	GACAGAGGGCTGCAGGTAG
KRTAP13-1	CAGTATCTGTTGTATAATTACCTTTTTAT	GGACAGGACCATTTATATACGTTC
KRTAP13-2	GCAATAGAAGATTGTCAGTTTAAAG	TGTGAGTTCAGCTGAGTTATAGT
KRTAP13-3	CTCTCTGCCTCTAAATCTCTAAA	TGGGAAGATTCTGAGTTTTAACG
KRTAP13-4	AGTTGATGTCTTTTATCAAGATTGGT	GATTCTGAGTTTTAATGTCGCATC
KRTAP15-1	TTAAAGAAAGTTTATAAAGTAACTGCTT	AGTTCTGAGCTCAGGTGATTTG
KRTAP19-1	AGTTTACACGGCTAGTGGACAA	GAGTGGTGAGTTGGTTTGTTGTT
KRTAP19-2	AACAAAAAATACTCTTGGCAAAATGT	GCTCACGGTGTCAGAGATAGT
KRTAP19-3	CCTGGCCAACAGCTTTCTTTATTC	CCTCCACAGCCATAGCCCAG
KRTAP19-5	TTGAGGTTTTCCATTTATATCTTGTTT	GAGTGGTGAGTTGGTATGCTG
KRTAP19-6	TTCATACTACCCAAAGTGCTCTAA	GTCAGGGACTAGAGATTCAGTTC
KRTAP19-7	GAATCATCCATAAAAATGTGTTTTTTTC	TCAGAAGTGGTTAGCTGTTGTAG
KRTAP20-1	TGTGGGTGGGAACCATACACAA	TGTCAGAAATGGATGGTTCAGGT
KRTAP20-2	AGTCTTCTCCAAATATCAGTGTTTC	GTTTCAAGAGTGGAGGATTTGGT
KRTAP21-2	TCCCTGATACCTCAATAAAGCTGGAAAA	CAGCAGTTTCTGTAGTAGTTGCAACACA
KRTAP22-1	CAGAATTGGTGAATCCCGATGT	AAAGGTAGGCTCAGTTGTATAGC
KRTAP23-1	CTATCTTAACAGGCCTTGGTTTAC	TGAGAAGATGCCCTGATGTCC

KRTAP6-1	CCATCTTCACAGGTCTTTTATACTTCCTC	TTTCCGTAGTAGCTGCCACACA
KRTAP6-2	CAATGTGGGCTAAATTTTAGTTGTC	GACTTCAGTATGGATACTTCCTGTA
KRTAP8-1	GAAAACCTTGCTCTTTCTTTGATA	GGAGTGGAGGGCTTAGGTAG
LIPI	AGTGTGTTTTATTTATTACACATGCTC	ATCAGAAGTTCACTAGCTCTTTG
LRRC3	CGAGATTGCCCCCACTGAAC	GGCGAAAGGCGGAGACAG
MCM3AP	GGCTGTCACCTTTCTTTTTCCT	GTACAAAATTAATTGGCTTCCTGAAA
MEFV	TGCCACCGTCCAGAGTCAG	CTGTCCTTGGTGAGCAAGAAAA
MRAP	GGGGGATGTCTTTGTGATATTTTG	TGATTTCTGCAGGAATCGTCC
MRPL39	TCTTTAGTCTCCTTTAATTTAGGACAGTTT	CCTCCATAGCAGCGGTGAGAA
MRPS6	AGCTGAGACCCGACTCTTTAT	AAGCTGGTTTTGGTTAACTTAAAATA
MX1	TCAGGGAAAGGTACAAACGTTTAT	GGGCTCACGCCTGTAATCT
MX2	AGGGGGCTTGCCACATCAG	CTTCACTTTTATGTGGCATATGAAC
NCAM2	AAAGAAAAGAGAAGACACATAATTATAA	AAGCTCTACTTTGCTAAGTGAAAT
NDUFV3	TTTTATTTTAATGCAGCACCTTATATTT	CCGACTGCGCCAATTACAG
NLRC3	AAAACAATATTAAAAACATTTCTAAAGGG	TGGTAGATGCCCTGGGAATCC
NLRC4	TATAAATATAGATATGGCAGAGGATAAT	CTCTTTCCTCTGTCTTTGCATAA
NLRP1	AAGGGTCAGGTTTCTGATCTATTT	TCTTCAGGAACTCCAAGTAACAG
NLRP10	TCCTGACATCAAATGATCCACG	CAGACCAGAAGACCAGTGACC
NLRP11	TTGTTGTTTTCATTTGCTTTTATCTG	ATCGGCCAGAAAGGTAACTCA
NLRP12	TCTCCATCTCCTGAAACATGAAG	GCCAGTGGTTGGAGGAGAGA
NLRP13	ACTTGCTGTTTTCACTGACCG	AACACATGCAGTTTCTCACTTCA
NLRP14	AAATCAAACTTTCAGGTCTGTTTTAA	CAGATTTGAAAATGAAATACTTATATAGA
NLRP2	CTCACTTTCTTATTCTTTCTAGGATA	AAGCTCTATTAAGGCTTCTTTCT
NLRP3	GAGCTGCAGAGTAGGTCTGT	GAAAGCAGAAGAGAATGTTTTCAG
NLRP4	AGGGAGGGACTTCCAGTTGTT	GACAGTGACCAGGAACAATAAATAA
NLRP5	TCTTGATGATCGTGTGGTTTTTG	ACATCGCCATGGAAAGTAACTG
NLRP6	CTGCCAAAGATGGTCTTGAACTCTAG	TCTCTTCCTCCCTGGGTCTAGCA
NLRP7	CCAGGGTCTTGCTATTTTGTCC	CTCCCTCAGGTCAGGTCTTG
NLRP8	TAGGAGTAGAACTACTGAATTCCA	TCAGTGTCCACGATAAAGACAAC
NLRP9	TGGAGAAATGTCTATTCTAGTCCT	AGGGAAAAGAAAGATCCTCAAAAAT
NLRX1	CCAAGCCTGGCTAATTTTTTTGTATTTCT	TAGTACTGCCCTCTGCAGGTTCC
NRG1	CTTTGGCATTGCAACTTTTCAGGTC	TTCTGGCTGTCGGTTCGGACTC
NRIP1	TGAGTCTACCCAAGAGCAAAAA	TTCAATAGAAGTGTTCACAAGGG
OLIG1	CCTACGGGCCCCACCTAG	ACCCCTTTAAACCCGGCTTG
OLIG2	GGCACCAGGGCCAAGGGA	GGCTCGGCCGGTTTTTATAG
PCBP3	CGGGCTCCATTCTGAGAAAC	TGAGGGTGGTGGCTTAAGGT
PCNT	ACTGTCACCTCGACTATAACTG	TTCGCTCTATTTACACTCCCTC
PCP4	TGAGAAAAATCCCCCTTAAGGTAA	ACACTGCTCAGGGGTCGC
PDXK	CCCTCCAGGGCTCTCTCCCT	TCCGAACCCGCGAGTTCCGC
PFKL	CCTCGGGCCTCAGTTTCCC	CGCGCCTTCCGCCCGCCC
PGLYRP1	TAATGAGATCTTGTTCTTTGGATTAG	TGGAGGGGCCGCCTTTATC
PGLYRP2	CTGCTCAATTTTGTCACGAACC	GAACTGAGCAGAGCCTTTGAC
PGLYRP3	TGGAACAGATAGATCTTCATTGGT	GGTCCCAGGACTCTGACCG
PGLYRP4	GGTGGGAAGTTCAGAGAATATTGT	CAGTCCCTTTGCTGGAAGAG
PIGP	GACAGTGAAAATTGGTACAACCACTTC	CCACCTCAGCGCTGAGTTCTCC
POFUT2	GCCTGAGGTTAAAAGATTACCCC	CCCCGCCCCCGTGAAATG
PPIAL3	CATTATATTTAGAGGTGCAGAATACA	GCGTCTGCAAAGCCAAGTAA
PRDM15	CTGATAGCCATGGAAGCTCTTGTCTGTT	CGGACTCCGGGTTGCGATCC
PRED15	TGTTGGTCTTGCAGTCAATGTTA	GAGTTTTAATTGTCCTTACAGTTTTG
PRED58	CCCATATACATAACAGAAACTCCTCATC	CTCGCACTGACCTGTAGAGA
PRMT2	TAGCCTCTGTTGCAGTAACTATT	TTGTGGGGAGAGGAAGTTTGG
PRSS7	CTTTATCACCCTCCATCTCTTCACTTCCC	TTTTGGTTTTGAAGGCTTGCTAATTTAAG
PTTG1IP	AGGCTGAGGCAGAGGAGTC	GCCTGCGCGATAGAGGAA
PWP2	GAGAAGGAGCACCGGGTTC	AGACACTTCCCGGGGGTCC

PYCARD	ACCCGTAGTTCCAGCTACACA	CTCTGGATTGGGGCCCCAG
PYDC1	CAATGGGATGAGGTCTCTGGG	GGATGTCCCGGGAAGGAGAC
RBM11	AAAAGAGAGTAAGAGTTGGCTGCACA	AGCAGGGAACATCCTCCGGTCTC
RCAN1	TACCTTCTCTCTCTGTGGATTTG	GCTCCGCCGTTAACCCCC
RIPK1	CATTTCGTTCATGCGGTTTTCA	TCAAGAACGCCCAAAATGGTAC
RIPK2	TTTCTTTAAAAAAAAAAAATGACATGGAA	TTTTCTAGAGACTGGTGCCAAC
RIPK3	TCTACTAAAAATACAAAAATTAGTAGCT	GACCAAGACGGTGAGTCTACTT
RIPK4	AGGAAACTGGCTTGGAATTCTT	GCTCCGCCCTGTTACGTC
RIPK4	AGTATAGGAGGAAACTGGCTTGGAAT	GCTCCGCCCTGTTACGTCACTT
RRP1	ATGGTGAAACTCCATCTCACTAAA	GGGCTGATGGCGTCACAAA
RRP1B	ACTGTGCATCTGACACGCC	CCTGGTACCCAGCACACTGA
RSPH1	GTCACTGCCAGAAAACTATAGTTAT	AGCCTTTGCGTACTGTCATAG
RUNX1	CAGACTCTTGGGTACAATCCCC	AGTTTTCACACAACCCAAATTACAA
RWDD2B	GCTATCATACCCAGCCATAATTTT	GACCGGAAGCTGGGGCTC
S100B	CTTGGCTTCTCCCTTCAATCA	GTCGTCTCTGCTGCGGGT
SAMSN1	AATAATTTGCTACTAGAAAGTAATGAG	TTGTTTTCTATGGCCAAGGAATG
SETD4	TTTTGTTTTGTTTTGTTTTGTTTTAGC	GCCCAAGGCACGTGGAAATA
SH3BGR	CTGGTCTTTGTGAACAGACATAAA	CACTGGACCCTTTCCCTATCAC
SIM2	TTGGGTTTCTTCCTTCACAATTCCAC	TGTGTGGCTAGGCGCAGCAC
SLC19A1	TCTCTCCAGCCAAGGCAGAGGA	CTCCGGGACTACAGCGCCCA
SOD1	GTGGCCTGTTCAGGATCCTATA	CCGAGGACTGCAACGGAAA
SON	AGCAGTGTAACCTAGGAGATTAAC	CGTCCGCTCCGTTCTCTC
STCH	TACTTCCCAACCCCAGTGCC	GACTGTACCAGACGTGAGGC
SUMO3	CGTAGCCCTTCCTGTACTGA	GCTCTGGGCGGTTGGAGA
SYNJ1	CTGTGTGCGTGTTTAATTTTCCT	CCGGCTTGCTCACCTCTTC
TFF1	TAAAATCCAGCTGGGCACAGT	CCCGAGCCCCGGATTTTAT
TFF2	GAAGACGGTGGGTGAAGTGA	GTCCACTGCCGGGTATGTTTT
TFF3	CACCACTTCATGCTGTCCTGTG	GCTCTGGCAGCCATGACCACC
TIAM1	CCTGGCCAGCACATTGTTAATA	CGGTCGTCCGGGATCTCC
TLR1	CGGGCAGAGGCTGCAATC	ATTCAAGCACTTCCTTGAATTTATTT
TLR2	TGAATAATAAAACCAAAAGGAGAGATAA	TAACAAAATACCAGAGACCAAGTAAT
TLR3	AAAAAAATAAAAACGGCTGTTTTATTAA	TTGAAAAAAAATGCTTTTATGTTTGGT
TLR4	GTCTGGTACCTGGACCTGTGATGATTA	CACAGCAATTGGTGTATTCAAAGCAG
TLR5	TTAAGAAGTACCTAATAAACCTTGGTATTTGTAT	GCTCCTCCCGGACGCAAAAG
TLR6	TGTGATAAGACTATGCTTAGTTTTATAA	AGTGGCTATCCTAAAGGGTTGT
TLR7	CAAAACCAGCCTGGCCAATATA	TGATGGGAGGCAGGAAGTGTA
TLR8	GAGAGGCTTGTCTGCTGGT	AAACTGGCTAACTTCCTGAAATG
TMPRSS3	CATCTGACACACGTGTTAACATG	CCAACGGCTTGCATCAAAATAA
TPH2	AAATTGGTCTGTTTTTTTAATTGCTG	CGATGGCCACAGTCAGATTAC
TRADD	ATTATTATTTTGAGATAAACTGTTACTCT	GCCAGGCTCCGCTCTACT
TRPM2	GAGCCGGGTGCGGTCTGT	TTCCAGACACCAGTGACTACTTT
TTC3	GTGCAATCTCAGCTCACGG	TGAAGCGAGGTGCGTCAC
TWF2	CTGGCTAATACATATCTATCTATTAATC	GTGGGTCTAAGGATGATGATTGA
UBASH3A	GTAGCTGGAACTAATGCCACCATGTCTTTT	CCCTTAGGCCAAGAAGCTCGC
USP25	TGAAGGCCTTCCTCCTGTATC	GCGTTCGCACGTTCCTTTG
WRB	TGGGGTGTGGTAGTGGACTTC	CAAGCGCCCCAGAACAGG

Table 9. Forward primers for the amplification of 500 bp promoter fragments.

HGNC Symbol	Primer Forward (0.5kb)
ABCG1	AATTTCATGAGGGACACGAGGAACTGGCA
ADAMTS1	GCTGGGAAGGTGGAGAAGTGGGGT
ADAMTS5	AAGGGACACAGACACGCCGCTTC
ATP5J	GACAACCCATTGTTATCTCCTTGGTT
BACE2	GCTGACCTCCGAGTCACCCCC
C21orf119	TGCCCGGCACACGGAGGC
C21orf123	CTCCACACCTCACAGGGCAGC
C21orf125	TGCAGCATACAGATCTGGGGGAGA
C21orf129	GCCATTGACTGGAGGTCAGCAGGT
C21orf13	TGCATTGTTTTGTCATCAGATTTTG
C21orf30	GCAGGTGAACTGTTGCTGGCCT
C21orf42	TCCCAATATATTTATAAGAGGTCTGGGT
C21orf45	AGGGAAGATCGCTTAAGGCTAGGAG
C21orf55	GCCTAAGGTGCAAGCAGAAGA
C21orf62	AGAGAATAATGTCTATCTTTGGAGTTCA
C21orf66	AGTCACACCTCAGGTCAAAGTT
C21orf77	TCTTGCTGGAAATCGAGGATGGCT
C21orf88	GGGCGTGGCGTGAGTAGCGATAAG
C21orf91	TGACACCTGAATATGAACTTTTCCTTCG
CBR1	TCCCACCATTATTGGAAAATCTGGCT
CBR3	CTCCCACTTGGGGGCACATTC
CCT8	CATTCCCAGCGCTTCCTATGGT
CHAF1B	TCTGCCTACCGACTGATTTACAGCC
CHODL	TCACTGAGCAGGTTTTGGCTCTCA
CLDN17	GATTACCATCTTAAGGAAAAGTGGCT
CLDN8	TCACATGTGTCCCAAAGAATGTGG
CLIC6	CAAGGACTTCTGGTCAACCCCAGG
COL6A2	CAAGCCCACCTGTGGTTTCCAGC
CXADR	TCGGTAGCTGCAGAGGGGCTCTATC
CYYR1	TGCCATACACTTTCTGTACTTCCCAATC
DSCR2	GCCCCTCTAGCCCGCAGTCCT
DSCR9	CTTTGTGTCAAACAGACCTGTCATTG
FAM3B	GAGGTGGGGCGCCTGCTAG
GART	GGCCTGAGCCGGGGAACGAC
HEMK2	AACCGCATCATCTTGCTTAAGGTTT
HLCS	GGCAATCCCTCAGTTTCTAGTGGGC
JAM2	TTTAAAGCACGCAACAGAGCATTCA
KCNJ6	CATGGAGCACCCACGTTTCCAG
LIPI	AACATTTGCATAAATGAGTTGCTGA
LRRC3	TGGGCGGAGATCTAGCCCTACTCA
MRPL39	CCCACTCCAGCCGAACTAAGAACC
MX2	GATCATGTACCTGAGAGCATGTCC
OLIG1	ATGGGTGGAGCTTCAGGGTTGC
OLIG2	GCTGCGTGGGGCATCAGGTTAGT
PCP4	AAGGGAAACACAGCTCCTGGGAA
PRSS7	GCAACACTTTTGATTGTTTCCATAAG
PTTG1IP	GGACATCGGCGGGAGCCCAG
PWP2	GTCATCCACCCGCCTCAGTC
RBM11	CAAAAACTCATCTAGGGAGCACTTTCT
RCAN1	CAAAAGTGGGCCCCTGTGCTTCAA
RWDD2B	CTGGGAAATGACAACGGGGTTGGCA
SETD4	AGCCCATAATCAAAGTGGGGGTGG

SH3BGR	GCCCATGGCTGGTCAGTTTCTGTT
SIM2	TTTCCCCTTTTCAGCTCTTGGACCTGCA
SLC19A1	GTCAGGCAGGAGGGCGGG
SOD1	CCGAATTCTGCCAACCAAATAAGAAACT
SON	GGGGGCGGGGGCTGAAA
STCH	GACCCTTCTACGGAGGATGGCTGG
SYNJ1	CTCCTCTCCCGTCAGGCCCGTG
TTC3	GGGCAGGGAACAAGGCTCGC
USP25	TTTCCCTTCAGGCAAGTTCAAGGC
WRB	CTGGAACGAAAACGTAAGGC

10.3 Promoter activities determined by two quantification approaches

Table 10. Comparison the results from two quantification approaches. The pattern of promoter activities determined by quantifications based on (i) the intensity of reporter fluorescence and (ii) the number of transfected cells.

HGNC Symbol	Standard condition		PMA treatment		TSA treatment		Serum depletion		ARS treatment		Serum depletion & ARS treatment	
	i	ii	i	ii	i	ii	i	ii	i	ii	i	ii
ADAMTS1	1	1	1	1	1	1	1	1	1	1	1	1
ATP5O	1	1	1	1	1	1	1	1	1	1	1	1
BIRC5	1	1	1	1	1	1	1	1	1	1	1	1
C21orf119	1	1	1	1	1	1	1	1	1	1	1	1
C21orf33	1	1	1	1	1	1	1	1	1	1	1	1
C21orf55	1	1	1	1	1	1	1	1	1	1	1	1
C21orf57	1	1	1	1	1	1	1	1	1	1	1	1
C21orf58	1	1	1	1	1	1	1	1	1	1	1	1
C21orf67	1	1	1	1	1	1	1	1	1	1	1	1
C21orf70	1	1	1	1	1	1	1	1	1	1	1	1
C21orf86	1	1	1	1	1	1	1	1	1	1	1	1
C21orf91	1	1	1	1	1	1	1	1	1	1	1	1
CBS	1	1	1	1	1	1	1	1	1	1	1	1
CCT8	1	1	1	1	1	1	1	1	1	1	1	1
DIP2A	1	1	1	1	1	1	1	1	1	1	1	1
DSCR4	1	1	1	1	1	1	1	1	1	1	1	1
DSCR8	1	1	1	1	1	1	1	1	1	1	1	1
HMGN1	1	1	1	1	1	1	1	1	1	1	1	1
MCM3AP	1	1	1	1	1	1	1	1	1	1	1	1
OLIG1	1	1	1	1	1	1	1	1	1	1	1	1
OLIG2	1	1	1	1	1	1	1	1	1	1	1	1
PRDM15	1	1	1	1	1	1	1	1	1	1	1	1
PWP2	1	1	1	1	1	1	1	1	1	1	1	1
RIPK4	1	1	1	1	1	1	1	1	1	1	1	1
RRP1	1	1	1	1	1	1	1	1	1	1	1	1
SETD4	1	1	1	1	1	1	1	1	1	1	1	1
SOD1	1	1	1	1	1	1	1	1	1	1	1	1
SON	1	1	1	1	1	1	1	1	1	1	1	1
STCH	1	1	1	1	1	1	1	1	1	1	1	1
SYNJ1	1	1	1	1	1	1	1	1	1	1	1	1

TTC3	1	1	1	1	1	1	1	1	1	1	1	1
DONSON	1	1	1	1	0	1	1	1	1	1	1	1
GART	1	1	1	1	0	1	1	1	1	1	1	1
LRRC3	1	1	1	1	0	1	1	1	1	1	1	1
PIGP	1	1	1	1	0	1	1	1	1	1	1	1
C21orf115	1	1	1	1	0	1	0	1	0	1	1	0
C21orf13	1	1	1	0	1	1	1	1	1	1	1	1
C21orf66	1	1	1	0	1	0	1	1	1	1	1	1
KRTAP19-1	1	1	1	0	0	0	1	1	1	1	1	1
BACH1	1	1	0	1	1	1	1	1	1	1	1	1
NDUFV3	1	1	0	1	1	1	1	1	1	1	1	1
NLRP4	1	1	0	1	1	1	1	1	1	1	1	1
TIAM1	1	1	0	1	1	1	1	1	1	1	1	1
BTG3	1	1	0	1	0	1	1	1	1	1	1	1
PTTG1IP	1	1	0	1	0	1	1	1	1	1	1	1
CBR1	1	1	0	1	0	1	0	1	1	1	1	1
C21orf56	1	1	0	1	0	1	0	1	1	1	0	1
RCAN1	1	1	0	0	1	1	1	1	1	1	1	1
ATP5J	1	1	0	0	0	1	1	1	1	1	1	1
KRTAP21-2	1	1	0	0	0	0	1	1	1	0	0	0
MRPL39	1	0	1	0	1	1	1	1	1	0	1	0
KRTAP20-1	1	0	1	0	1	0	0	0	1	0	1	0
RBM11	1	0	0	1	1	0	0	1	1	0	0	0
C21orf29	1	0	0	1	0	1	0	0	0	0	0	0
MRAP	1	0	0	1	0	0	1	1	0	0	1	0
C21orf30	1	0	0	1	0	0	0	1	1	1	0	0
SuMO3	1	0	0	1	0	0	0	1	0	0	0	0
C21orf120	1	0	0	0	0	0	0	1	0	0	0	0
BCL10	1	0	0	0	0	0	0	0	0	0	0	0
DEFA6	1	0	0	0	0	0	0	0	0	0	0	0
NLRP14	1	0	0	0	0	0	0	0	0	0	0	0
ARC	0	1	1	1	1	1	1	1	1	1	1	1
IFNAR2	0	1	1	1	0	1	1	1	1	1	1	1
BIRC2	0	1	0	1	1	1	0	1	0	1	0	1
PCNT	0	1	0	1	0	1	1	1	1	1	1	1
CXADR	0	1	0	1	0	1	0	1	1	1	1	1
COL6A2	0	1	0	1	0	1	0	1	1	0	0	1
RSPH1	0	1	0	1	0	1	0	1	0	1	1	1
SIM2	0	1	0	1	0	1	0	1	0	1	1	1
CARD8	0	1	0	1	0	1	0	1	0	1	0	1
DTNBP1	0	1	0	1	0	1	0	1	0	1	0	1
IFIH1	0	1	0	1	0	1	0	1	0	1	0	1
NLRP2	0	1	0	1	0	1	0	1	0	1	0	1
RIPK2	0	1	0	1	0	1	0	1	0	1	0	1
TLR5	0	1	0	1	0	1	0	1	0	1	0	1
TRADD	0	1	0	1	0	1	0	1	0	1	0	1
POFUT2	0	1	0	1	0	1	0	1	0	1	0	0
BIRC3	0	1	0	1	0	1	0	0	0	0	0	0
CRADD	0	1	0	0	0	1	0	1	0	1	0	1
DSCR2	0	1	0	0	0	1	0	1	0	0	0	0
TLR1	0	1	0	0	0	1	0	0	0	0	0	0
AR	0	1	0	0	0	0	0	1	0	0	0	1
C21orf87	0	0	1	1	0	1	0	1	0	1	0	0
CBR3	0	0	1	1	0	1	0	0	0	1	0	0

TFF1	0	0	1	1	0	0	0	1	0	1	1	0
B3GALT5	0	0	1	1	0	0	0	1	0	1	0	0
TFF3	0	0	1	1	0	0	0	0	0	1	1	0
KRTAP19-5	0	0	1	0	0	1	0	1	1	0	0	0
USP25	0	0	0	1	1	1	0	1	1	1	1	0
C21orf88	0	0	0	1	0	1	1	1	0	0	0	0
C21orf45	0	0	0	1	0	1	0	1	1	1	1	1
SLC19A1	0	0	0	1	0	1	0	1	1	1	1	0
ETS2	0	0	0	1	0	1	0	1	0	1	1	0
IFNGR2	0	0	0	1	0	1	0	1	0	1	0	0
KRTAP11-1	0	0	0	1	0	1	0	1	0	1	0	0
KCNJ6	0	0	0	1	0	1	0	1	0	0	1	0
AIRE	0	0	0	1	0	1	0	1	0	0	0	0
C21orf25	0	0	0	1	0	1	0	1	0	0	0	0
CHAF1B	0	0	0	1	0	1	0	1	0	0	0	0
GRIK1	0	0	0	1	0	1	0	1	0	0	0	0
MX2	0	0	0	1	0	1	0	1	0	0	0	0
CLDN14	0	0	0	1	0	1	0	0	0	1	1	0
HLCS	0	0	0	1	0	1	0	0	0	1	0	0
BACE2	0	0	0	1	0	1	0	0	0	0	0	0
C21orf2	0	0	0	1	0	1	0	0	0	0	0	0
CARD11	0	0	0	1	0	1	0	0	0	0	0	0
PCP4	0	0	0	1	0	1	0	0	0	0	0	0
RRP1B	0	0	0	1	0	1	0	0	0	0	0	0
C21orf129	0	0	0	1	0	0	1	1	0	0	0	0
ADAMTS5	0	0	0	1	0	0	0	1	0	1	1	0
H2BFS	0	0	0	1	0	0	0	1	0	1	0	0
C21orf123	0	0	0	1	0	0	0	1	0	0	0	0
C21orf125	0	0	0	1	0	0	0	1	0	0	0	0
FTCD	0	0	0	1	0	0	0	1	0	0	0	0
HSF2BP	0	0	0	1	0	0	0	1	0	0	0	0
PRED58	0	0	0	1	0	0	0	1	0	0	0	0
PRMT2	0	0	0	1	0	0	0	1	0	0	0	0
C21orf84	0	0	0	1	0	0	0	0	1	0	0	0
TMPRSS3	0	0	0	1	0	0	0	0	0	1	1	0
WRB	0	0	0	1	0	0	0	0	0	1	1	0
C21orf19	0	0	0	1	0	0	0	0	0	1	0	0
DOPEY2	0	0	0	1	0	0	0	0	0	1	0	0
C21orf77	0	0	0	1	0	0	0	0	0	0	1	0
C21orf99	0	0	0	1	0	0	0	0	0	0	1	0
CLIC6	0	0	0	1	0	0	0	0	0	0	1	0
ABCG1	0	0	0	1	0	0	0	0	0	0	0	0
C21orf100	0	0	0	1	0	0	0	0	0	0	0	0
C21orf116	0	0	0	1	0	0	0	0	0	0	0	0
C21orf128	0	0	0	1	0	0	0	0	0	0	0	0
C21orf32	0	0	0	1	0	0	0	0	0	0	0	0
C21orf42	0	0	0	1	0	0	0	0	0	0	0	0
C21orf74	0	0	0	1	0	0	0	0	0	0	0	0
C21orf82	0	0	0	1	0	0	0	0	0	0	0	0
CHODL	0	0	0	1	0	0	0	0	0	0	0	0
CRYAA	0	0	0	1	0	0	0	0	0	0	0	0
CYYR1	0	0	0	1	0	0	0	0	0	0	0	0
DNMT3L	0	0	0	1	0	0	0	0	0	0	0	0
DSCR10	0	0	0	1	0	0	0	0	0	0	0	0

DSCR6	0	0	0	1	0	0	0	0	0	0	0	0
DSCR9	0	0	0	1	0	0	0	0	0	0	0	0
FAM3B	0	0	0	1	0	0	0	0	0	0	0	0
FGF7P2	0	0	0	1	0	0	0	0	0	0	0	0
IL10RB	0	0	0	1	0	0	0	0	0	0	0	0
ITGB2	0	0	0	1	0	0	0	0	0	0	0	0
KRTAP19-6	0	0	0	1	0	0	0	0	0	0	0	0
KRTAP19-7	0	0	0	1	0	0	0	0	0	0	0	0
PCBP3	0	0	0	1	0	0	0	0	0	0	0	0
PFKL	0	0	0	1	0	0	0	0	0	0	0	0
S100B	0	0	0	1	0	0	0	0	0	0	0	0
TFF2	0	0	0	1	0	0	0	0	0	0	0	0
TRPM2	0	0	0	1	0	0	0	0	0	0	0	0
KRTAP13-4	0	0	0	0	1	1	0	1	1	0	0	1
CASP1	0	0	0	0	1	0	1	0	0	0	1	0
DEFB127	0	0	0	0	1	0	1	0	0	0	0	0
NLRC4	0	0	0	0	1	0	1	0	0	0	0	0
NLRP9	0	0	0	0	1	0	0	0	0	0	0	0
SH3BGR	0	0	0	0	1	0	0	0	0	0	0	0
RUNX1	0	0	0	0	0	1	0	1	0	0	0	0
RWDD2B	0	0	0	0	0	1	0	1	0	0	0	0
KRTAP13-2	0	0	0	0	0	1	0	0	0	0	0	0
MEFV	0	0	0	0	0	0	1	0	0	0	1	0
CIITA	0	0	0	0	0	0	1	0	0	0	0	0
GABPA	0	0	0	0	0	0	0	1	0	1	0	0
UBASH3A	0	0	0	0	0	0	0	1	0	1	0	0
C21orf93	0	0	0	0	0	0	0	1	0	0	0	0
KCNE1	0	0	0	0	0	0	0	1	0	0	0	0
MX1	0	0	0	0	0	0	0	1	0	0	0	0
C21orf114	0	0	0	0	0	0	0	0	0	0	1	0
CARD14	0	0	0	0	0	0	0	0	0	0	1	0
DAOA	0	0	0	0	0	0	0	0	0	0	1	0
ABCC13	0	0	0	0	0	0	0	0	0	0	0	0
C21orf109	0	0	0	0	0	0	0	0	0	0	0	0
C21orf110	0	0	0	0	0	0	0	0	0	0	0	0
C21orf121	0	0	0	0	0	0	0	0	0	0	0	0
C21orf15	0	0	0	0	0	0	0	0	0	0	0	0
C21orf34	0	0	0	0	0	0	0	0	0	0	0	0
C21orf41	0	0	0	0	0	0	0	0	0	0	0	0
C21orf62	0	0	0	0	0	0	0	0	0	0	0	0
C21orf9	0	0	0	0	0	0	0	0	0	0	0	0
C21orf90	0	0	0	0	0	0	0	0	0	0	0	0
C21orf94	0	0	0	0	0	0	0	0	0	0	0	0
CARD18	0	0	0	0	0	0	0	0	0	0	0	0
CARD6	0	0	0	0	0	0	0	0	0	0	0	0
CLDN17	0	0	0	0	0	0	0	0	0	0	0	0
CLDN8	0	0	0	0	0	0	0	0	0	0	0	0
DEFA4	0	0	0	0	0	0	0	0	0	0	0	0
DEFA5	0	0	0	0	0	0	0	0	0	0	0	0
DEFB1	0	0	0	0	0	0	0	0	0	0	0	0
DEFB103A	0	0	0	0	0	0	0	0	0	0	0	0
DEFB104A	0	0	0	0	0	0	0	0	0	0	0	0
DEFB105A	0	0	0	0	0	0	0	0	0	0	0	0
DEFB106A	0	0	0	0	0	0	0	0	0	0	0	0

DEFB118	0	0	0	0	0	0	0	0	0	0	0	0
DEFB119	0	0	0	0	0	0	0	0	0	0	0	0
DEFB123	0	0	0	0	0	0	0	0	0	0	0	0
DEFB125	0	0	0	0	0	0	0	0	0	0	0	0
DEFB126	0	0	0	0	0	0	0	0	0	0	0	0
DEFB128	0	0	0	0	0	0	0	0	0	0	0	0
DEFB129	0	0	0	0	0	0	0	0	0	0	0	0
DEFB4	0	0	0	0	0	0	0	0	0	0	0	0
DSCAM	0	0	0	0	0	0	0	0	0	0	0	0
ERG	0	0	0	0	0	0	0	0	0	0	0	0
HEMK2	0	0	0	0	0	0	0	0	0	0	0	0
IGSF5	0	0	0	0	0	0	0	0	0	0	0	0
JAM2	0	0	0	0	0	0	0	0	0	0	0	0
KCNE2	0	0	0	0	0	0	0	0	0	0	0	0
KRTAP13-1	0	0	0	0	0	0	0	0	0	0	0	0
KRTAP13-3	0	0	0	0	0	0	0	0	0	0	0	0
KRTAP15-1	0	0	0	0	0	0	0	0	0	0	0	0
KRTAP19-2	0	0	0	0	0	0	0	0	0	0	0	0
KRTAP19-3	0	0	0	0	0	0	0	0	0	0	0	0
KRTAP20-2	0	0	0	0	0	0	0	0	0	0	0	0
KRTAP22-1	0	0	0	0	0	0	0	0	0	0	0	0
KRTAP23-1	0	0	0	0	0	0	0	0	0	0	0	0
KRTAP6-1	0	0	0	0	0	0	0	0	0	0	0	0
KRTAP6-2	0	0	0	0	0	0	0	0	0	0	0	0
KRTAP8-1	0	0	0	0	0	0	0	0	0	0	0	0
LIPI	0	0	0	0	0	0	0	0	0	0	0	0
MRPS6	0	0	0	0	0	0	0	0	0	0	0	0
NCAM2	0	0	0	0	0	0	0	0	0	0	0	0
NLRC3	0	0	0	0	0	0	0	0	0	0	0	0
NLRP1	0	0	0	0	0	0	0	0	0	0	0	0
NLRP10	0	0	0	0	0	0	0	0	0	0	0	0
NLRP11	0	0	0	0	0	0	0	0	0	0	0	0
NLRP12	0	0	0	0	0	0	0	0	0	0	0	0
NLRP13	0	0	0	0	0	0	0	0	0	0	0	0
NLRP3	0	0	0	0	0	0	0	0	0	0	0	0
NLRP5	0	0	0	0	0	0	0	0	0	0	0	0
NLRP6	0	0	0	0	0	0	0	0	0	0	0	0
NLRP7	0	0	0	0	0	0	0	0	0	0	0	0
NLRP8	0	0	0	0	0	0	0	0	0	0	0	0
NLRX1	0	0	0	0	0	0	0	0	0	0	0	0
NRG1	0	0	0	0	0	0	0	0	0	0	0	0
NRIP1	0	0	0	0	0	0	0	0	0	0	0	0
PDXK	0	0	0	0	0	0	0	0	0	0	0	0
PGLYRP1	0	0	0	0	0	0	0	0	0	0	0	0
PGLYRP2	0	0	0	0	0	0	0	0	0	0	0	0
PGLYRP3	0	0	0	0	0	0	0	0	0	0	0	0
PGLYRP4	0	0	0	0	0	0	0	0	0	0	0	0
PPIAL3	0	0	0	0	0	0	0	0	0	0	0	0
PRED15	0	0	0	0	0	0	0	0	0	0	0	0
PRSS7	0	0	0	0	0	0	0	0	0	0	0	0
PYCARD	0	0	0	0	0	0	0	0	0	0	0	0
PYDC1	0	0	0	0	0	0	0	0	0	0	0	0
RIPK1	0	0	0	0	0	0	0	0	0	0	0	0
RIPK3	0	0	0	0	0	0	0	0	0	0	0	0

SAMSN1	0	0	0	0	0	0	0	0	0	0	0	0
TLR2	0	0	0	0	0	0	0	0	0	0	0	0
TLR3	0	0	0	0	0	0	0	0	0	0	0	0
TLR4	0	0	0	0	0	0	0	0	0	0	0	0
TLR6	0	0	0	0	0	0	0	0	0	0	0	0
TLR7	0	0	0	0	0	0	0	0	0	0	0	0
TLR8	0	0	0	0	0	0	0	0	0	0	0	0
TPH2	0	0	0	0	0	0	0	0	0	0	0	0
TWF2	0	0	0	0	0	0	0	0	0	0	0	0

Each row in the panel represents one promoter. Two columns in each growth condition represent the promoter activity status determined by the quantifications on the basis of fluorescence inetsnsity (i) and tranfected cell number (ii). Value 1 and 0 represent active promoters (green boxes) and silent promoters, respectively.

10.4 Summary of the results

Table 11. Summary of promoter activities, endogenous gene expression, Pol IIa binding sites occupancy, presence of core promoter elements and EST expression for 255 tested 2.5 kb promoter fragments.

HGNC Symbol	Standard growth condition	PMA treatment	TSA treatment	Serum depletion	ARS treatment	Serum depletion & ARS treatment	Gene expression	PolIIa	CGI	TATA	INR	DPE	Tissue Hits
ADAMTS1	1	1	1	1	1	1	1	1	1	1	0	1	34
SYNJ1	1	1	1	1	1	1	1	1	1	0	1	0	28
HMGN1	1	1	1	1	1	1	1	1	1	0	0	1	45
SOD1	1	1	1	1	1	1	1	1	1	0	0	1	44
TTC3	1	1	1	1	1	1	1	1	1	0	0	1	44
SON	1	1	1	1	1	1	1	1	1	0	0	1	43
GART	1	1	1	1	1	1	1	1	1	0	0	1	40
RRP1	1	1	1	1	1	1	1	1	1	0	0	1	36
C21orf33	1	1	1	1	1	1	1	1	1	0	0	1	35
MCM3AP	1	1	1	1	1	1	1	1	1	0	0	1	34
STCH	1	1	1	1	1	1	1	1	1	0	0	1	34
DIP2A	1	1	1	1	1	1	1	1	1	0	0	1	31
C21orf57	1	1	1	1	1	1	1	1	1	0	0	1	27
C21orf91	1	1	1	1	1	1	1	1	1	0	0	1	26
C21orf67	1	1	1	1	1	1	1	1	1	0	0	1	7
CCT8	1	1	1	1	1	1	1	1	1	0	0	0	43
BIRC2	1	1	1	1	1	1	1	1	1	0	0	0	38
CBR1	1	1	1	1	1	1	1	1	1	0	0	0	37
BIRC5	1	1	1	1	1	1	1	1	1	0	0	0	36
PWP2	1	1	1	1	1	1	1	1	1	0	0	0	36
RIPK2	1	1	1	1	1	1	1	1	1	0	0	0	36
PCNT	1	1	1	1	1	1	1	1	1	0	0	0	33
BTG3	1	1	1	1	1	1	1	1	1	0	0	0	32
TIAM1	1	1	1	1	1	1	1	1	1	0	0	0	30
BACH1	1	1	1	1	1	1	1	1	1	0	0	0	28
C21orf58	1	1	1	1	1	1	1	1	1	0	0	0	28
PIGP	1	1	1	1	1	1	1	1	1	0	0	0	28
SETD4	1	1	1	1	1	1	1	1	1	0	0	0	28
DONSON	1	1	1	1	1	1	1	1	1	0	0	0	26
C21orf70	1	1	1	1	1	1	1	1	1	0	0	0	21
C21orf119	1	1	1	1	1	1	1	1	1	0	0	0	15
C21orf55	1	1	1	1	1	1	1	1	1	0	0	0	9
ATP5O	1	1	1	1	1	1	1	1	0	0	0	1	42
NDUFV3	1	1	1	1	1	1	1	1	0	0	0	1	33
OLIG2	1	1	1	1	1	1	1	0	1	1	0	0	5
RIPK4	1	1	1	1	1	1	1	0	1	0	0	1	26
CXADR	1	1	1	1	1	1	1	0	1	0	1	1	33
IFNAR2	1	1	1	1	1	1	1	0	1	0	0	1	27
OLIG1	1	1	1	1	1	1	1	0	1	0	0	1	4
PTTG1IP	1	1	1	1	1	1	1	0	1	0	0	0	45
CBS	1	1	1	1	1	1	1	0	1	0	0	0	32
DTNBP1	1	1	1	1	1	1	1	0	1	0	0	0	32
TRADD	1	1	1	1	1	1	1	0	1	0	0	0	30
PRDM15	1	1	1	1	1	1	1	0	1	0	0	0	24
SIM2	1	1	1	1	1	1	1	0	1	0	0	0	11
ARC	1	1	1	1	1	1	1	0	1	0	0	0	9
CARD8	1	1	1	1	1	1	1	0	0	0	0	1	32
C21orf86	1	1	1	1	1	1	1	0	0	0	0	1	14
C21orf56	1	1	1	1	1	1	1	0	0	0	0	0	24
LRRC3	1	1	1	1	1	1	1	0	0	0	0	0	3
IFIH1	1	1	1	1	1	1	0	0	1	0	1	0	31
TLR5	1	1	1	1	1	1	0	0	1	0	0	1	14
NLRP2	1	1	1	1	1	1	0	0	1	0	0	0	26
RSPH1	1	1	1	1	1	1	0	0	1	0	0	0	11
NLRP4	1	1	1	1	1	1	0	0	1	0	0	0	7
DSCR8	1	1	1	1	1	1	0	0	0	0	0	1	6
DSCR4	1	1	1	1	1	1	0	0	0	0	0	1	3
POFUT2	1	1	1	1	1	0	1	1	1	0	0	1	32
C21orf115	1	1	1	1	1	0	0	0	0	0	0	1	n/a
COL6A2	1	1	1	1	0	1	1	0	1	1	0	0	35
BIRC3	1	1	1	0	0	0	0	0	1	0	0	0	30
RCAN1	1	0	1	1	1	1	1	0	1	0	0	1	41
ATP5J	1	0	1	1	1	1	1	0	1	0	0	0	39
CRADD	1	0	1	1	1	1	1	0	1	0	0	0	25
C21orf13	1	0	1	1	1	1	0	1	1	0	1	1	17
DSCR2	1	0	1	1	0	0	1	1	1	0	0	0	30
TLR1	1	0	1	0	0	0	0	0	0	0	0	1	21
C21orf66	1	0	0	1	1	1	1	1	1	0	0	1	32
KRTAP19-1	1	0	0	1	1	1	1	0	0	0	0	0	1
AR	1	0	0	1	0	1	0	0	1	0	0	0	10

KRTAP21-2	1	0	0	1	0	0	0	0	0	0	0	0	n/a
C21orf45	0	1	1	1	1	1	1	1	1	0	0	0	32
USP25	0	1	1	1	1	0	1	1	1	0	0	0	33
IFNGR2	0	1	1	1	1	0	1	0	1	0	0	1	36
SLC19A1	0	1	1	1	1	0	1	0	1	0	0	1	31
ETS2	0	1	1	1	1	0	1	0	1	0	0	0	42
C21orf87	0	1	1	1	1	0	0	0	1	0	0	1	5
KRTAP11-1	0	1	1	1	1	0	0	0	0	1	0	0	n/a
CHAF1B	0	1	1	1	0	0	1	0	1	0	0	1	27
AIRE	0	1	1	1	0	0	0	0	1	0	1	0	2
GRIK1	0	1	1	1	0	0	0	0	1	0	0	1	10
KCNJ6	0	1	1	1	0	0	0	0	1	0	0	1	5
C21orf25	0	1	1	1	0	0	0	0	1	0	0	0	36
C21orf88	0	1	1	1	0	0	0	0	1	0	0	0	4
MX2	0	1	1	1	0	0	0	0	0	0	0	1	30
HLCS	0	1	1	0	1	0	1	1	1	0	0	0	34
CBR3	0	1	1	0	1	0	1	0	1	0	0	1	19
CLDN14	0	1	1	0	1	0	0	0	0	0	0	1	3
C21orf2	0	1	1	0	0	0	1	1	1	0	0	1	27
RRP1B	0	1	1	0	0	0	1	1	1	0	0	0	30
BACE2	0	1	1	0	0	0	1	0	1	0	0	0	38
CARD11	0	1	1	0	0	0	0	0	1	0	0	0	18
PCP4	0	1	1	0	0	0	0	0	0	0	0	0	12
C21orf29	0	1	1	0	0	0	0	0	0	0	0	0	2
C21orf30	0	1	0	1	1	0	1	0	0	0	0	1	8
ADAMTS5	0	1	0	1	1	0	0	0	1	0	0	0	23
TFF1	0	1	0	1	1	0	0	0	0	0	1	1	10
B3GALT5	0	1	0	1	1	0	0	0	0	0	1	0	7
H2BFS	0	1	0	1	1	0	0	0	0	0	0	0	2
HSF2BP	0	1	0	1	0	0	1	1	1	0	0	1	10
SUMO3	0	1	0	1	0	0	1	0	1	0	1	1	41
RBM11	0	1	0	1	0	0	1	0	1	0	0	0	8
PRMT2	0	1	0	1	0	0	1	0	0	0	0	1	42
FTCD	0	1	0	1	0	0	0	0	0	1	0	0	9
PRED58	0	1	0	1	0	0	0	0	0	0	0	1	n/a
C21orf123	0	1	0	1	0	0	0	0	0	0	0	1	1
C21orf125	0	1	0	1	0	0	0	0	0	0	0	0	9
C21orf129	0	1	0	1	0	0	0	0	0	0	0	0	3
MRAP	0	1	0	1	0	0	0	0	0	0	0	0	0
WRB	0	1	0	0	1	0	1	1	1	1	0	0	35
C21orf19	0	1	0	0	1	0	1	0	0	0	0	0	n/a
DOPEY2	0	1	0	0	1	0	1	0	0	0	0	0	25
TFF3	0	1	0	0	1	0	0	0	0	0	0	0	21
TMPRSS3	0	1	0	0	1	0	0	0	0	0	0	0	15
IL10RB	0	1	0	0	0	0	1	1	1	0	0	1	37
PFKL	0	1	0	0	0	0	1	0	1	0	0	1	41
ABCG1	0	1	0	0	0	0	1	0	0	0	0	1	26
PCBP3	0	1	0	0	0	0	1	0	0	0	0	0	17
C21orf99	0	1	0	0	0	0	1	0	0	0	0	0	2
CHODL	0	1	0	0	0	0	0	0	1	0	1	1	10
CLIC6	0	1	0	0	0	0	0	0	1	0	0	1	14
CYYR1	0	1	0	0	0	0	0	0	1	0	0	0	26
FAM3B	0	1	0	0	0	0	0	0	1	0	0	0	15
DSCR6	0	1	0	0	0	0	0	0	1	0	0	0	1
C21orf42	0	1	0	0	0	0	0	0	0	0	1	0	3
C21orf77	0	1	0	0	0	0	0	0	0	0	1	0	3
C21orf82	0	1	0	0	0	0	0	0	0	0	0	1	n/a
KRTAP19-6	0	1	0	0	0	0	0	0	0	0	0	1	n/a
TRPM2	0	1	0	0	0	0	0	0	0	0	0	1	22
S100B	0	1	0	0	0	0	0	0	0	0	0	1	18
C21orf84	0	1	0	0	0	0	0	0	0	0	0	1	6
C21orf32	0	1	0	0	0	0	0	0	0	0	0	1	4
C21orf116	0	1	0	0	0	0	0	0	0	0	0	0	n/a
FGF7P2	0	1	0	0	0	0	0	0	0	0	0	0	n/a
KRTAP19-7	0	1	0	0	0	0	0	0	0	0	0	0	n/a
ITGB2	0	1	0	0	0	0	0	0	0	0	0	0	35
CRYAA	0	1	0	0	0	0	0	0	0	0	0	0	7
DSCR9	0	1	0	0	0	0	0	0	0	0	0	0	6
TFF2	0	1	0	0	0	0	0	0	0	0	0	0	6
C21orf74	0	1	0	0	0	0	0	0	0	0	0	0	3
DNMT3L	0	1	0	0	0	0	0	0	0	0	0	0	2
C21orf100	0	1	0	0	0	0	0	0	0	0	0	0	1
C21orf128	0	1	0	0	0	0	0	0	0	0	0	0	1
DSCR10	0	1	0	0	0	0	0	0	0	0	0	0	0
KRTAP13-4	0	0	1	1	0	1	0	0	0	0	0	0	n/a
RWDD2B	0	0	1	1	0	0	1	1	1	0	0	1	36
MRPL39	0	0	1	1	0	0	1	1	1	0	0	1	34
RUNX1	0	0	1	1	0	0	1	0	0	1	0	1	33
KRTAP19-5	0	0	1	1	0	0	0	0	0	1	0	0	n/a
KRTAP13-2	0	0	1	0	0	0	0	0	0	0	0	0	1
GABPA	0	0	0	1	1	0	1	0	1	0	0	1	27
UBASH3A	0	0	0	1	1	0	0	0	0	0	0	0	11
C21orf93	0	0	0	1	0	0	0	0	0	0	0	1	4
MX1	0	0	0	1	0	0	0	0	0	0	0	0	39
KCNE1	0	0	0	1	0	0	0	0	0	0	0	0	12

C21orf120	0	0	0	1	0	0	0	0	0	0	0	0	n/a
HEMK2	0	0	0	0	0	0	1	1	1	1	0	0	18
PDXK	0	0	0	0	0	0	1	1	1	0	0	0	36
TWF2	0	0	0	0	0	0	1	1	1	0	0	0	33
NLRX1	0	0	0	0	0	0	1	1	1	0	0	0	27
BCL10	0	0	0	0	0	0	1	0	1	0	0	0	32
JAM2	0	0	0	0	0	0	1	0	1	0	0	0	19
SH3BGR	0	0	0	0	0	0	1	0	0	0	1	0	18
NRIP1	0	0	0	0	0	0	1	0	0	0	0	1	34
PPIAL3	0	0	0	0	0	0	1	0	0	0	0	1	n/a
MRPS6	0	0	0	0	0	0	1	0	0	0	0	0	42
RIPK1	0	0	0	0	0	0	1	0	0	0	0	0	33
NCAM2	0	0	0	0	0	0	1	0	0	0	0	0	13
C21orf90	0	0	0	0	0	0	1	0	0	0	0	0	4
NRG1	0	0	0	0	0	0	0	0	1	0	1	0	20
ERG	0	0	0	0	0	0	0	0	1	0	0	1	26
CARD14	0	0	0	0	0	0	0	0	1	0	0	1	15
PYCARD	0	0	0	0	0	0	0	0	1	0	0	0	23
C21orf121	0	0	0	0	0	0	0	0	1	0	0	0	3
PGLYRP1	0	0	0	0	0	0	0	0	1	0	0	0	3
PYDC1	0	0	0	0	0	0	0	0	1	0	0	0	3
CASP1	0	0	0	0	0	0	0	0	0	1	0	0	30
TLR2	0	0	0	0	0	0	0	0	0	1	0	0	21
TLR7	0	0	0	0	0	0	0	0	0	1	0	0	10
TLR6	0	0	0	0	0	0	0	0	0	1	0	0	7
DEFA5	0	0	0	0	0	0	0	0	0	1	0	0	3
DEFB127	0	0	0	0	0	0	0	0	0	1	0	0	1
PGLYRP3	0	0	0	0	0	0	0	0	0	1	0	0	1
DEFB4	0	0	0	0	0	0	0	0	0	1	0	0	n/a
KRTAP13-1	0	0	0	0	0	0	0	0	0	1	0	0	n/a
KRTAP20-1	0	0	0	0	0	0	0	0	0	1	0	0	n/a
KRTAP20-2	0	0	0	0	0	0	0	0	0	1	0	0	n/a
KRTAP6-1	0	0	0	0	0	0	0	0	0	1	0	0	n/a
KRTAP6-2	0	0	0	0	0	0	0	0	0	1	0	0	n/a
NLRP3	0	0	0	0	0	0	0	0	0	0	1	1	10
TLR4	0	0	0	0	0	0	0	0	0	0	1	0	24
NLRC4	0	0	0	0	0	0	0	0	0	0	1	0	12
NLRP1	0	0	0	0	0	0	0	0	0	0	0	1	27
TLR3	0	0	0	0	0	0	0	0	0	0	0	1	23
C21orf34	0	0	0	0	0	0	0	0	0	0	0	1	21
TLR8	0	0	0	0	0	0	0	0	0	0	0	1	12
C21orf62	0	0	0	0	0	0	0	0	0	0	0	1	9
C21orf109	0	0	0	0	0	0	0	0	0	0	0	1	5
DEFA4	0	0	0	0	0	0	0	0	0	0	0	1	4
KCNE2	0	0	0	0	0	0	0	0	0	0	0	1	4
NLRP14	0	0	0	0	0	0	0	0	0	0	0	1	3
ABCC13	0	0	0	0	0	0	0	0	0	0	0	1	2
DEFB125	0	0	0	0	0	0	0	0	0	0	0	1	1
DEFB128	0	0	0	0	0	0	0	0	0	0	0	1	n/a
KRTAP15-1	0	0	0	0	0	0	0	0	0	0	0	1	n/a
RIPK3	0	0	0	0	0	0	0	0	0	0	0	0	29
SAMSN1	0	0	0	0	0	0	0	0	0	0	0	0	25
CIITA	0	0	0	0	0	0	0	0	0	0	0	0	19
CARD6	0	0	0	0	0	0	0	0	0	0	0	0	16
NLRC3	0	0	0	0	0	0	0	0	0	0	0	0	16
C21orf15	0	0	0	0	0	0	0	0	0	0	0	0	10
NLRP12	0	0	0	0	0	0	0	0	0	0	0	0	10
CLDN8	0	0	0	0	0	0	0	0	0	0	0	0	9
DEFB1	0	0	0	0	0	0	0	0	0	0	0	0	9
C21orf41	0	0	0	0	0	0	0	0	0	0	0	0	7
IGSF5	0	0	0	0	0	0	0	0	0	0	0	0	7
NLRP6	0	0	0	0	0	0	0	0	0	0	0	0	7
NLRP7	0	0	0	0	0	0	0	0	0	0	0	0	7
PGLYRP2	0	0	0	0	0	0	0	0	0	0	0	0	7
DEFB119	0	0	0	0	0	0	0	0	0	0	0	0	6
NLRP11	0	0	0	0	0	0	0	0	0	0	0	0	6
DSCAM	0	0	0	0	0	0	0	0	0	0	0	0	5
PRSS7	0	0	0	0	0	0	0	0	0	0	0	0	5
C21orf110	0	0	0	0	0	0	0	0	0	0	0	0	4
LIPI	0	0	0	0	0	0	0	0	0	0	0	0	4
PGLYRP4	0	0	0	0	0	0	0	0	0	0	0	0	4
C21orf9	0	0	0	0	0	0	0	0	0	0	0	0	3
CARD18	0	0	0	0	0	0	0	0	0	0	0	0	3
NLRP9	0	0	0	0	0	0	0	0	0	0	0	0	3
C21orf94	0	0	0	0	0	0	0	0	0	0	0	0	2
DEFA6	0	0	0	0	0	0	0	0	0	0	0	0	2
MEFV	0	0	0	0	0	0	0	0	0	0	0	0	2
NLRP10	0	0	0	0	0	0	0	0	0	0	0	0	2
TPH2	0	0	0	0	0	0	0	0	0	0	0	0	2
DEFB118	0	0	0	0	0	0	0	0	0	0	0	0	1
DEFB123	0	0	0	0	0	0	0	0	0	0	0	0	1
DEFB126	0	0	0	0	0	0	0	0	0	0	0	0	1
DEFB129	0	0	0	0	0	0	0	0	0	0	0	0	1
KRTAP19-3	0	0	0	0	0	0	0	0	0	0	0	0	1
NLRP5	0	0	0	0	0	0	0	0	0	0	0	0	1

DEFB106A	0	0	0	0	0	0	0	0	0	0	0	0	0
KRTAP8-1	0	0	0	0	0	0	0	0	0	0	0	0	0
C21orf114	0	0	0	0	0	0	0	0	0	0	0	0	n/a
CLDN17	0	0	0	0	0	0	0	0	0	0	0	0	n/a
DAOA	0	0	0	0	0	0	0	0	0	0	0	0	n/a
DEFB103A	0	0	0	0	0	0	0	0	0	0	0	0	n/a
DEFB104A	0	0	0	0	0	0	0	0	0	0	0	0	n/a
DEFB105A	0	0	0	0	0	0	0	0	0	0	0	0	n/a
KRTAP13-3	0	0	0	0	0	0	0	0	0	0	0	0	n/a
KRTAP19-2	0	0	0	0	0	0	0	0	0	0	0	0	n/a
KRTAP22-1	0	0	0	0	0	0	0	0	0	0	0	0	n/a
KRTAP23-1	0	0	0	0	0	0	0	0	0	0	0	0	n/a
NLRP13	0	0	0	0	0	0	0	0	0	0	0	0	n/a
NLRP8	0	0	0	0	0	0	0	0	0	0	0	0	n/a
PRED15	0	0	0	0	0	0	0	0	0	0	0	0	n/a

Each row in the panel represents one promoter. The columns from left to right, respectively, show HGNC symbol of promoters, promoter reporter gene activities for untreated and HEK293T cells treated with different external stimuli measured on transfected-cell arrays (standard growth condition, PMA treatment, TSA treatment, Serum depletion, ARS treatment and Serum depletion & ARS treatment), corresponding endogenous gene expression (gene expression), RNA polymerase IIA occupancy (Pol IIA), the presence of core promoter elements in the cloned fragments (CGI, TATA, INR and DPE) and the corresponding ESTs expression.

Value 1 represents active promoters driving reporter gene expression (green boxes), transcripts detected in HEK293T (purple boxes), Pol IIa binding regions identified in promoter fragments (blue boxes) and presence of CpG islands, TATA boxes, INR and DPE elements in promoter fragments (orange boxes). Value 0 represents silent promoters, no transcripts in HEK293T, no overlapping of Pol IIa site in promoter fragments and absence of promoter elements in cloned fragments. Numbers shown in the last column indicate the number of different tissues (out of 45) in which corresponding ESTs are present.

Table 12. Summary of promoter activities for 62 fragments with length of 500 bp under standard growth condition and after treatments.

HGNC Symbol	Standard growth condition	PMA treatment	TSA treatment	Serum depletion	ARS treatment	Serum depletion & ARS treatment
ABCG1	0	1	0	0	0	0
ADAMTS1	1	1	1	1	1	1
ADAMTS5	0	1	0	0	0	0
ATP5J	1	1	0	1	1	1
BACE2	0	1	0	0	0	0
C21orf119	1	1	1	1	1	1
C21orf123	0	1	0	0	0	0
C21orf125	0	0	0	0	0	0
C21orf129	0	0	0	0	0	0
C21orf13	1	1	1	1	1	1
C21orf30	0	1	0	0	0	0
C21orf42	0	0	0	0	0	0
C21orf45	1	1	1	1	0	1
C21orf55	1	1	1	1	1	1
C21orf62	0	0	0	0	0	0
C21orf66	1	1	1	1	1	1
C21orf77	0	1	0	0	0	0
C21orf88	0	0	0	0	0	0
C21orf91	1	1	1	1	1	1
CBR1	1	1	1	1	1	1
CBR3	0	1	0	0	0	0
CCT8	1	1	1	1	1	1
CHAF1B	1	1	1	1	1	1
CHODL	0	1	0	0	0	0
CLDN17	0	0	0	0	0	0
CLDN8	0	0	0	0	0	0
CLIC6	0	1	0	0	0	0
COL6A2	1	1	1	1	1	1
CXADR	1	1	1	1	1	1
CYYR1	0	1	0	0	0	0
DSCR2	0	1	0	0	0	0
DSCR9	0	0	0	0	0	0
FAM3B	0	1	0	0	0	0
GART	1	1	1	1	1	1
HEMK2	0	1	0	0	0	0
HLCS	1	1	1	1	0	1
JAM2	0	1	0	0	0	0
KCNJ6	0	1	1	1	0	0
LIPI	0	0	0	0	0	0
LRRC3	1	1	1	1	1	1
MRPL39	1	1	1	1	1	1
MX2	0	0	0	0	0	0
OLIG1	0	1	0	0	0	0
OLIG2	1	1	1	1	1	1
PCP4	0	1	1	0	0	0
PRSS7	0	0	0	0	0	0
PTTG1IP	1	1	1	1	0	0

PWP2H	1	1	1	1	1	1
RBM11	1	1	1	1	0	1
RCAN1	1	1	1	1	1	1
RWDD2B	0	0	0	0	0	0
SETD4	1	1	1	1	1	1
SH3BGR	1	1	1	1	0	1
SIM2	0	0	0	0	0	0
SLC19A1	0	1	0	0	0	0
SOD1	1	1	1	1	1	1
SON	1	1	1	1	1	1
STCH	1	1	1	1	1	1
SYNJ1	1	1	1	1	1	1
TTC3	1	1	1	1	1	1
USP25	0	1	1	1	0	0
WRB	0	0	0	0	0	0

The panels represent the activities of 0.5 kb promoter fragments for untreated and HEK293T cells treated with different external stimuli measured on transfected-cell arrays. Each row in the panel represents one promoter. The columns indicate the growth conditions. Value 1 and 0 represent active promoters (green boxes) and silent promoters, respectively.

ACKNOWLEDGEMENTS

In the first place I would like to thank Prof. Dr. Hans Lehrach for providing me the opportunity to carry out my Ph.D. program in his department at the Max-Planck-Institute for Molecular Genetics.

I would like to express my great appreciation to Prof. Dr. Gerd Multhaup for being my second supervisor and for reviewing my thesis as representative of the Free University Berlin.

I am exceptionally grateful to Dr. Michal Janitz, my group leader, for inviting me to be here, for all his supervision, discussion and support on my Ph.D. program during the last four years, as well as for his correction of this dissertation and the translation of the Zusammenfassung.

I would like to thank all the collaborators involved in this project. Many thanks to Dr. Anna Guerasimova for the stimulating discussions, valuable suggestions and her all-embracing inspiration at any step of my work, to Dr. Robert Querfurth, Dr. Hans-Jörg Warnatz from the group of Dr. Marie-Laure Yaspo for primers design, promoter cloning and preparation of the endogenous transcript and Pol IIa data, to Dr. Stefan Haas and Dr. Thomas Manke from the group of Dr. Martin Vingron for the promoter prediction, PCR primer design and the sequence analysis of cloned fragments and helpful discussions, to Dr. Wilfried Nietfeld for reagent supply and inspiring discussions, to Prof. Dr. med. Philip Rosenstiel from Christian-Albrechts University Kiel for providing the information of inflammation genes. Special thanks go to Dr. Wilfried Nietfeld and Dr. Robert Querfurth for their correction and critical comments on my thesis.

I would like to thank my colleagues for the friendly atmosphere. Many thanks to Dr. Andrea Fiebitz and Dr. Dominique Vanhecke for their technical supports on cell arrays, to Sabine Thamm for her daily support and kindness for my numerous requests for German translation during my time in the lab as well as a lot of very generous help in my private life, to Irina Girnus and Kerstin Zurth for their assistance during the DNA cloning, to Dr. Andreas Dahl, Dr. Yuhui Hu, Dr. Jianping Liu and many people from the Institute and beyond for their technical support, reagent and material supply and inspiring discussions, to Prof. Dr. Manfred Schweiger and Prof. Dr. Monica Hirsch-Kauffmann and a lot of people in Fabeckstrasse for many scientific and non-scientific help and discussions.

I acknowledge the M2-automation and Olympus for their professional technical support on microarray printing and microscopy.

Particular thanks go to Dr. Ulrich Barenbrock, Dr. Cornelia Platzer, Anna Skotarczyk and Kathleen Müller for their administrative helps on my life here as a foreigner.

I would additionally like to thank my friends wherever in the world for their concerns and encouragement during these years.

Last, but above all, I owe my deepest gratitude to my parents in China for their everlasting love and care. Over these years I could not visit them as often as I wished and should. But no matter where I am, I always know that they are right there supporting me, proud of me and believing in me.

CURRICULUM VITAE

For reasons of data protection,
the curriculum vitae is not included in the online version

For reasons of data protection,
the curriculum vitae is not included in the online version

SELBSTÄNDIGKEITSERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Arbeit eigenständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Ich versichere, dass diese Arbeit in dieser oder anderer Form noch keiner anderen Prüfungsbehörde vorgelegt wurde.

Hiermit erkläre ich, dass ich die Arbeit selbst verfasst habe sowie keine anderen als die angegebenen Quellen und Hilfsmittel in Anspruch genommen habe. Ich versichere, dass diese Arbeit in dieser oder anderer Form keiner anderen Prüfungsbehörde vorgelegt wurde.

Berlin, den

_____

(Xi Cheng)