CHARACTERIZATION OF GENE EXPRESSION
AND TRANSCRIPTION KINETICS UNDER THE
REGULATION OF MODIFIED CORE
PROMOTERS

Dan Even

Submitted in partial fulfillment of the requirements for the Master's
Degree in the Faculty of Life Sciences, Bar-Ilan University

Ramat Gan, Israel 2014
This work was carried out under the supervision of Dr. Tamar Juven-Gershon, from the Mina and Everard Goodman.

Faculty of Life Sciences, Bar Ilan University
Acknowledgments

A major research project like this is never the work of anyone alone. The contributions of many different people, in their different ways, have made this possible. I would like to extend my appreciation especially to the following.

First and foremost, I would like to express my sincere gratitude to my advisor Dr. Tamar Juven-Gershon, who has supported me throughout my thesis with her patience and knowledge whilst allowing me the room to work in my own way. I appreciate her continuous support, her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. One simply could not wish for a better or friendlier supervisor.

A special thanks is reserved to Dr. Diana Ideses, our lab manager, for her kindness, encouragement, insightful comments, and hard questions. Her help and friendship are invaluable to me.

In my daily work I have been blessed with a friendly, cheerful and very much helpful group of fellow labmates: Yonathan Zehavi, Avital Ovadia, Adi Kedmi, Hila Shir, Miri Pinhasov, Anna Sloutskin, and the former lab members Julia Sharabany and Ola Kuznetsova. Particularly, I would like to especially thank Yahuda Danino for the many valuable discussions, support and help. I would also like to thank Shani Basch- Barzilay for working close to me on my first research period. I wish everyone the best in their future research and work.

I am also grateful to Dr. Sharon Yunger and Pinhas Kafri, for spending their valuable time and efforts helping me to progress with my research.

Finally, none of this would have been possible without the love and patience of my family. My immediate family, has been a constant source of love, concern, support and strength all these years.
# Table of Contents

Abstract .............................................................................................................................................. 1  
Introduction ....................................................................................................................................... 3  
Core promoter .................................................................................................................................... 3  
Core promoter elements ....................................................................................................................... 4  
SCP - Super Core Promoter .................................................................................................................. 7  
Research importance ............................................................................................................................ 10  
Research Goals ................................................................................................................................ 12  
Results ................................................................................................................................................ 13  
Characterization of Gene Expression Under the Regulation of Modified Core Promoters 13
  pRC/CMV based constructs ................................................................................................................... 13
  pEGFP-N1 based constructs ................................................................................................................. 22
The Mechanism of SCPs Action ............................................................................................................... 28
  Examination whether DNA methylation is involved in the potent activity of the SCPs ...................... 28
  Examination whether histone acetylation is involved in the potent activity of SCPs ......................... 30
  Examination whether synergism between DNA methylation and histone acetylation is involved in the potent activity of the SCPs ................................................................. 33
  The involvement of TAF1 in regulating the activity of SCPs .............................................................. 36
Characterization of the Influence of Core Promoter Elements on Transcription Activity and Kinetics................................................................................................................................. 40
Abstract

The transcription of protein-coding genes is a pivotal process underlying proper cellular function. The accurate initiation of transcription by RNA polymerase II (Pol II) is a critical step in the regulation of gene expression, in which Pol II is recruited to the core promoter via the basal transcription machinery. The core promoter, which plays a key role in the regulation of gene expression, is typically 80 nucleotides long, encompassing from -40 to +40 relative to the transcription start site. The core promoter consists of several functional subregions, termed core promoter elements or motifs (e.g. the TATA-box, initiator and downstream core promoter element (DPE)), which confer specific functional properties to each promoter.

Earlier studies that examined the ability to modulate gene expression levels via the core promoter, led to the design of strong synthetic core promoters, termed super core promoters (SCPs), which combine different core promoter elements into a single core promoter. Based on both viral and cellular promoters, a variety of modified core promoters were engineered, including the novel super core promoter, termed SCP3. Here, we characterized the gene expression driven by these modified core promoters. Our results provide additional support for the importance of core promoter elements that are located downstream of the initiator in the regulation of transcription, and demonstrate for the first time, robust and long-term gene expression in transiently transfected cells.

With the aim of understanding the mechanism underlying the potent transcriptional activity of these modified core promoters, we examined the possible involvement of DNA methylation, histone acetylation and the TAF1 basal transcription factor. Our data suggests that the architecture of the core promoters, and particularly the sequence downstream of the transcription start site, is a vital factor that dictates promoter identity, such as its dependence or independence on a specific transcription factor, as well as the promoter’s mode of action in response to various stimuli.

The most exciting part of this work in my opinion is the construction and utilization of an in vivo imaging system, based on the system developed by Prof. Yaron Shav-Tal and
Prof. Yuval Garini. This in vivo system provides the ability to detect and analyze real time transcriptional activity and kinetics of a gene, at single-cell, single-gene and single-allele levels. The strength of this system lies in the fact that it eliminates position effects, hence all changes in activity and kinetics only result from the predetermined investigated variable. Here, we present the construction of the system with modified core promoters as the predetermined variable. The use of these modified core promoters enables us to investigate the individual contribution of each core promoter element as well as the connections and associations among the elements when they are combined in a single promoter. The data obtained so far indicates substantial differences between the modified core promoters, which as explained earlier, can only result of the investigated individual or associated contribution of each core promoter element.

To conclude, the architecture of core promoters is vital for proper transcriptional regulation, as the composition of core promoter elements dictates the promoter’s identity and mode of action. Each individual core promoter element confers specific functional properties to the promoter. However, since core promoters are comprised of combinations of different core elements, these elements interact and function in a different conjunct manner. The characterization of the individual, as well as multiple core promoter elements, will greatly contribute to the knowledge of various aspects of transcription and to the understanding of strong synthetic modified core promoters and their applications, such as the described non-viral approach for long-term gene expression in transiently transfected cells.
**Introduction**

A pivotal process underlying proper cellular functions, such as homeostasis, growth, differentiation and development, is the transcription of protein-coding genes\(^1\). The accurate initiation of transcription by RNA polymerase II (Pol II) is a critical step in the regulation of gene expression, in which Pol II is recruited to the core promoter via the basal transcription machinery\(^2\). This recruitment of Pol II is preceded by multiple events that include chromatin decondensation, nucleosome remodeling, histone modifications and binding of transcriptional activators and coactivators to enhancers and promoters, which ultimately leads to the recruitment of the basal transcription machinery and the assembly of the transcription preinitiation complex (PIC)\(^2\). The basal transcription machinery is comprised of many protein complexes such as the TFIID complex, which contains the TBP (TATA-box Binding Protein) and 13 TBP - associated factors (TAFs) that can recognize sequential elements in the core promoter\(^3\text{-}^5\). The protein complexes of the basal transcription machinery recognize and bind the core promoter and thus recruit Pol II to the transcription start region\(^2\).

**Core promoter**

The core promoter is generally defined as the minimal DNA region required for basal transcription and for accurate initiation of transcription by RNA polymerase II. The core promoter is typically 80 nucleotides long, encompassing the region from -40 to +40 relative to the transcription start site. In the past, it was presumed that all core promoters function via a single universal mechanism, however, it is now well established that core promoters differ in both structure and function \(^5\text{-}^12\). Initiation of transcription can occur in two different patterns - focused and dispersed \(^7\text{-}^8,^{11-14}\) (Figure 1):

**Focused initiation:** The transcription begins at a single nucleotide or at a cluster of several nucleotides within a narrow region. This pattern is the predominant mode of transcription in simple organisms, however it can also occur in developed organisms. Focused initiation is
usually associated with regulated genes, and one can postulate that it is easier to regulate transcription of genes with a single start site.

**Dispersed initiation:** The transcription begins from multiple weak start sites scattered over a broad region of about 50 to 100 nucleotides. In vertebrates, about 70% of the genes have dispersed promoters, typically found in CpG islands. Dispersed promoters usually regulate constitutive genes (such as housekeeping genes), as the use of multiple start sites can minimize variations of expression.

![Figure 1. Schematic representation of focused and dispersed transcription initiation strategies](image)

Although dispersed promoters comprise the majority of promoters in vertebrates, most studies of the RNA Pol II transcription were performed with focused promoters. Multiple studies analyzed focused promoters due to the biological significance of many genes that are regulated by the focused promoters.

**Core promoter elements**

The core promoter consists of several functional subregions, termed core elements or motifs. Some of the known core promoter elements are the TATA box, TFIIB recognition elements (BREu and BReD), DCE, initiator (Inr), TCT, MTE, DPE and XCPE\(^{15-27}\) (Figure 2). There is no universal core promoter and the different combinations of core promoter motifs confer specific functional properties. Moreover, there are core promoters that lack any of the known core motifs, suggesting the existence of additional unknown core promoter elements that remain to be discovered.
This work puts a heavy emphasis on four core promoter elements:

- **initiator (Inr):**
  The Inr motif is the most common element, and it encompasses the transcription start site, which is designated as the A nucleotide in the Inr consensus (usually designed as position “+1”)\(^27-29\). Some other core promoter motifs, such as the MTE and DPE, function in a manner that depends on their exact spacing from the Inr\(^12\). There are several factors that can interact with the Inr motif, however the binding of TFIID (via TAF1 and TAF2) is particularly important since it correlates best with the Inr activity\(^7\). The human Inr consensus is YYANWYY and the *Drosophila* consensus is TCAKTY.

- **TATA box:**
  The TATA box motif (also named the Goldberg-Hogness box after its discoverers) is the first element that was discovered as well as the best-known element\(^7,11,12,20\). Despite the extensive knowledge gained with regards to the TATA box over the years, it is only present in about 10-15% of the mammalian core promoters. Only about one-third of vertebrate core promoters are focused and only some of them contain the TATA box motif\(^11\). This motif is conserved from archaeabacteria to humans\(^30\). The TATA box is an A/T-rich sequence with a consensus of TATAWAAR, where the upstream T nucleotide
is located at −31 or −30 relative to the A+1 position in the Inr. In eukaryotes, the TATA box is recognized and bound by the TBP (TATA box-binding protein) subunit of the TFIID complex.2,7,11,12,31.

- **DPE (Downstream Core Promoter Element):**

  The DPE motif was identified as a downstream TFIID recognition site that is important for basal transcription activity. This motif is conserved from *Drosophila* to humans, and DPE-dependent promoters are typically (but not exclusively) found in TATA-less promoters. The *Drosophila* DPE consensus is RGWYVT and the human consensus is yet to be determined. Nevertheless, several mammalian promoters that contain sequences that match the *Drosophila* consensus, have been shown to possess DPE activity. The DPE, which is located precisely from +28 to +33 relative to the A+1 position in the Inr, serves as a recognition site for the TAF6 and TAF9 subunits of TFIID and functions in a cooperative manner with it. A precise spacing between the Inr and DPE motifs is crucial for transcriptional activity, thus even an increase or decrease of a single nucleotide or three nucleotides in the spacing between the Inr and the DPE, results in a major decrease of transcriptional activity.3,4,18,33.

- **MTE (Motif Ten Element):**

  The MTE motif, similarly to the DPE, is located downstream of the Inr motif and servers as recognition site for TFIID. The MTE is located immediately upstream of the DPE at precisely +18 to +27 relative to the A+1 position in the Inr, and its exact spacing from the Inr motif is crucial since it functions cooperatively with the Inr. Furthermore, the MTE can function independently of the TATA box and DPE motifs, however its collaboration with the TATA box or the DPE, results in strong synergy. The MTE is conserved from *Drosophila* to humans and its consensus sequence is CSARCSSAAC.
**SCP - Super Core Promoter**

Previous studies of the MTE revealed that cooperative function of the MTE motif with the TATA box, as well as with the DPE, results in strong synergy\(^1\). This synergy raised the idea that by combining different core promoter elements into a single core promoter, it could be possible to design new unusually strong synthetic core promoters \(^{1,14,36}\). Consequently, synthetic core promoters designated Super Core Promoters, which contain the TATA box, Inr, MTE and DPE elements, were engineered by Dr. Tamar Juven-Gershon, and were shown to drive high levels of transcription both *in vivo* and *in vitro*\(^1\).

**SCP1**

SCP1 was the first Super Core Promoter to be rationally designed and engineered. It comprises the sequences from –36 to +45 relative to the A+1 position in the Inr. SCP1 contains the TATA box, Inr, MTE and DPE elements in a manner by which each of these four core motifs contributes to the full SCP1 activity. The BRE motif is not included in the SCP as it been reported to have both positive and negative effects on transcriptional activity. As the structure and function of core promoters, as well as basal transcription factors, are highly conserved from the fruit fly to humans, the SCP is designed with no distinction between the origin of the core promoter sequences, hence it is comprised of sequences derived from several different core promoters that exhibit strong activity *in vitro* (**Figure 3**). SCP1 is distinctly stronger than the cytomegalovirus (CMV) IE1 and adenovirus major late (AdML) core promoters, both *in vitro* and *in vivo*. There is substantially higher template usage with SCP1 (40%) than with the CMV (15%) or AdML (6%) core promoters\(^{14}\). Additionally, it is bound very efficiently by TFIID and exhibits high propensity to form productive transcription complexes. In addition to strong basal transcription activity, SCP1 also enhances activated transcription in conjunction with the SV40 enhancer or the CMV enhancer\(^{14}\).

![Figure 3. Schematic representation of SCP1](image-url)
**The natural CMV promoter & SCP2**

In order to explore the function of the SCPs in conjunction with the CMV enhancer, four parallel core promoter variations were constructed (Figure 4). All four core promoters are based on the commercial pRC/CMV vector (Life Technologies) that contains the CMV enhancer and promoter region, but lacks core promoter sequences that are downstream of -16 relative to the A+1 position, including the Inr motif (hence the pRC/CMV vector contains only the TATA box element). In order to explore the complete CMV core promoter (in conjunction to its enhancer), the promoter region of the commercial vector was replaced with the complete core promoter sequence of CMV IE1, from -36 to +45 relative to the A+1 position, thus creating the natural CMV construct that includes both the TATA box and Inr motifs (Figure 5). The third promoter construct is the above mentioned SCP1, which was cloned into the commercial vector instead of the incomplete CMV promoter. As the SCPs function in conjunction with the CMV enhancer was explored, the fourth construct, termed SCP2, which also encompasses the region from -36 to +45 relative to the A+1 position, was constructed (Figure 5). SCP2 is almost identical to the natural CMV apart from the sequence from +17 to +34, which was replaced with a *D. melanogaster Tollo* MTE and a human *Calm2 DPE*.

![Figure 4](image)

**Figure 4. Schematic representation of the four pRC/CMV-based core promoters**

Upon comparison between SCP1 and SCP2 it was found that SCP1 is more active in the absence of an enhancer or in conjunction to the SV40 enhancer, while SCP2 is more active than SCP1 in conjunction with the CMV enhancer. This may be explained by optimization of
function between enhancer and core promoter, as SCP2 contains more CMV sequences than SCP1, which might work in concert with the CMV enhancer14.

Figure 5. Schematic representation of the natural CMV and of SCP2

**SCP3**

SCP3 is a modified version of the previously engineered SCP2. The design of SCP3 was done by Dr. Tamar Juven-Gershon. Similarly to SCP2, SCP3 contains the CMV TATA and Inr motifs, along with the *D.melanogaster Tollo* MTE and the human *Calm2* DPE. Unlike SCP2, SCP3 does not contain the T7 promoter that is downstream of +45 relative to the +1 transcription start site (to position the reporter gene immediately downstream of the promoter) and includes four nucleotide changes (**Figure 6**):

1. To improve the Inr sequence (A to T in position +3).
2. To improve the Inr sequence (T to C in position +4).
3. To improve the sequence adjacent to the MTE (G to C in position +16; based on single nucleotide analysis of the *Tollo* MTE16).
4. To improve the DPE (T to C in position +31; as in many DPE driven genes as well as in the MTE-dependent *Tollo* core promoter15,16,19).
The study of SCP3, which is described in part of this work, has been submitted for publication to *Scientific Reports*, a journal of the Nature Publishing Group (Appendix 2). The submitted article will be the first publication of SCP3.

**Research importance**

SCPs are unusually strong core promoters that yield long-term expression and high levels of transcription. As such, the SCPs can be useful for applications that range from studies of transcription to increased gene expression and protein production in cells.

The high template usage achieved with the SCPs, as well as the high affinity of TFIID binding to the SCPs, make these promoters ideal for biophysical studies of TFIID binding to the DNA and for structural analyses of transcription complexes. An example is provided by a recent study performed by the Nogales lab\(^{37}\), which employed the high affinity of TFIID to the SCPs in order to investigate the conformational landscape of the human TFIID. This study discovered that the human basal transcription factor TFIID coexists in two distinct structural conformations, where the rearranged state is important for the assembly of the preinitiation complex. The authors have shown that the core promoter architecture dictates TFIIA-dependent and -independent interactions of TFIID with core promoter DNA.

SCPs can also be useful in the biotechnology industry and in molecular biology studies, as demonstrated in a previous study\(^{38}\) in which a modified core promoter (*Drosophila*-SCP that was designed based on the super core promoters) was used to identify different enhancers and the expression patterns they drive in small subsets of neurons, for manipulation and dissection of neural circuitry. Additionally, the long-term expression of genes that are
regulated by SCPs provide a non-viral way for long-term gene expression in transiently transfected cells. This can be relevant in cases where long-term follow-up is important, such as the response to drugs and cytokines.

Investigation of the mechanism that enable the strong and prolonged gene expression driven by the SCPs, may greatly contribute to the understanding of both the molecular processes underling transcriptional regulation and the factors that distinguish between different promoters.

Characterization of the Influence of core promoter elements on the levels of transcription activity as well as on the kinetics of transcription driven by modified core promoters, will allow us to investigate the individual contribution of each core promoter element as well as the connections and functional interactions among the elements when they are combined in a single promoter.
Research Goals

1. Characterization of gene expression under the regulation of modified core promoters:
   Identification and comparison of the transcriptional activity driven by SCP2, SCP3
   and other core promoters, such as the CMV IE1 promoter and its commercial
   versions.
   The characterization and comparison of activities is done with regards to both
duration and levels of expression.

2. Investigation of the mechanism that enables the strong and prolonged gene
   expression driven by SCP2 and SCP3:
   Examination of the possible involvement of post-translational modifications, as well
   as of basal transcription factors in the mechanism underling the potent transcriptional
   activity of the modified core promoters.

3. Characterization of the Influence of core promoter elements on transcriptional activity
   levels and transcription kinetics of the Cyclin D1 gene:
   • Analysis of transcription driven by viral promoters that are based on the CMV
     promoter (including SCP2 and SCP3).
   • Analysis of transcription driven by cellular promoters that are based on the
     Cyclin D1 promoter.
   This characterization is done by the utilization of an in vivo system for visualization
   and analysis of real time transcriptional activity and kinetics of a single-copy, single-
   allele gene.
Results

Characterization of Gene Expression Under the Regulation of Modified Core Promoters

pRC/CMV based constructs

A previous study\textsuperscript{14}, done by Dr. Juven-Gershon, showed the potential in combining different core promoter elements for the creation of unusually strong synthetic core promoters. In this study, synthetic core promoters based on the commercial pRC/CMV vector (Life Technologies) were engineered. The study utilized the chloramphenicol acetyltransferase (CAT) and luciferase reporter genes to assay the transcription activity of the promoters. Unlike these assays that require harvesting of the cells at a specific time point, in this work we used the \textit{EGFP} reporter gene. The \textit{EGFP} reporter gene enables us to qualitatively follow the activity of the various core promoters in the same population of transfected living cells over time, and to quantitatively analyze the activity of each promoter. In addition, by the use of the \textit{EGFP} reporter gene, we were able to add a new, temporal dimension to the characterization of the SCPs.

To this end, the firefly luciferase sequence was replaced by Ravid Tikotzki and Shani Basch-Barzilay (former M.Sc. students in our lab) by the \textit{EGFP} reporter gene sequence, cloned from the commercial pEGFP-N1 vector (Clonetech), thus creating a set of four new expression vector constructs. All of these constructs are pRC/CMV vector-based (which includes the CMV enhancer), each driving the expression of the \textit{EGFP} reporter gene (Figure 7):

- **pRC/CMV-\textit{EGFP}:** Commercial vector containing the CMV IE1 promoter. The core promoter contains the TATA box element, but lacks CMV sequences that are downstream of -16 relative to the +1 transcription start site, including the Inr element.
- **pRC/CMV-natural CMV-\textit{EGFP}:** The complete CMV promoter sequence (from -36 to +45) that contains both the TATA box and Inr elements.
• **pRC/CMV-SCP2-EGFP**: The core promoter contains the CMV TATA box and Inr elements, the Tollo MTE, and the calm2 DPE motif.

• **pRC/CMV-SCP3-EGFP**: The core promoter elements composition is the same as in SCP2, except for the absence of the T7 promoter and for four nucleotide changes (as described in the introduction).

![Diagram of promoter elements](image)

**Figure 7. Schematic representation of the four pRC/CMV-based core promoters**

<table>
<thead>
<tr>
<th></th>
<th>-36</th>
<th>+1</th>
<th>+45</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRC/CMV</td>
<td>TATA</td>
<td>Inr</td>
<td>EGFP</td>
</tr>
<tr>
<td>natural CMV</td>
<td>TATA</td>
<td>Inr</td>
<td>EGFP</td>
</tr>
<tr>
<td>SCP2</td>
<td>TATA</td>
<td>Inr</td>
<td>MTE DPE</td>
</tr>
<tr>
<td>SCP3</td>
<td>TATA</td>
<td>Inr</td>
<td>MTE DPE</td>
</tr>
</tbody>
</table>

Characterization and comparison of the transcription activity of the four promoters was done by analysis of the fluorescence intensity and the number of cells expressing EGFP, hence the level of transcription activity and the number of cells in which EGFP is expressed under the regulation of the different promoters. In order to refer to the changes in activity over a period of time, all experiments were analyzed on a daily basis for 8 days post-transfection (divided into short-term (d 1-4) and long-term (d 4-8)). Moreover, as short- and long-term experiments were performed separately, day 4 is common to both analyses, thus serving as a linkage between the experiments. In this section we used two types of human cell lines, HeLa S3 (human cervical carcinoma cells) and SH-SY5Y (human neuroblastoma cells originating from bone marrow tissue). The HeLa S3 cell line was chosen as a continuation to the previous study14, done by Dr. Juven-Gershon. The SH-SY5Y cell line was chosen based on previous experiments done by Ravid Tikotzki, who showed positive EGFP expression driven by the SCPs in this cell line (even after 5 days post-transfection).
Qualitative Live Cell Imaging

Live Cell Imaging enabled us to qualitatively observe the activity of the different core promoters in live cell population.

HeLa S3 and SH-SY5Y cells were transiently transfected with the pRC/CMV, natural CMV, SCP2 or SCP3-driven EGFP-expression vectors, and the fluorescence signals expressed by the transfected cell populations were imaged every 24 h. Between the imaging sessions the cells were incubated at 37°C, allowing them to proliferate.

Representative results of 3 independent experiments for short-term and 4 independent experiments for long-term expression are presented in Figures 8 and 9 (respectively).

Figure 8. Live cell EGFP imaging of short-term expression of pRC/CMV-based constructs, in HeLa S3 and SH-SY5Y cells.

HeLa S3 and SH-SY5Y cells were transiently transfected with either the pRC/CMV, natural CMV, SCP2 or SCP3 vector expressing EGFP. The cells were imaged once a day during 1-4 days post-transfection (P.T.). Each circle displays the whole well image constructed by stitching individual microscopic fields. (a) HeLa S3 cells. (b) SH-SY5Y cells. Data shown are representative of 3 independent experiments for each cell type.
Live cell imaging for short-term expression, performed during days 1-4 post-transfection, reveals that the full-length core promoters have a substantial advantage over the commercial pRC/CMV core promoter, both in terms of fluorescence intensity and number of fluorescent cells (Figure 8). During the 4 days following transfection, an increase in the fluorescence levels, thus the transcriptional activity, can be observed in all four promoters, though only slightly in the pRC/CMV promoter. Notably, SCP3 is consistently the most active core promoter.

Figure 9. Live cell **EGFP** imaging of long-term expression of pRC/CMV-based constructs, in HeLa S3 and SH-SY5Y cells.

HeLa S3 and SH-SY5Y cells were transiently transfected with either the pRC/CMV, natural CMV, SCP2 or SCP3 vector expressing **EGFP**. The cells were imaged once a day during 4-8 days post-transfection (P.T.). Each circle displays the whole well image constructed by stitching individual microscopic fields. (a) HeLa S3 cells. (b) SH-SY5Y cells. Data shown are representative of 4 independent experiments for each cell type.
Remarkably, the long-term expression imaging performed during days 4-8 post-transfection, exhibits a distinct increase in activity using the complete natural CMV, SCP2 and especially SCP3, as compared to the commercial pRC/CMV core promoter (Figure 9). As predicted for transient transfections, we observed a decrease in fluorescence intensity during the long-term. However, in days 7 and 8, there are still relatively high fluorescence intensities driven by the full-length core promoters, and particularly the SCP3 core promoter.

**Quantitative FACS analysis**

To investigate the transcription activity of the various core promoters in a quantitative manner, we performed FACS analysis. HeLa S3 and SH-SY5Y cells were transiently transfected with the pRC/CMV, natural CMV, SCP2 or SCP3-driven *EGFP*-expression vectors, and analyzed by FACS on a daily basis for 8 days post-transfection (divided into short-term (d 1-4) and long-term (d 4-8) follow-up). To quantitatively measure the reporter levels of transcription activity and the number of cells expressing the reporter under the regulation of the four core promoters, both the fluorescence intensity and the number of cells expressing *EGFP* were examined on a daily basis. As the fluorescence intensity and the number of *EGFP*-expressing cells differ between high fluorescence expressors and low fluorescence expressors, we analyzed the cells that express high fluorescence levels (designated “HIGH EXP”) in addition to analyzing the entire population of fluorescent cells (designated “ALL EXP”). A non-parametric Kruskal-Wallis test was performed to examine the statistical differences between any pair of promoters. A p-value ≤ 0.05 was considered to be statistically significant. The FACS analyses presented in Figures 10 and 11 depict representative experiments, whereas the statistical significance was calculated for 5-6 individual experiments for each cell line for short- as well as long-term expression.

Representative results, for short-term expression, of 5 independent experiments using HeLa S3 cells and 6 independent experiments using SH-SY5Y cells, are presented in Figure 10. Representative results, for short-term expression, of 6 independent experiments using HeLa S3 cells and 5 independent experiments using SH-SY5Y cells, are presented in Figure 11.
HeLa S3 and SH-SY5Y cells were transiently transfected with pRC/CMV, natural CMV, SCP2 or SCP3 vector expressing EGFP. The cells were collected 1-4 days post-transfection (P.T.) for FACS analysis. (a) FACS analysis of fluorescence intensity of all HeLa S3 fluorescent cells and SH-SY5Y fluorescent cells. (b) FACS analysis of fluorescence intensity of high intensity HeLa S3 fluorescent cells and SH-SY5Y fluorescent cells. (c) FACS analysis of the number of all HeLa S3 fluorescent cells and SH-SY5Y fluorescent cells. (d) FACS analysis of the number of high intensity HeLa S3 fluorescent cells and fluorescent SH-SY5Y cells. Data shown are representative of 5 independent experiments using HeLa S3 cells and 6 independent experiments using SH-SY5Y cells. Statistical comparisons between the promoters were done using the Kruskal–Wallis test with pairwise comparisons. A p-value ≤ 0.05 was considered to be statistically significant.
FACS analysis of short-term EGFP-expression reveals a constant advantage in favor of the SCP3 promoter (Figure 10):

This advantage is reflected in both HeLa S3 and SH-SY5Y cell lines, by the fluorescence intensity levels of all fluorescent cells. Analysis of the fluorescence intensity of all HeLa S3 fluorescent cells reveals that SCP3 is significantly stronger than the pRC/CMV promoter during all four days. Similarly, analysis of all EGFP-expressing SH-SY5Y cells, reveals that SCP3 is significantly stronger than pRC/CMV in days 1 and 2 post-transfection (Figure 10a). Overall, we did not observe significant difference in the fluorescence intensity driven by either promoter when we analyzed the high expressors in both cell lines (Figure 10b).

The advantage of SCP3 is also reflected in the number of all EGFP-expressing SH-SY5Y cells, where significant differences between SCP3 and pRC/CMV were observed during days 1, 2 and 4 post-transfection (Figure 10c). The analysis of the number of high fluorescence intensity HeLa S3 and SH-SY5Y cells, reveals significant advantage favoring SCP3 over the pRC/CMV promoter, in all four examined days (Figure 10d).

In addition to the advantage of SCP3, short-term analysis also reveals a significant advantage in favor of SCP2-driven transcription compared to the pRC/CMV promoter. This advantage is observed by the fluorescence intensity levels of all fluorescent HeLa S3 cells during days 2 and 4 (Figure 10a).

In terms of the number of EGFP-expressing HeLa S3 cells, we observed temporal variations upon comparison of SCP2 to pRC/CMV, however in all EGFP-expressing SH-SY5Y cells, SCP2 expression is significantly stronger in days 2 and 4 (Figure 10c, d).
Figure 11. FACS analysis of long-term fluorescence intensity and number of fluorescent cells in HeLa S3 and SH-SY5Y cells.

HeLa S3 and SH-SY5Y cells were transiently transfected with pRC/CMV, natural CMV, SCP2 or SCP3 vector expressing EGFP. The cells were collected 4-8 days post-transfection (P.T.) for FACS analysis. (a) FACS analysis of fluorescence intensity of all HeLa S3 fluorescent cells and SH-SY5Y fluorescent cells. (b) FACS analysis of fluorescence intensity of high intensity HeLa S3 fluorescent cells and SH-SY5Y fluorescent cells. (c) FACS analysis of the number of all HeLa S3 fluorescent cells and SH-SY5Y fluorescent cells. (d) FACS analysis of the number of high intensity HeLa S3 fluorescent cells and SH-SY5Y fluorescent cells. Data shown are representative of 6 independent experiments using HeLa S3 cells and 5 independent experiments using SH-SY5Y cells. Statistical comparisons between the promoters were done using the Kruskal–Wallis test with pairwise comparisons. A p-value ≤ 0.05 was considered to be statistically significant.
FACS analysis of long-term expression (performed during days 4-8 post-transfection), reveals an advantage of SCP3 over the pRC/CMV promoter in both HeLa S3 and SH-SY5Y cells (Figure 11). SCP3 exhibits a significant advantage over the pRC/CMV promoter in the entire population of fluorescent HeLa S3 cells in days 4-7. The fluorescence intensity of all EGFP-expressing SH-SY5Y cells also reveals stronger transcriptional activity of the SCP3 promoter as compared to pRC/CMV, in days 4, 5 and 8 (Figure 11a). Analysis of the fluorescence intensity of HeLa S3 and SH-SY5Y cells that express high levels of EGFP does not indicate constant significant differences among the promoters (Figure 11b).

SCP3 directs the long-term expression of a larger number of HeLa S3 and SH-SY5Y cells that express EGFP as compared to the pRC/CMV promoter (Figure 11c, d). This potent activity of SCP3 is observed in days 5-8 of all fluorescent HeLa S3 cells and in days 4, 6 and 8 of all fluorescent SH-SY5Y cells (Figure 11c). The advantage of SCP3 over pRC/CMV is also observed during days 4, 5 and 7 of the high EGFP-expressing HeLa S3 cells, and days 4, 5, 6 and 8 of the high EGFP-expressing SH-SY5Y cells (Figure 11d).

SCP2-driven transcription is significantly stronger as compared to the pRC/CMV promoter in terms of fluorescence intensity levels of all EGFP-expressing HeLa S3 cells during days 4, 6 and 7 (Figure 11a). The advantage of SCP2 is also reflected by the number of EGFP-expressing cells in days 5, 7 and 8 (Figure 11c).
**pEGFP-N1 based constructs**

Transcriptional activity of a specific gene following a transient transfection may be influenced by the vector backbone that contains the gene. Hence, even if the cloned gene, core promoter and enhancer are the same, the expression levels may be different when using different vector backbones. Thus, we wanted to investigate the transcriptional activity of the four promoters in a different vector backbone, the pEGFP-N1 vector (Clonetech).

The commercial pEGFP-N1 vector contains an incomplete CMV promoter, a CMV enhancer and also the *EGFP* reporter gene. Unlike the CMV promoter present in the commercial pRC/CMV vector, in this vector the promoter lacks only the sequence that is downstream of +5 relative to the +1 transcription start site, hence it contains the CMV TATA box and Inr elements.

The designated core promoters, from the pRC/CMV based constructs, were cloned by Shani Basch- Barzilay into the pEGFP-N1 vector, thus creating a new set of constructs based on the pRC/CMV promoter integrated into the pEGFP-N1 vector (Figure 12):

- pEGFP-N1
- pEGFP-N1-natural CMV
- pEGFP-N1-SCP2
- pEGFP-N1-SCP3

![Figure 12. Schematic representation of the four pEGFP-N1-based core promoters](image)

pEGFP-N1, natural CMV, SCP2 and SCP3
Similarly to the experiments done with the pRC/CMV constructs, we examined both the fluorescence intensity and the number of cells expressing \textit{EGFP} during 8 days after transfection (divided into short- and long-term).

To investigate the transcription activity of the core promoters, several human cells types (as described below) were transiently transfected with the pEGFP-N1, pEGFP-natural CMV, pEGFP-SCP2 or pEGFP-SCP3 vectors, and subjected to short-term and long-term FACS analysis. The following FACS results were statistically analyzed with ANOVA, Bonferroni multiple comparisons. A p-value $\leq 0.05$ was considered to be statistically significant. The presented FACS analyses depict representative experiments, whereas the statistical significance was calculated for at least 3 individual experiments (the exact number of experiments is indicated below for each experiment).

The following FACS analyses, performed with the various pEGFP-N1 vectors, were conducted in collaboration with Shani Basch-Barzilay.

\textbf{HeLa S3 cells}

A human derived cervical carcinoma cell line. The reason for using this cell line is the same as described in pRC/CMV -based constructs section. Representative results of 3 independent experiments for short-term and long-term expression are presented in Figure 13.
HeLa S3 cells were transiently transfected with pEGFP-N1, natural CMV, SCP2 or SCP3 vector expressing \textit{EGFP}. The cells were collected 1-8 days post-transfection (P.T.) for FACS analysis. (a) Short-term FACS analysis of fluorescence intensity. (b) Long-term FACS analysis of fluorescence intensity. (c) Short-term FACS analysis of the number of fluorescent cells. (d) Long-term FACS analysis of the number of fluorescent cells. Data shown are representative of 3 independent experiments. Statistical comparisons between the promoters were done using the ANOVA, Bonferroni multiple comparisons posttest. A p-value ≤ 0.05 was considered to be statistically significant.

FACS analysis of the fluorescence intensity, reveals significant advantages in favor of the full-length core promoters over the pEGFP-N1 during days 4 and 8, and also between pEGFP-N1 and the natural CMV on day 3. On days 2 and 3, the differences between pEGFP-N1 and SCP3 are not significant, but their P value are very low (p < 0.072 and p < 0.069 respectively) (Figure 13a, b).

Analysis of the number of cells expressing EGFP, exhibits a significant advantage over the pEGFP-N1 favoring the natural CMV during days 3 and 4, the SCP2 on days 3-4 and 8, and also the SCP3 promoter on days 5, 7 and 8. The difference between pEGFP-N1 and SCP2 on day 7 is not significant, however the obtained P value is very low (p < 0.067) (Figure 13c, d). The results reveal a distinct advantage in favor of the full-length core promoters over the
incomplete promoter of the commercial pEGFP-N1, both in terms of fluorescence intensity and number of fluorescent cells.

**SH-SY5Y cells**

A human derived neuroblastoma cells line. The reason for using this cell line is the same as described in pRC/CMV -based constructs section. Representative results of 6 independent experiments for short-term and long-term expression are presented in Figure 14.

**Figure 14. FACS analysis of short-term and long-term fluorescence intensity and number of fluorescent cells in SH-SY5Y cells**

SH-SY5Y cells were transiently transfected with pEGFP-N1, natural CMV, SCP2 or SCP3 vector expressing *EGFP*. The cells were collected 1-8 days post-transfection (P.T.) for FACS analysis. 

(a) Short-term FACS analysis of fluorescence intensity. (b) Long-term FACS analysis of fluorescence intensity. (c) Short-term FACS analysis of the number of fluorescent cells. (d) Long-term FACS analysis of the number of fluorescent cells. Data shown are representative of 6 independent experiments. Statistical comparisons between the promoters were done using the ANOVA, Bonferroni multiple comparisons posttest. A p-value ≤ 0.05 was considered to be statistically significant.
FACS analysis of the fluorescence intensity reveals that all three full core promoters are significantly stronger than the pEGFP-N1 promoter during days 1, 2, 4, 5 and 8. In addition, significant preferential activity over the pEGFP-N1 promoter is also observed favoring SCP2 on day 3 as well as SCP3 during days 3 and 7. Comparison between SCP3 and pEGFP-N1, on day 6, resulted with a very low P value ($p < 0.056$) though insignificant (Figure 14a, b).

SCP2 directs the expression of a significantly larger number of cells that express $EGFP$, as compared to the pEGFP-N1 during all eight days examined. Additionally, both the natural CMV promoter and SCP3 are significantly stronger than the pEGFP-N1 promoter during days 3-8 (Figure 14c, d).

Hence, short-term as well as long-term analyses indicate that there is a distinct increase in transcriptional activity in favor of the full-length core promoters (natural CMV, SCP2 and SCP3) with regards to the fluorescence intensity level. Moreover, analysis of the fluorescence intensity in the vast majority of the experiments indicates a slight advantage in favor of either SCP2 or SCP3 compared to the natural CMV. Similarly, regarding the number of cells, in all 8 days examined there are substantially more cells expressing $EGFP$ under the regulation of the three full core promoters. Additionally, the numbers of either SCP2 or SCP3 expressing cells in days 4-8 is higher than the natural CMV regulated cells.

**HOP-92 cells**

A human derived non-small Lung Carcinoma cells line. This cells have a particularly long doubling time of 72.5 h (in comparison to the 24-48 h doubling time of the other cells lines we used). As the experiments were done by transient transfections, the transfected vectors do not replicate during cell division, thus are diluted within the cell population throughout the generations. For this reason, we chose to use the HOP-92 cell line, to eliminate the untransfected daughter cell's impact on the results, by reducing the dilution of the transfected vectors.
Unfortunately, contrary to our expectations and despite the cells' long doubling time, \textit{EGFP} expression has quickly decreased reaching very low levels (Figure 15). Since the low expression has already been observed on day 6, we examined the fluorescence intensity and the number of cells expressing \textit{EGFP} only during 5 days post-transfection. Representative results of 3 independent experiments are presented in Figure 15.

\textbf{Figure 15. FACS analysis of fluorescence intensity and number of fluorescent cells in HOP-92 cells}

HOP-92 cells were transiently transfected with pEGFP-N1, natural CMV, SCP2 or SCP3 vector expressing \textit{EGFP}. The cells were collected 1-5 days post-transfection (P.T.) for FACS analysis. (a) FACS analysis of fluorescence intensity. (b) FACS analysis of the number of fluorescent cells. Data shown are representative of 3 independent experiments. Statistical comparisons between the promoters were done using the ANOVA, Bonferroni multiple comparisons posttest. A $p$-value $\leq 0.05$ was considered to be statistically significant.

FACS analysis of the fluorescence intensity reveals a significant advantage, during all five examined days, in favor of SCP3-driven transcription compared to the pEGFP-N1 promoter. The natural CMV promoter and SCP2 also exhibits significant advantages over the pEGFP-N1 promoter. These advantages are reflected during days 1, 3, 4 and 5 regarding the natural CMV promoter, and during days 4 and 5 with regards to SCP2 (Figure 15a).

The analysis of the number of cells expressing \textit{EGFP} reveals significant advantages favoring all full-length core promoters over the pEGFP-N1, in almost all examined days. An
exception is SCP2 that doesn’t exhibit a significant advantage over the pEGFP-N1 in day 1 (Figure 15b).

Using the HOP-92 cell, unlike the previously examined cell lines, revealed that SCP3 has an advantage both in terms of fluorescence intensity and number of fluorescent cells, over the other three core promoters. Additionally, SCP2 is slightly better than the natural CMV, especially on days 4 and 5.

**The Mechanism of SCPs Action**

Characterization of the incomplete CMV, natural CMV, SCP2 and SCP3, during short and long time periods and in different backbones and cell lines, revealed a high and prolong transcriptional activity of the full-length core promoters. The fact that the transcriptional activity of these potent core promoters is much higher than that of other commercial versions of considerably strong viral promoters (such as the CMV IE1 promoter within the pEGFP-N1 vector), led us to try and investigate the mechanism that enables the strong activities of these promoters.

**Examination whether DNA methylation is involved in the potent activity of the SCPs**

As hypermethylation of promoter’s sequences, by the DNA methyl-transferase enzyme, is a known and common reason for promoter silencing, we first examined whether different methylation patterns are involved in the mechanism that enables the potent activity of the full-length core promoters. To this end, we used a demethylating agent named – 5’-Aza-deoxycytidine.

5’-Aza-deoxycytidine (5’-Aza-dC), is a cytidine analog that serves as a demethylating agent. It inhibits the DNA methyl-transferase enzymatic activity, and thus prevents DNA methylation (Figure 16).
If DNA methylation contribute to the transcriptional activities of the full-length core promoters, we expect that upon treatment with 5′-Aza-dC, the advantage of the full core promoters (natural CMV, SCP2 and SCP3) over the incomplete CMV promoter will substantially decrease, resulting in similar transcription levels of all the examined promoters. SH-SY5Y cells were transiently transfected with the pEGFP-N1, pEGFP-natural CMV, pEGFP-SCP2 or pEGFP-SCP3 vectors, and treated with 1-7.5µM of 5′-Aza-dC for 72 h (as is customary in the literature\(^{35,39}\)). The growth medium was changed and fresh 5′-Aza-dC was added every 24 h, and the cells were subjected to FACS analysis after 72 h. Representative results of 3 independent experiments are presented in Figure 17.

Figure 17. FACS analysis of fluorescence intensity and number of fluorescent cells after 5′-Aza-dC treatment in SH-SY5Y cells

SH-SY5Y cells were transiently transfected with pEGFP-N1, natural CMV, SCP2 or SCP3 vector expressing EGFP. The cells were treated with various concentrations of 5′-Aza-dC for 72 h post-transfection (P.T.), and collected for FACS analysis. (a) FACS analysis of fluorescence intensity. (b) FACS analysis of the number of fluorescent cells. Data shown are representative of 3 independent experiments. Statistical comparisons between the promoters were done using the ANOVA, Bonferroni multiple comparisons posttest. A p-value ≤ 0.05 was considered to be statistically significant.
FACS analysis reveals significant fluorescence intensity advantages of the full-length core promoters over the pEGFP-N1 promoter. However, no significant differences in the number of cells expressing EGFP under the regulation of the four core promoters are observed. These results are essentially the same with all examined doses of 5’-Aza-dC (Figure 17a, b).

The treatment with 5’-Aza-dC also led to a considerable increase in the EGFP expression, regardless of the core promoter and the dose of 5’-Aza-dC. Since demethylation is a known cause for promoter’s activation, this observation is a confirmation of proper 5’-Aza-dC activity.

As the 5’-Aza-dC treatment had no impact on the differences in transcriptional activity between the various promoters, it can be assumed that DNA methylation is not involved in the mechanism of action of the full-length core promoters.

**Examination whether histone acetylation is involved in the potent activity of SCPs**

Histone deacetylation is an epigenetic mechanism associated (among other mechanisms, such as DNA hypermethylation) with gene silencing. Unlike DNA methylation, histone acetylation is usually correlated with transcriptionally active promoters, as acetylation leads to chromatin relaxation - a structure that provides the transcription factors an easy access to the DNA. Since DNA methylation does not appear to be involved in the mechanism of the full-length core promoters, we next decided to examine whether inhibition of histone deacetylations influences the transcriptional activity of the promoters.

Trichostatin A (TSA), is a *Streptomyces* metabolite that specifically inhibits mammalian histone deacetylase (HDAC), thus resulting in histone hyperacetylation that leads to chromatin relaxation (Figure 18).
As was initially assumed with the 5’-Aza-dC treatment, we predicted that if histone acetylation contributes to the full-length core promoters activity, the TSA treatment will result in similar activity levels regarding of the examined promoters, hence the advantage of the full core promoters over the incomplete CMV promoter will decrease.

SH-SY5Y and HeLa S3 cells were transiently transfected with the pEGFP-N1, pEGFP-natural CMV, pEGFP-SCP2 or pEGFP-SCP3 vectors and treated with 100-1000nM of TSA for 48 h (as is customary in the literature \textsuperscript{3,33}). The cells were harvested and subjected to FACS analysis 48 h post-transfection. Representative results of and 2 independent experiments using HeLa S3 cells and 3 independent experiments using SH-SY5Y cells are presented in Figure 19.
HeLa S3 and SH-SY5Y cells were transiently transfected with pEGFP-N1, natural CMV, SCP2 or SCP3 vector expressing EGFP. The cells were treated with various concentrations of TSA for 48 h post-transfection (P.T.), and collected for FACS analysis. (a) FACS analysis of fluorescence intensity of HeLa S3 cells. (b) FACS analysis of the number of HeLa S3 fluorescent cells. (c) FACS analysis of fluorescence intensity of SH-SY5Y cells. (d) FACS analysis of the number of SH-SY5Y fluorescent cells. Data shown are representative of 2 independent experiments using HeLa S3 cells and 3 independent experiments using SH-SY5Y cells. Statistical comparisons between the promoters were done using the ANOVA, Bonferroni multiple comparisons posttest. A p-value ≤ 0.05 was considered to be statistically significant.

FACS analysis of the fluorescence intensity reveals significant advantages in favor of the full-length core promoters over the pEGFP-N1, upon treatments with low TSA doses of 1-175nM. On the other hand, treatment with higher doses of TSA does not result in significant differences among the promoters. Moreover, treatments with doses of 500nM and 1000nM TSA completely eliminates any differences between the promoters. These results are observed in both HeLa S3 and SH-SY5Y cells (Figure 19a, c).
All three full core promoters direct the expression of a larger number of HeLa S3 and SH-SY5Y cells that express EGFP, as compared to pEGFP-N1, when no TSA treatment is used (0nM). These differences, however, are not significant (with a very low P value - p < 0.058). No significant differences in the number of fluorescent cells were observed upon TSA treatment, regardless of the dose used (Figure 19b, d). Additionally, a direct correlation between the TSA doses and the number of HeLa S3 and SH-SY5Y cells that express EGFP can be observed, as increasing concentrations of TSA treatment result in a decrease in the number of fluorescent cells. This correlation results from the toxic properties of TSA that lead to an overall decrease of living cells.

In addition to the main activity of TSA as an HDAC inhibitor, TSA has also been shown to induce DNA demethylation\(^4\). Nevertheless, these findings, together with the 5’-Aza-dC results, support the assumption that acetylation, and not the DNA methylation, is at least partially the cause for the differences in promoter activity levels. Thus, it is possible that histone acetylation contribute to the potent transcription activity of the full-length core promoters, either in a direct or indirect manner, as discussed in detail below.

**Examination whether synergism between DNA methylation and histone acetylation is involved in the potent activity of the SCPs**

DNA methylation and regulation of chromatin structure (by histone acetylation patterns) are typically considered as two distinct epigenetic mechanisms that are involved in the regulation of gene expression. However, there are studies indicating a synergistic effect of DNA demethylation and histone hyperacetylation in the reactivation of silenced promoters, even when each one of them can’t induce reactivation independently \(^4,34,41,42\).

In terms of mechanism, this synergism is attributed to methylated cytosines that can be bound by methyl-CpG binding proteins (MBD). These MBDS can recruit multiprotein complexes, such as HDACs, which eventually prevent gene transcription. In addition, chromatin modifying enzymes also have the ability to increase DNA methylation by recruiting DNA methyltransferases (DNMTs). Therefore, it appears that DNA methylation and histone
deacetylation (which can stimulate each other) can modify local chromatin, into a more dense structure, thus repressing genes transcription. The opposite mechanism, in which histone hyperacetylation (induced by histone acetyltransferases (HATs) and HDACs inhibition) promotes DNA demethylation, is still unknown. However, the assumption is that hyperacetylated DNA regions are more exposed, and therefore prone, to demethylation processes. Regarding the TSA treatment, as explained above, it is possible that TSA influences the acetylation of proteins that are involved in DNA methylation or demethylation, thus changing their activity.

If the advantage of the full-length core promoters over the incomplete promoter is achieved due to synergism between DNA demethylation and histone hyperacetylation, then a combined treatment of both 5'-Aza-dC and TSA would eliminate the differences between the full-length promoters and the pEGFP-N1.

SH-SY5Y cells were transiently transfected with the pEGFP-N1, pEGFP-natural CMV, pEGFP-SCP2 or pEGFP-SCP3 vectors, and treated with 500nM TSA and with 1-7.5µM of 5'-Aza-dC. The growth medium containing the TSA and 5'-Aza-dC was changed after 24 h, and the cells were subjected to FACS analysis after 48 h. This experiment was performed three times, however one repetition yielded exceptional results. Thus, it is necessary to conduct additional experiments in order to draw unequivocal conclusions. Representative results of 2 independent experiments are presented in Figure 20.
SH-SY5Y cells were transiently transfected with pEGFP-N1, natural CMV, SCP2 or SCP3 vector expressing EGFP. The cells were either not treated at all (as control) or treated with 500nM TSA together with changing 5'-Aza-dC concentrations of 0-7.5µM for 48 h post-transfection (P.T.), and collected for FACS analysis. (a) FACS analysis of fluorescence intensity. (b) FACS analysis of the number of fluorescent cells. Data shown are representative of 2 independent experiments. Statistical comparisons between the promoters were done using the ANOVA, Bonferroni multiple comparisons posttest was performed on all three experiments. A p-value ≤ 0.05 was considered to be statistically significant.

FACS analysis of the control group (untreated cells) reveals that differences in the fluorescence intensity between pEGFP-N1 and each of the three full-length core promoters are not significant, but with a low P value (p < 0.062). The reason that the full promoters are not significantly stronger is probably due to the variations in repetitions (as described above). As expected of the control group, a significantly larger number of EGFP-expressing cells are detected in cells that express EGFP that is driven by the full core promoters, as compared to the pEGFP-N1 (Figure 20a, b).

The analyses of the cells that were treated with TSA, together or without a 5'-Aza-dC treatment, reveal no significant differences between the four core promoters. These insignificances are reflected both by the fluorescence intensity and by the number of cells.
that express *EGFP*. In addition, the differences are mostly lost, as observed by the combined treatment of 500nM TSA and 5µM or 7.5µM (the high concentrations) of 5’-Aza-dC (Figure 20a, b).

Consistent with the previous results of the 5’-Aza-dC and the TSA treatments, the combined treatment indicates that the reduction of differences between the transcriptional activities of the promoters is obtained due to the TSA activity. In the combine treatment, the differences between the promoters were reduced once the TSA was added (with or without the 5’-Aza-dC), while the addition of the 5’-Aza-dC merely caused a slight change in the differences, and even that was obtained only with high 5’-Aza-dC concentrations. Therefore, based on the 5’-Aza-dC, TSA and the combined treatments, it is reasonable to assume that acetylation patterns are involved in the mechanism that grants the full-length core promoters their transcriptional advantage over the incomplete core promoter.

**The involvement of TAF1 in regulating the activity of SCPs**

Aiming to deepen our understanding of the mechanism that confers the natural CMV, SCP2 and SCP3 promoters their advantage over the incomplete CMV promoter, we searched for a lead that will connect TSA, histone and transcription factors acetylations, core promoter’s composition and the transcription process.

A prominent candidate that meets these criteria is TAF1.

**TAF1 (TATA Box Binding Protein (TBP) Associated Factor I)**, is the largest subunit of TFIID (the first protein complex to bind the transcription start region and recruit RNA polymerase II). TAF1 serves as an element that catalyzes TFIID formation and regulates TBP’s binding to the TATA box motif. In addition, TAF1 possesses a variety of enzymatic activities (such as protein phosphorylation, ubiquitin activating/conjugating enzyme and acetylation) that provide TAF1 the ability to modify histones and transcription factors. Additionally, TAF1 has a bromodomain motif that enables it to recognize and bind acetylated lysine residues.
The fact that TAF1 is an essential basal transcription factor in addition to its ability to bind acetylated lysine residues in different proteins and to acetylate histones and transcription factors, led us to consider TAF1 as part of the full-length core promoter's potent activity mechanism.

In order to investigate the potential role of TAF1’s HAT activity in the full-length core promoters’ mechanism, we used the ts13 cell line.

**ts13 cells**

A BHK-21 (Baby Hamster Kidney) (see below)-derived temperature sensitive cell line that harbors a single amino acid substitution mutation (G690D) in the HAT domain of TAF1. At the permissive temperature (33.5°C) the HAT domain of TAF1 is functional, while at the restrictive temperature (39.5°C) the HAT activity is disrupted so TAF1 can not perform acetyl transferase activities. This impaired activity of TAF1 at the restrictive temperature causes late G1 arrest of the cells and disrupts the transcription of some protein-coding genes, including cell cycle regulators such as cyclin D1, cyclin A and cyclin E.

**BHK-21 cells**

A Baby Hamster Kidney cell line, from which the ts13 cell line was derived. This cell line does not contain the G690D mutation, so that the HAT domain of TAF1 is active at both the permissive and the restrictive temperatures. The BHK-21 cell line was used as a control for changes in EGFP expression that result from the temperature change.

Assuming that the HAT activity of TAF1 is indeed involved in the full-length core promoters mechanism, we expect transfected ts13 cells at the permissive temperature (33.5°C) to result in similar differences between the promoters as in the previous experiments, while at the restrictive temperature (39.5°C) we expect the differences between the promoters to be less pronounced. Regarding the control BHK-21 cells, we expect the differences between the promoters to be similar to the previous experiments at both temperatures.
BHK-21 and ts13 cells were transiently transfected with the pEGFP-N1, pEGFP-natural CMV, pEGFP-SCP2 or pEGFP-SCP3 vectors (in duplicates), and incubated at 33.5°C for three hours. After the three hours incubation, one sample of each duplicate was kept incubated at 33.5°C while the second was transferred to incubation at 39.5°C. The cells were harvested and subjected to FACS analysis 48 h post-transfection. As the number of EGFP-expressing cells is high under the regulation of all the promoters, we analyzed the number of cells that express high fluorescence levels (designated “HIGH EXP”) in addition to the entire population of fluorescent cells (designated “ALL EXP”). Representative results of 2 independent experiments (hence statistical analysis was not performed) are presented in Figure 21.

**Figure 21.** FACS analysis of fluorescence intensity and number of fluorescent cells at 33.5°C and 39.5°C in BHK-21 and ts13 cells

BHK-21 and ts13 cells were transiently transfected with pEGFP-N1, natural CMV, SCP2 or SCP3 vector expressing EGFP. The cells were incubated at either the permissive (33.5°C) or the restrictive (39.5°C) temperature for 48 h post-transfection (P.T.), and collected for FACS analysis. **(a)** FACS analysis of fluorescence intensity of all BHK-21 fluorescent cells. **(b)** FACS analysis of the number of all BHK-21 fluorescent cells. **(c)** FACS analysis of the number of high intensity BHK-21 fluorescent cells. **(d)** FACS analysis of fluorescence intensity of all ts13 fluorescent cells. **(e)** FACS analysis of the number of all ts13 fluorescent cells. **(f)** FACS analysis of the number of high intensity ts13 fluorescent cells. Data shown are representative of 2 independent experiments using both BHK-21 and ts13 cells.
FACS analysis of the BHK-21 cells reveals an advantage in favor of SCP2 and SCP3 over both the pEGFP-N1 and the natural CMV promoters. This advantage is reflected by the fluorescence intensity levels as well as by the number of high fluorescence intensity cells, at both temperatures (Figure 21a, c). Overall the BHK-21 cells display a similar trend of differences between the promoters, at both the permissive and the restrictive temperatures. In addition, at the restrictive temperature a noticeable decrease in the fluorescence intensity levels of all fluorescent cells as well as in the number of high EGFP-expressing BHK-21 cells is observed, probably due to the heat shock response of the cells (Figure 21a-c).

As expected of the ts13 cells at the permissive temperature (33.5°C), the results are similar to those obtained by the BHK-21 cells, hence indicating an advantage in favor of SCP2 and SCP3 that is reflected by the fluorescence intensity levels and by the number of high fluorescence intensity cells (Figure 21d, f). Interestingly, at the restrictive temperature (39.5°C) the differences between the promoters are substantially reduced, both in terms of fluorescence intensity levels and number of high EGFP-expressing ts13 cells. These reduction of differences resulted with roughly similar transcription levels in all the examined promoters (as obtained previously with the TSA treatment). Surprisingly, unlike the BHK-21 heat shock response, the fluorescence intensity levels of all ts13 cells increased at the restrictive temperature (Figure 21d-f).

Using both BHK-21 and ts13 cell lines, we observe a substantial advantage in favor of the SCP2 and SCP3 promoters over the pEGFP-N1 and the natural CMV (unlike the HeLa S3 and SH-SY5Y cell lines, here the fluorescence intensity of pEGFP-N1 and the natural CMV are almost the same). Thus, using these cell lines we observe a transcriptional priority to promoters that comprise all four promoter elements (TATA box, Inr, MTE and DPE) in comparison to promoters that contain only some of the promoter elements (for example the TATA box and Inr in the natural CMV promoter). The results obtained at the restrictive temperature with the ts13 cell line indicating that the differences between the promoters are substantially reduced, are reminiscent of the results observed upon treatment
with the histone deacetylase inhibitor, TSA. This resemblance may indicate a relation between TAF1 activity, TSA treatment and core promoter’s composition.

**Characterization of the Influence of Core Promoter Elements on Transcription Activity and Kinetics**

This section was by utilizing a system developed by Prof. Yaron Shav-Tal, Prof. Yuval Garini, Dr. Sharon Yunger and the Ph.D. student Liat Rosenfeld in collaboration with Prof. Yaron Shav-Tal’s lab.

Aiming to investigate the individual, as well as the combined contribution and influence of each core promoter element on promoter-driven transcriptional activity and kinetics, we used an *in vivo* system that allows visualization and analysis of mammalian mRNA in real-time. This system provides the ability to avoid the major impact of the inconsistent copy number and the inexact integration locus of transfected genes on their expression. The system is based on a single allele - specific locus integration, thus enabling the visualization and analysis of mRNA synthesis from a single specific allele. To that end, gene integration is performed using the Flp-In™ system (Life Technologies), which enables the integration of a single copy of a gene into a specific genomic location. The integration is catalyzed by the Flp-recombinase enzyme, which mediates homologous recombination between two FRT sites (Flp-Recombinase Target): one in the plasmid together with the gene and promoter of interest, and the second in a specific location at the genome of the cell. As a result of the recombination event, the cells contain the transfected gene under the regulation of the desired promoter, and they also acquire resistance to Hygromycin B and become sensitive to Zeocin (*Figure 22*), hence a selection process can be done in order to generate Flp-In stable cell lines.
HEK 293 Flp-In cells are co-transfected with the pcDNA5/FRT vector (containing a desired promoter and gene) and the pOG44 plasmid (Flp-recombinase expression plasmid). Flp-recombinase catalyzes homologous recombination between the FRT site in the pcDNA5/FRT vector and the FRT site in the genome of the cells. Transfected cells express the desired gene and become Hygromycin B resistant and Zeocin sensitive.

Once the stable cell line is established, it is possible to investigate the activity of the gene (which is detected as a specific fluorescent locus in the nucleus of the cell) in real time, by the use of visual approaches, such as live cell imaging. Real-time visualization of the activity of the gene is performed by labeling mRNAs with MS2-GFP. The use of the MS2 protein provides a mean to indirectly label specific mRNAs with a fluorescent protein. The MS2 protein is a phage RNA-binding protein that normally forms the capsid of the bacteriophage, however the MS2 protein can also bind MS2 stem-loop structured RNA. Incorporation of several MS2-binding sites to the gene’s RNA will enable many MS2-GFP fusion proteins to bind the MS2-binding sites and thus to the desired RNA (Figure 23).
In this manner, expression of RNA containing MS2-binding sites and MS2-GFP proteins at the same time, provides a good method to detect mRNPs\textsuperscript{46}. As the MS2-GFP proteins also contain a nuclear localization signal (NLS) sequence that targets them to the nucleus, they can bind transcripts (of RNA that contain MS2-binding sites) as they are transcribed, resulting in an accumulation of transcripts bound by MS2-GFP proteins in the nucleus. This accumulation is created in a specific site within the nucleus that is the transcription site. This in vivo system of single allele integration into a specific locus, and fluorescence labeling of the desired gene’s transcripts and transcription site, will provides us with the ability to analyze the transcription of a single copy exogenous gene, and thus investigate the kinetics of transcription driven by different promoters. It will also provide evidence for the contribution of various core promoter elements to gene expression.

The construction of the system and part of the analyses were done together with Shani Basch- Barzilay (a former M.Sc. student in our lab), with the assistance of Dr. Sharon Yunger, who developed the system together with Prof. Yaron Shav-Tal and Pinhas Kafri, a Ph.D. student in the Shav-Tal lab who uses the system in his research.

**Plasmids**

Dr. Sharon Yunger and Prof. Yaron Shav-Tal have previously used the HEK293 Flp-In cell line to generate two cell clones, each containing a single copy of the human Cyclin D1 gene (CCND1) under the regulation of either the Cyclin D1 endogenous promoter (CCND1\textsubscript{pr}) or the CMV promoter (CMV\textsubscript{pr})\textsuperscript{44}. These cell clones were used to compare the transcriptional kinetics of CCND1 driven by the two promoters. In order to generate these clones two vectors were constructed:
• **FRT-CMVpr–CCND1**: Contains the CCND1 gene under the regulation of the commercial CMV promoter (pcDNA5\FRT). The core promoter contains the TATA box element, but lacks the sequence that is downstream of -16 relative to the +1 transcription start site, including the Inr element.

• **FRT-CCND1pr–CCND1**: Contains the CCND1 gene under the regulation of its own endogenous promoter (CCND1pr). The core promoter contains the Inr and the DPE elements.

As our aim is to investigate the impact of the core promoter and particularly the contribution of different core elements on transcription activity and kinetics, two sets of constructs were build (by Shani Basch- Barzilay):

1. A set based on the FRT- CMVpr-CCND1 construct, in which the CCND1 gene is under the regulation of the CMV promoter (Figure 24):

   • **FRT- CMVpr–CCND1**: The core promoter of the commercial CMV promoter that contains only the TATA box element. This construct was received from Prof. Yaron Shav-Tal's lab.
   • **FRT-NAT CMV–CCND1**: The complete CMV core promoter (from -36 to +45) that contains the TATA box and the Inr elements.
   • **FRT-SCP2–CCND1**: The synthetic SCP2 core promoter that contains the TATA box, Inr, MTE and the DPE elements.
   • **FRT- SCP3–CCND1**: The synthetic SCP3 core promoter that contains the TATA box, Inr, MTE and the DPE elements (the differences between SCP2 and SCP3 are specified at the introduction section).
2. A set based on the FRT-CCND1pr-CCND1 construct, in which the CCND1 gene is under the regulation of its own endogenous promoter (Figure 25):

- **FRT-CCND1pr–CCND1**: The endogenous core promoter of Cyclin D1 that contains the Inr and the DPE elements. This construct was received from Prof. Yaron Shav-Tal’s lab.
- **FRT-ENHANCED CCND1pr-CCND1**: The core promoter contains the Inr, MTE and DPE elements.
- **FRT-SUPER CCND1pr-CCND1**: The core promoter contains the TATA box, Inr and DPE elements.
- **FRT-SUPER DUPER CCND1pr-CCND1**: The core promoter contains the TATA box, Inr, MTE and DPE elements (similarly to the SCP promoter).
Figure 25. Schematic representation of four construct that contain the CCND1 gene under the regulation of different promoters based on the CCND1 endogenous promoter

**Generation of stable cell clones**

In order to generate the eight cell clones (containing a single copy of the CCND1 gene driven by one of the above promoters in a specific locus), HEK 293 Flp-In cells were transfected with each of the described constructs and selected with Hygromycin B for three weeks. After selection, single cell clones were isolated and grown in order to expand the population. Once the cell clones were stable, verifications of the accurate single allele integrations were done (as described below). Importantly, in order to conduct an accurate comparison between the promoters, two of the eight cell clones we generated are identical to those previously used by Dr. Sharon Yunger (FRT- CMVpr–CCND1 and FRT- CCND1pr-CCND1), hence their characteristics should be the same.
Verification of proper system set-up

In order to verify that the single allele integrations are accurate, so that the mRNA of CCND1 is indeed transcribed from a single copy of the gene, we performed RNA Fluorescent *In Situ* Hybridization (FISH), X-gal (lacZ) assays, PCR and DNA sequencing of genomic DNA.

1. RNA Fluorescence *In Situ* Hybridization (FISH)

   Detection of a single active transcription site can verify the accurate single allele integration in a visual manner. To this end, FISH experiments were performed with a Cy3-labeled probe to the MS2 repeats in the CCND1 transcripts. FISH results for all eight cell clones display a single bright dot inside the nucleus that arises from the accumulation of CCND1 transcripts bound by labeled MS2 proteins. These detected dots represent the location where the CCND1 transcripts are transcribed, thus the active transcription site (*Figure 26, 27*). The CMV promoter is a potent viral promoter that constantly drives expression, while the CCND1 promoter driven transcription is not constant so the transcription site may fluctuate between “on” and “off” states. Accordingly, observations of cell populations from the first set of constructs (based on the CMV promoter) displays an active transcription site in the vast majority of the cells, whereas in cell populations from the second set of constructs (based on the CCND1 promoter) the active transcription site can only be detected in some of the cells. Hence, the detection of the transcription site depends on the activity state of the CCND1 gene in each individual cell at the time of fixation.
Figure 26. Visualization of the transcription site of Cyclin D1 gene in cell clones that expresses the gene under the regulation of different CMV-based promoters.

RNA FISH experiments were performed with a Cy3-labeled probe against the MS2 repeats in the CCND1 transcripts. Arrows indicate the transcription site of Cyclin D1.
**Figure 27. Visualization of the transcription site of Cyclin D1 gene in cell clones that expresses the gene under the regulation of different CMV-based promoters**

RNA FISH experiments were performed with a Cy3-labeled probe against the MS2 repeats in the CCND1 transcripts. Arrows indicate the transcription site of Cyclin D1.

2. **lacZ (β-galactosidase activity) assay**

Proper integration should yield a recombined genomic region in the HEK 293 Flp-In cells that disrupts the lacZ gene, leading to the lack of β-galactosidase activity. Thus, an X-gal staining experiment can verify that the homologous recombination in the cell clones was indeed correct (positive clones that lost the β-galactosidase activity will not be stained, whereas clones that are negative for the integration, will be stained in blue).

An X-gal staining experiment was conducted in all eight cell clones and also in the maternal HEK 293 Flp-In cells (as control). The results indicate that all of the eight cell clones are positive (not stained), hence verifying that the integration event was correct (**Figure 28**).
3. PCR and DNA sequencing of genomic DNA

In order to verify at the DNA level that only a single copy of the construct was integrated at the correct location, PCR followed by sequencing of the amplified DNA was done. Genomic DNA of each of the eight cell clones (and also HEK 293 Flp-In cells as a negative control) were used as a template for two separate PCRs, each with a set of primers designated to amplify one of the two FRT integration regions. The first PCR is designed to amplify the region between the SV40 promoter and the Hygromycin resistance gene, and the second PCR is designed to amplify the region between the BGH polyA and the lacZ-Zeocin gene. The products of both PCRs show the correct integration of the constructs (Figure 29). As expected of the control HEK 293 Flp-In, no specific products were obtained. The amplified PCR products were sequenced, verifying that the ATG start codon is in frame with the Hygromycin gene (first PCR product) and that the BGH region (downstream of the CCND1 gene) is integrated upstream of the lacZ-Zeocin gene (second PCR product).
Quantitative analyses

By the use of quantitative analyses it is possible to quantify and measure mRNAs located in the nucleus and cytoplasm of a cell. Analysis of mRNAs transcribed under the regulation of different core promoters, will allow us to investigate the contribution and effect of each core promoter element on promoter-driven transcriptional activity as well as the synergistic mechanism of action of combinations of different core promoter elements within the same promoter.

The analysis was done by quantitative RNA Fluorescence In Situ Hybridization (FISH) performed with an MS2-Cy3 probe that binds to MS2 repeats located at the 3' UTR region of CCND1. As a result of MS2-Cy3 probes binding to the transcribed mRNAs, the mRNAs are fluorescently labeled and thus they can be detected (as fluorescent dots) under a microscope. Using wide field microscope, the fluorescent mRNAs in the entire cell volume are identified. Images of total cell volume are acquired by 3D stacks of 76 slices with a step width of 0.2µm (Figure 30).
Figure 30. Visualization of the transcription site and mRNAs of Cyclin D1 in different 3D stacks

Quantitative RNA FISH experiments were performed with a Cy3-labeled probe against the MS2 repeats in the CCND1 transcripts. 2 slices of the 76 stacks are presented in each panel. Top panel- mRNAs can be detected in the cell. Bottom panel- mRNAs in the cell and the transcription site can be detected. (a) A HEK 293 flp-in- FRT-SCP3-CCND1 cell. (b) A HEK 293 flp-in-FRT-ENHANCED-CCND1 cell.
Since it is necessary for the collected fluorescent mRNAs data to be of the best quality possible, each series of stacks (76 slices of one cell) is subjected to deconvolution process (deconvolution was done by Huygens commercial software). Deconvolution is a mathematical process that improves the observed spatial resolution, contrast and signal-to-noise ratio (SNR) of images acquired by fluorescence microscopy. The process restores out-of-focus light back to its original point, thus creating a more accurate image. Next, the deconvolved 3D stacks are processed and analyzed using the Imaris software (Bitplane) that allows visualization and processing of multidimensional microscopy images. The program enables the detection of real signals (which are the labeled mRNAs) over background noise. Furthermore, the program enables the marking of each real mRNA signal by a single spot (as the total volume of a single mRNA might have been imaged by few stacks, every spot is comprised of several rendered stacks). The spotted signals are then analyzed providing a measured fluorescent intensity for each individual mRNA (Figure 31).

Figure 31. Images processing by the Imaris software

Quantitative RNA FISH experiment was performed with a Cy3-labeled probe against the MS2 repeats in the CCND1 transcripts. All presented images are of the same cell (HEK 293 flp-in-FRT-ENHANCED-CCND1). (a) Deconvolved image. (b-e) 3D presentation (f) Each signal (mRNA) is marked by a single spot. (g) Each mRNA is analyzed as a three-dimensional volume.
The measured intensities vary among different individual mRNAs and it is thus necessary to find the most probable intensity mean of a single mRNA. These variations of intensities are derived from variability in the number of probes that bind each mRNA. In order to address the variations in intensity and find the mean of a single mRNA, all measured mRNA intensities (of an individual cell) are plotted as a frequency histogram. Histogram of images that yielded high quality data will result in a Gaussian distribution, where the low-value tail of the histogram represents noise, the high tail represents mRNA aggregations and the most frequent value represents the most probable intensity level of a single mRNA (Figure 32).

![Histogram example](image)

**Figure 32.** Example of histogram plot showing the frequencies of all the mRNA intensities

Once the mean intensity of single mRNA is acquired, it is possible to quantify the number of mRNAs in the nucleus and cytoplasm and also to determine the number of transcripts located at the transcription site (by division of the fluorescent intensity of the transcription site by the fluorescent intensity of a single mRNA). The use of histograms is meant to address the intensity variations, however since this solution is not absolute and since there are variables that are set arbitrarily in each cell analysis, one must analyze large cells population in order to obtain conclusive results. Hence, for each tested cell clone, at list 25 successful repetitions (each repeat is of one individual cell) are required.

It is of high importance to note that although a single quantitative FISH experiment can yield many images, eventually only few of these images provide successful results (this is
because of technical reasons that can occur in any step on the quantitative analysis).

Moreover, after acquiring high numbers of successful results, a statistical analysis should be done on the obtained values in order to eliminate atypical values. As a result, in order to obtain 25 successful results, many more repetitions should be performed.

It should be taken into account that at this point a statistical analysis to eliminate the atypical values hasn’t been performed. For this reason, we present the mean values of all successful results (designated “All repetitions”) as well as the mean values of successful results that we assume would be relevant after the statistical analysis (designated “Estimated repetitions”). In the results we refer to the mean values obtained from the repetitions that we estimate to be relevant, and not to all of the repetitions. The complete table of all the acquired results is presented in Appendix 1.

As a first step of the quantitative analysis we decided to focus on the four cell clones based on the FRT-CCND1pr-CCND1 construct (CCND1 endogenous promoter, Enhanced promoter, Super promoter and Super Duper promoter). We have so far acquired more than 25 successful results for the Endogenous and the Super cell clones, and the analyses of the Enhanced and Super Duper cell clones are still in progress. Results of the obtained mean values are presented in Table 1.
Table 1. Mean values quantification of mRNA molecules at the cell and the transcription site of cell clones based on Cyclin D1 endogenous promoter

Quantitative RNA FISH experiments were imaged by 3D stacks of 76 slices. 3D stacks series were then subjected to deconvolution by the Huygens commercial software. Deconvolved 3D stacks are processed by the Imaris software providing with a measured fluorescent intensity value for each individual mRNA. All measured values (of individual cell) are plotted as a frequency histogram to provide the most probable fluorescent intensity value of a single mRNA. The number of mRNAs located at the transcription site is calculated according to the fluorescent value of a single mRNA. All repetitions refer to the mean values of all successful results. Estimated repetitions refer to the mean values of successful results that we assume would be relevant after the statistical analysis.

As explained previously, the EP cell clone that we generated (in which the Cyclin D1 is regulated by its own endogenous promoter), is identical to the cell clone previously examined by Dr. Sharon Yunger, thus our results are expected to be the same as the previously published results. Indeed, our results of about 8 mRNA molecules at the transcription sites and about 86 mRNA molecules per cell, are compatible with Dr. Sharon Yunger’s results of 7 ± 4 mRNA molecules at the transcription sites and 41 ± 30 mRNAs per cell. The results of the Super cell clone (in which the Cyclin D1 is regulated by a promoter that contains the TATA box, Inr and DPE elements) from which we obtained 26 successful results measurements, indicates about 16 mRNA molecules at the transcription sites and

<table>
<thead>
<tr>
<th></th>
<th>ENDObENOUS</th>
<th>ENHANCED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inr, DPE</td>
<td>Inr, MTE, DPE</td>
</tr>
<tr>
<td>mRNAs per</td>
<td>mRNAs per</td>
<td>mRNAs per</td>
</tr>
<tr>
<td>transcription</td>
<td>transcription</td>
<td>cell</td>
</tr>
<tr>
<td>site</td>
<td>site</td>
<td></td>
</tr>
<tr>
<td>All repetitions</td>
<td>8.44614814</td>
<td>90.6</td>
</tr>
<tr>
<td>Estimated</td>
<td>8.228713231</td>
<td>86.15384615</td>
</tr>
<tr>
<td>repetition</td>
<td>30</td>
<td>26</td>
</tr>
<tr>
<td>number of</td>
<td>19.4861197</td>
<td>121.25</td>
</tr>
<tr>
<td>repetitions</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SUPER</th>
<th>SUPER DUPER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TATA, Inr, DPE</td>
<td>TATA, Inr, MTE, DPE</td>
</tr>
<tr>
<td>mRNAs per</td>
<td>mRNAs per</td>
<td>mRNAs per</td>
</tr>
<tr>
<td>transcription</td>
<td>transcription</td>
<td>cell</td>
</tr>
<tr>
<td>site</td>
<td>site</td>
<td></td>
</tr>
<tr>
<td>All repetitions</td>
<td>15.24049096</td>
<td>156.9354839</td>
</tr>
<tr>
<td>Estimated</td>
<td>16.66262461</td>
<td>158.6538462</td>
</tr>
<tr>
<td>repetition</td>
<td>31</td>
<td>26</td>
</tr>
<tr>
<td>number of</td>
<td>21.30533464</td>
<td>120.5</td>
</tr>
<tr>
<td>repetitions</td>
<td>20</td>
<td>14</td>
</tr>
</tbody>
</table>

|               | 25.01474175 | 142.571429 |
|               | 14          | 14         |
about 158 mRNA molecules per cell, thus two times higher than the endogenous promoter (Table 1).

As for the two remaining CCND1pr-based cell clones (the ENHANCED promoter that contain the Inr, MTE and DPE elements, and the SUPER DUPER promoter that contain the TATA box, Inr, MTE and DPE elements), a sufficient number of successful experiments has not been reached up to this point, and additional experiments and analyses are in progress. However, even these insufficient numbers of experiments suggest a possible advantage in the transcription activity in favor of these two cell clones over the Super cell clone, as well as a substantial advantage over the EP cell clone that consist of Cyclin D1 endogenous promoter (Table 1).
Discussion

Characterization of gene expression under the regulation of modified core promoters

In this section we investigated and compared the transcription activity of four core promoters: the incomplete CMV, the natural CMV, SCP2 and SCP3. All four promoters drive the expression of an EGFP reporter gene, which enabled us to follow the activity of the core promoters over time, both in qualitative (using live cell imaging) and quantitative (using flow cytometry) manners. As gene expression may be influenced by different cellular environments as well as by different vector-backbones (that contain the gene), we used three different types of cell lines and two different vector-backbones.

The results of the pRC/CMV-based expression vectors, reveal a distinct increase in transcriptional activity in favor of the full-length core promoters, particularly the SCP3. This advantage is reflected in HeLa S3 and SH-SY5Y cells, both by high expression levels and by the high number of cells that express EGFP. An advantage in transcriptional activity was also observed with the SCP2, however less pronounced than that of SCP3.

The results of the pEGFP-N1-based expression vectors in HeLa S3 and SH-SY5Y cells reveal a distinct increase in transcriptional activity favoring the full-length core promoters as well, however neither of the full-length promoters displayed a pronounced and consistent preference. Yet, it is of note that in most experiments at least one of the SCPs exhibited some advantage over the natural CMV. The pEGFP-N1-based vectors were also transfected into the HOP-92 cell line that have a long doubling time, with the aim to reduce the dilution effect of the transfected vectors. Similarly to the other two cell lines, an advantage in transcriptional activity was observed in favor of the full promoters, but in this cell line SCP3 is superior to both the natural CMV and SCP2.

All of these experiments were performed by transient transfections and no antibiotic selection was used, suggesting that the transfected DNA did not integrate into the
chromosome and its expression is limited by time, typically considered to be 24-96 h after transfection. Nevertheless, we discovered that under the regulation of the full-length core promoters and particularly SCP2 and SCP3, gene expression is unusually strong and prolonged. Thus, the SCPs are not only unusually strong core promoters with respect to transcription levels, they also provide a novel non-viral way for long-term gene expression in transiently transfected cells. Regarding the two SCPs, SCP3 is at least as potent as SCP2 in terms of expression levels, however its advantage over pRC/CMV is more consistent than SCP2, whose expression is more variable at certain time points. The observed EGFP expression levels driven by the four core promoters (with respect to both vector-backbones), varies between the different types of cell lines indicating that, as expected, the cell lines used may influence gene expression levels.

SCP3 differs from SCP2 in the absence of the T7 promoter as well as in substitutions of four nucleotides. While the core promoter elements composition is the same, the high transcription activity observed for SCP3-driven EGFP indicates that the entire core promoter sequence, and not just the presence of the core promoter elements themselves, is important for core promoter function. Furthermore, the high levels and long-term transcription driven by the SCPs, provides additional support for the importance of downstream core promoter elements, as well as to the central role of the core promoter composition, in the regulation of transcription initiation and gene expression.

**The mechanism of SCPs activity**

Aiming to discover the mechanism that grants the full-length core promoters their potent properties, we investigated whether DNA methylation and/or histone/protein acetylation play a role in it.

Initially, since DNA hypermethylation is a known and common reason for promoter silencing, we examined if DNA methylation is involved in the mechanism. To this end, we treated the cells with 5’-Aza-dC, which inhibits the activity of DNA methyl-transferases, and thus
prevents DNA methylation. The results of the 5′-Aza-dC treatment indicate that DNA methylation is probably not involved in the potency of the full-length promoters.

Another modification that may be involved in the mechanism is histone acetylation, which is usually correlated with transcriptionally active promoters. In order to investigate the role of histone acetylation we treated the cells with TSA, which inhibits histone deacetylase (HDAC), thus resulting in histone hyperacetylation that leads to chromatin relaxation. The TSA treatment resulted in the reduction of differences in transcriptional activity between the promoters. Thus, it can be assumed that changes in histone acetylation patterns are involved, in a direct or indirect manner, in the mechanism that grants the full-length core promoters their strong properties.

TSA has also been implicated in the induction of DNA demethylation\textsuperscript{40}. Our findings, however, (of 5′-Aza-dC, TSA and the combined treatments) support the assumption that it is the HDAC inhibitory activity of TSA that contributes to the potency of the full-length core promoters.

A possible explanation for the connection between the full-length core promoters and histone acetylation patterns, is that the core promoter motifs comprising these promoters (most likely the downstream motifs), are able to recruit some factors, which maintain histones at the promoter region in a hyperacetylated state. These factors may have a HAT activity or may recruit other histone acetyltransferases (HATs), and thus induce hyperacetylation in a direct manner. Alternatively, these factors may induce the hyperacetylation indirectly by blocking the access of HDACs to the promoter region.

As transcription factors are regulated, among others, by acetylations and deacetylations\textsuperscript{42,47,48}, and since the deacetylation inhibition activity of TSA is not only specific to histones, but also to cellular proteins such as transcription factors, another explanation may be suggested. It is possible that the reduction of differences between the promoters, induced by the TSA treatment, is due to changes in acetylation of transcription factors and not histones. If this is the case, then one can speculate an example in which the full-length
core promoters are able to recruit, through the downstream sequences, a specific transcription factor in its unacetylated state, whereas the incomplete promoter would only be able to recruit the same factor when it is acetylated.

It is important to note, that despite the fact that hyperacetylation is generally correlated with transcriptionally active promoters, not all promoters are activated upon TSA treatment. Some promoters (such as those of Cyclin D1, Cyclin A, Cyclin E, MMTV, Hmga2 and SRC) respond in an opposite manner, hence HDAC inhibitors suppresses their activity.\textsuperscript{43,49}

Taken together, we assume that the core promoter composition, and particularly the downstream sequences, is vital to the way in which TSA treatment affects transcriptional activity. It is very likely that the mechanism granting the potency to the full-length core promoters includes both the structure of the promoter and specific acetylation patterns.

\textbf{The involvement of TAF1 in regulating the activity of SCPs}

The results obtained by the TSA treatment led us to try and find a link between TSA, acetylation and core promoter composition. A suitable candidate who meets these criteria is TAF1.

In order to investigate the potential role of TAF1 in the mechanism of the full-length core promoters, we used the ts13 cell line. This cell line is temperature sensitive, so that at the permissive temperature (33.5°C) TAF1 has wild type HAT activity, while at the restrictive temperature (39.5°C) the HAT activity is disrupted.

The results obtained at the restrictive temperature, display a substantial reduction of the differences between the promoters – this is of high resemblance to the results of the TSA treatment. Consequently, it is likely that a correlation between the HAT activity of TAF1, TSA and core promoter architecture exists.

According to a study done by the Wang lab\textsuperscript{43}, core promoters can be divided to groups by their response to HDAC inhibitors, as well as by their ability to function properly upon
disruption of TAF1 activity. Promoters that function properly without TAF1 may be referred as TAF1-independent promoters, whereas promoters that require TAF1 proper function may be referred as TAF1-dependent promoters. In this study using the ts13 cell line, the authors have shown that transcription driven by TAF1-dependent promoters is repressed as a result of TSA treatment, whereas the activity of TAF1-independent promoters is induced after such treatment. Furthermore, the authors found that TAF1-dependent promoters (TSA-repressed) lack the TATA box element, while the TAF1-independent promoters (TSA-induced) are characterized by the existence of a distinct functional TATA box element. Finally, this study indicates that the core promoter architecture, as well as some proximal promoter sequences, dictates the identity of the promoter, as TAF1-dependent or –independent, as well as the impact of TSA treatment on the activity of the promoter.

In a follow-up study, it was discovered that deficient HAT activity of TAF1 leads to reduction of histone H3 acetylation at the TAF1-dependent promoter of Cyclin D1, but not at the TAF1-independent promoter of the C-fos gene. Moreover, hypoacetylation of H3 at the Cyclin D1 promoter can be reversed by treatment with TSA, thus restoring the full activity of the Cyclin D1 gene. These findings provide evidence for a possible role of histone acetylation by TAF1 in the activation of gene transcription. If so, it appears that the activities of both TSA and TAF1 are equivalent in the dynamic equilibrium among HATs and HDACs, hence it is likely that their mechanism of action includes the same promoter sequence elements.

A third study indicates that in ts13 cells line, at the restrictive temperature, not only the HAT activity of TAF1 is disrupted, but also its ability to bind the Inr element and recruit TFIID. This data was obtained from investigation of the Cyclin D1 gene, which is regulated by a TAF1-dependent promoter, hence a TATA box lacking promoter. These results suggest that TAF1 plays a critical role in the efficient recruitment of TFIID through an Inr element, in TAF1-dependent promoters. Interestingly, upon insertion of a consensus TATA box motif to the promoter, the recruitment of TFIID was once again proper and basal transcription activity was restored. Based on these results, the authors propose the following model for the
function of TAF1. TAF1-independent promoters contain a TATA box element such that the primary recruitment of TFIID is mediated by the binding of TBP to the TATA box, and thus disrupted TAF1 activity does not affect TFIID recruitment. On the other hand, TAF1-dependent promoters lack a TATA box element, so the recruitment of TFIID depends on the interaction of TAF1 with the Inr.

In light of all of these findings, it will be interesting to examine the effect of additional combinations of core promoter elements on both the basal and activated transcription. We have constructed several Cyclin D1-based promoters that differ by the combinations of core promoter motifs (these constructs were initially designed for the project of characterization of the influence of core promoter elements on transcription activity levels and kinetics driven by modified core promoters). Transfection of these core promoters to the BHK-21 and ts13 cell lines, will enable us to investigate the Cyclin D1 expression driven by the various core promoters, using Real Time PCR analysis.

Due to time constraints, in the past year most efforts were devoted to the first and third goals of this work (indeed we have recently submitted an article describing our findings towards the first goal of characterization of gene expression under the regulation of SCPs). Thus, the above Real Time PCR experiments are yet to be performed.

Characterization of the influence of Core Promoter Elements on Transcription Activity and Kinetics

In this section we utilized an in vivo system, developed by Prof. Yaron Shav-Tal and Prof. Yuval Garini, which allows the detection of transcription activity from a single gene, thus providing the ability to analyze the activity levels and kinetics of gene expression at single-cell, single-gene and single-allele levels. The strength of this system lies in the fact that it eliminates position effects, so that observed changes in transcription activity may result only from the investigated parameter. In our study, this parameter is the core promoter composition.
Our aim is to investigate both the individual as well as the combined contribution of each core promoter element on promoter driven transcriptional activity and kinetics. To this end, we generated 2 sets of stable cell clones, each contain 4 clones that express the Cyclin D1 gene under the regulation of different core promoters. In the first set of cell clones, Cyclin D1 is regulated by promoters that are based on the CMV promoter, including SCP2 and SCP3. These cell clones are intended to provide further insight (in addition to that described in the first two parts of the Discussion) into the activity of the SCPs. The second set of cell clones is comprised of four clones in which the Cyclin D1 is regulated by promoters based on its own endogenous promoter. The reason for choosing the promoter of Cyclin D1 as the basis of this set, is that this promoter was used earlier (in the same system) by Dr. Sharon Yunger, and because its core promoter comprise both the Inr and DPE (compatible with the consensus) motifs. Moreover, as the promoter of Cyclin D1 is TAF1-dependent and lacks a TATA box motif, this promoter as well as its variants are extremely relevant to the investigation of TAF1, in the ts13 cell line, described earlier.

In an RNA Fluorescence *In Situ* Hybridization (FISH) experiment that was conducted as one of the controls for the accurate single allele integration, it was possible to detect an active transcription site in the vast majority of the cells from the clones based on the CMV promoters (first set of cell clones). However, in cell clones that are based on the Cyclin D1 promoter (second set of cell clones), an active transcription site could be observed only in part of the cells. These different observations are due to the fact that the CMV promoter-driven transcription (and thus its variants) is constantly active as it is a strong viral promoter, whereas the Cyclin D1 endogenous promoter-driven transcription (and thus its variants) is not constant, so the transcription site fluctuates between "on" and "off" states. Importantly, observations of the four cell clones that are based on the endogenous Cyclin D1 promoter exhibit a stronger transcription activity in favor of the three variants over the endogenous promoter. These advantages were reflected by strong brightness of the transcription site as
well as by high number of detectable mRNAs. Nevertheless, these are only qualitative preliminary observations that may be indicative of future quantitative results.

Up to this point, we conducted quantitative analysis by a quantitative RNA Fluorescence In Situ Hybridization (FISH) experiment on the endogenous CCND1, Enhanced, Super and Super Duper cell clones (the four clones of which Cyclin D1 is regulated by promoters based on its own endogenous promoter). We have so far acquired more than 25 successful results with the Endogenous CCND1 promoter (which includes the Inr and DPE motifs) and the Super promoter (which includes the TATA box, Inr and DPE motifs) cell clones. The analysis of the Enhanced and Super Duper cell clones is still in progress.

The Endogenous CCND1 cell clone we generated is identical to that previously examined by Dr. Sharon Yunger\textsuperscript{44}, thus the compatibility of our results and the previously published results confirm the proper implementation of the method and the accuracy of our results. The results obtained so far reveal that the transcription activity driven by the Super promoter is much higher than that driven by the endogenous CCND1 promoter, hence the addition of the TATA box element greatly improves the activity of the promoter.

The preliminary results obtained from the Enhanced and Super Duper cell clones indicate that the transcriptional activity driven by both of these promoters is higher than that of the Super promoter. As both the Enhanced and Super Duper promoters contain the MTE and DPE motifs, while the other two promoters lack the MTE motif, these results indicates that the DPE functions in a preferential manner with the MTE motif, thus supporting the previous findings of the strong synergistic cooperative function of the MTE and DPE motifs\textsuperscript{15}.

Additionally, since it appears that the Enhanced promoter is stronger than the Super promoter it may be postulated that the contribution of the TATA box motif to promoters-driven transcription is not as pronounced as the contribution derived from the cooperative function of the MTE and DPE motifs. It is too early to estimate whether the TATA box motif strengthens, weakens or has no affect on promoters that contain both the MTE and DPE
motifs. Nevertheless, future analyses of the Enhanced and Super Duper promoters will allow us to address the matter.

In addition to the completion of the quantitative analysis of the Enhanced and Super Duper cell clones, as well as of the four cell clones in which Cyclin D1 is regulated by promoters based on the CMV promoter (the second set), we would also like to address to the kinetics of transcription driven by these eight promoters in future experiments. Analyses of transcription kinetics will be performed by live cell imaging and by FRAP experiments. Live cell imaging will allow us to follow the real-time gene activity patterns driven by the various promoters, which will be detected as a specific fluorescent locus in the nucleus of the cells. This approach will enable us to address issues such as whether the mode of transcription driven by a specific promoter is constitutive or bursting, and whether a promoter activity is ordered or stochastic. FRAP (Flourescence Recovery after Photobleaching) experiments will provide a proof that the active transcription sites detected by microscopy are indeed actively transcribing. By photobleaching of the transcription site, which is visible as a green spot, the fluorescence signal of the MS2-GFP proteins (associated with transcribed mRNA) decreases and essentially blackened. If the transcription site is indeed active, then new mRNAs will be transcribed and bound by unbleached MS2-GFP proteins, resulting in a gradual recovery of the green spot that can once again be detected. The fluorescence recovery rate of the MS2-GFP proteins is proportional to the activity rate of RNA pol II, thus the recovery kinetics can provide information about the elongation rate of RNA polymerase II.
Methods

Cell culture

HeLa S3 and SH-SY5Y cells were cultured in DMEM/F12 (Biological Industries) supplemented with 10% FBS and grown at 37°C with 5% CO2.

HOP-92 cells were cultured in RPMI-1640 (Biological Industries) supplemented with 10% FBS and grown at 37°C with 5% CO2.

BHK-21 cells were cultured in DMEM/F12 (Biological Industries) supplemented with 10% FBS and grown at 37°C for maintenance, and at 33.5°C or 39.5°C for experiments, with 5% CO2. The cells were received from Dr. Marcelo Ehrlich lab’s (Department of Cell Research and Immunology, Tel Aviv University).

ts13 cells were cultured in DMEM/F12 (Biological Industries) supplemented with 10% FBS and grown at 33.5°C for maintenance or experiments, and at 39.5°C for experiments, with 5% CO2. The cells were received from Prof. Rivka Dikstein lab’s (Department of Biological Chemistry, The Weizmann Institute of Science).

HEK 293 flp- in cells were cultured in DMEM High-Glucose (Biological Industries) supplemented with 10% FBS and grown at 37°C with 5% CO2. The cells were received from Prof. Yaron Shav-Tal lab’s.

Transfection

HeLa S3 or SH-SY5Y cells were plated in 24-well plates one day prior to transfection. Cells were transfected with the various promoter–EGFP constructs by using the TransFast reagent (Promega) according to the manufacturer’s instructions. HeLa S3 cells were transfected with 1µg and SH-SY5Y cells were transfected with 0.75µg of each of the different promoter–EGFP constructs.

HOP-92 cells were plated at 10 x 10⁴ cells per well in 24-well plates one day prior to transfection. Cells were transfected with the various promoter–EGFP constructs by using the Lipofectamine-2000 reagent (Life Technologies) according to the manufacturer’s
instructions. Prior to transfection the growth medium was replaced with fresh supplemented medium. Cells were transfected with 1µg of each of the different promoter–EGFP constructs. BHK-21 and ts13 cells were plated at 7.5 x 10^4 or 50 x 10^4 cells per well in 24-well plates or 60mm plates (respectively) one day prior to transfection. Cells were transfected with the various promoter–EGFP constructs by using the Lipofectamine-2000 reagent (Life Technologies) according to the manufacturer’s instructions. Prior to transfection, the growth medium was replaced with an empty medium (without FBS and antibiotics). Transfections were conducted in duplicates, and the transfected cells were grown at 33.5°C during 3 h post-transfection. After 3 h of incubation, 10% FBS as well as 1% antibiotics (Pen/Strep) were added to the medium, and one repetition of each duplicate was transferred to incubation at 39.5°C while the other repetition was kept incubated at 33.5°C. Cells were transfected with 1µg in 24-well plates and 3µg in 60mm plates of each of the different promoter–EGFP constructs.

HEK 293 Flp-In cells were plated in either 60mm or 100mm plates one or two days prior to transfection, so a confluence of 60-80% is obtained upon transfection. Cells were transfected with the various promoter–EGFP constructs by using the Calcium-Phosphate reagent. Prior to transfection the growth medium was replaced with fresh supplemented medium 25 µM Chloroquine (in order to increase transfection efficiency). Six to eight hours following transfection, the growth medium containing chloroquine was replaced with fresh supplemented medium. Cells were transfected with 3µg DNA per 60mm plates and 10µg DNA per 100mm plates, of each of the different promoter–EGFP constructs.

Live cell imaging

For live cell imaging, HeLa-S3 and SH-SY5Y cells were transfected with the different promoter–EGFP constructs as described, and incubated for 4 or 8 days. Cells were imaged every 24 h, starting 24 h after transfection for days 1-4, and 98 h after transfection for days 4-8 (imaging of days 1-4 and 4-8 was performed in separate experiments, in order to prevent cell overcrowding). Images were acquired using the Zeiss Observer Z1 inverted microscope
with a 5X ECPlan-Neofluar objective, equipped with a 37°C 5% CO₂ incubation chamber. In order to image a whole well, images were acquired with Zeiss AxioVision software’s “titles” function, allowing stitching images corresponding to a field of a complete well. Resulting images were processed using Adobe Photoshop software.

**Flow cytometry**

FACS analyses for short- and long-term experiments were performed separately (1-4 and 4-8 days post-transfection, respectively, so that day 4 is common to both analyses), to prevent cell overcrowding. Cells were transfected and incubated according to specific experiment as described. At the indicated time points, cells were harvested, centrifuged at 1000 rpm, resuspended in 0.3 ml PBS and subjected to flow cytometry analysis (Gallios, Beckman Coulter). The fluorescence of EGFP was measured in 10,000 cells per sample. The data was analyzed using the FlowJo software.

**Generation of stable cell clones**

Stable cell clones were generated according to the previously published protocol by Prof. Yaron Shav-Tal and Dr. Sharon Yunger⁴⁵. Briefly, HEK 293 Flp-In cells were either transfected by the Calcium-Phosphate reagent or by electroporation with a Gene pulserXcell, Biorad/ Neon™ Transfection System. Stably expressing HEK 293 Flp-In cells were established by co-transfection of a pcDNA5/FRT vector (containing an FRT site as well as the Cyclin D1 gene under the regulation of the various promoters) with the pOG44 plasmid (Flp-recombinase expression plasmid) using a ratio of 9µg : 1µg –of pOG44 : pcDNA5/FRT, followed by selection of at least 3 weeks (starting 24 h post-transfection) with 100 µg/ml Hygromycin B.
**X-gal staining (lacZ assay)**

The eight cell clones, as well as the HEK 293 Flp-In cells (as control) were grown in 24 well plates for 24 h. After 24 h, the cells were washed in PBS, fixed with 3.7% formaldehyde for 5 minutes, washed again and incubated in X-gal solution (containing 5mM K$_3$Fe(CN)$_6$, 5mM K$_4$Fe(CN)$_6$, 2mg MgCl$_2$ and 1mg/ml X-gal) at 37°C for at least two hours.

**PCR amplification**

Genomic DNA was extracted from the eight cell clones as well as from the HEK 293 Flp-In cells (as control) using the Archive Pure DNA Cell/Tissue kit (5 PRIME). The genomic DNA was used as a template for two separate PCRs. The first PCR reaction was done with a forward SV40 primer (F1 primer) and a reverse Hygromycin-resistance gene primer (R1 primer). The second PCR reaction was done with a forward BGH polyA primer (F2 primer) and a reverse lacZ-Zeocin gene primer (R2 primer). PCR products were detected in an agarose gel and the amplified DNA was extracted using the Nucleo Spin PCR clean-up Gel extraction kit (Macherey Nagel). The amplified DNA was sequenced using the same primers of the PCR reactions.

**Primers:**

- **SV40 sense (F1 primer):** 5′-CCAGTTCCGCCCATCTCC-3′
- **Hygromycin antisense (R1 primer):** 5′-CTGTATGCGGCCATTGTCC-3′
- **BGH pA sense (F2 primer):** 5′- CGA GTC TAG AGG GCC CGT TTA AAC -3′
- **lacZ-Zeocin antisense (R2 primer):** 5′- GTA ACC GTG CAT CTG CCA GTT TG -3′

**RNA Fluorescence In Situ Hybridization (FISH)**

The FISH experiments were performed as one of the verifications of the proper generation of the stable cell clones. The experiment was conducted is a similar manner as the Quantitative RNA Fluorescent In Situ Hybridization (FISH) that is described below.
Quantitative RNA Fluorescence \textit{In Situ} Hybridization (FISH)

Quantitative RNA FISH was performed according to the previously published protocol by Prof. Yaron Shav-Tal and Dr. Sharon Yunger\textsuperscript{45}. Briefly, HEK 293 Flp-In cells (as control) and the eight cell clones were grown on round coverslips and fixed in 4\% PFA for 20 min. After fixation, cells were washed in PBS and kept in 70\% ethanol at 4°C overnight. On the next day, cells were washed in PBS for 20 min, permeabilized with 0.5\% Triton X-100 and washed in 40\% formamide for 10 min. Hybridization with the MS2-Cy3 probe was done overnight at 37°C. On the following day cells were washed in 40\% warm formamide for 30 min and then washed for 2 h in PBS. Following the 2 h wash, cells were stained with a nuclear staining dye (Hoechst) and mounted on appropriate slide.

3D stacks images of total cell volumes were acquired using a wide-field fluorescence microscope (Zeiss Observer Z1 inverted microscope) with a 100X Plan-Apochromat objective. Each cell’s total volume was imaged by 76 slices with a step width of 0.2 µm.

Deconvolution

Deconvolution was done by the Huygens commercial software.

Quantitative RNA FISH analyses

Visualization and analysis of deconvolved multidimensional microscopy images was performed using the Imaris software (Bitplane).

Statistical analysis

- pRC/CMV based constructs

  The fluorescence intensity and number of cells expressing \textit{EGFP} for each promoter in each experiment, was compared to the pRC/CMV data obtained in the corresponding experiment (by division of the obtained absolute value by that of the pRC/CMV promoter). Statistical comparisons between the promoters were done using the non-parametric Kruskal-Wallis test for independent samples, with pairwise comparisons. The Kruskal-Wallis test was chosen since it is a non-
parametric test, (which does not require the assumption that the samples are
distributed normally) and since it can be applied to examine groups of unequal
size. A \( p \)-value \( \leq 0.05 \) was considered to be statistically significant.

- **pEGFP-N1 based constructs**
  
The fluorescence intensity and number of cells expressing \( EGFP \) for each
promoter in each experiment, was compared to the pEGFP-N1 data obtained in
the corresponding experiment (by division of the obtained absolute value by that of
the pEGFP-N1 promoter). Statistical comparisons between the promoters were
done using the ANOVA test, with a Bonferroni comparison for multiple
comparisons. Statistical significance was tested only in experiments that included
at least three repetitions. A \( p \)-value \( \leq 0.05 \) was considered to be statistically
significant.
References


3 Subramanian, C., Jarzembowski, J. A., Opipari, A. W., Jr., Castle, V. P. & Kwok, R. P. CREB-binding protein is a mediator of neuroblastoma cell death induced by the histone deacetylase inhibitor trichostatin A. *Neoplasia* **9**, 495-503.(2007)


Burke, T. W. & Kadonaga, J. T. The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. *Genes & development* **11**, 3020-3031.(1997)


41 Chiurazzi, P. *et al.* Synergistic effect of histone hyperacetylation and DNA demethylation in the reactivation of the FMR1 gene. *Human molecular genetics* **8**, 75


Appendices

Appendix 1.

<table>
<thead>
<tr>
<th>ENDOGENOUS</th>
<th>ENHANCED</th>
<th>SUPER</th>
<th>SUPER DUPER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inr, DPE</td>
<td>Inr, MTE, DPE</td>
<td>TATA, Inr, DPE</td>
<td>TATA, Inr, MTE, DPE</td>
</tr>
<tr>
<td>mRNAs per transcription site</td>
<td>mRNAs per cell</td>
<td>mRNAs per cell</td>
<td>mRNAs per transcription site</td>
</tr>
<tr>
<td>13.61649167</td>
<td>59</td>
<td>27.34009539</td>
<td>117</td>
</tr>
<tr>
<td>12.20733416</td>
<td>125</td>
<td>26.0977366</td>
<td>198</td>
</tr>
<tr>
<td>12.05487032</td>
<td>79</td>
<td>23.7380422</td>
<td>171</td>
</tr>
<tr>
<td>11.98653233</td>
<td>69</td>
<td>20.22445603</td>
<td>83</td>
</tr>
<tr>
<td>10.89340236</td>
<td>124</td>
<td>20.05707381</td>
<td>117</td>
</tr>
<tr>
<td>10.35832819</td>
<td>81</td>
<td>18.53864825</td>
<td>169</td>
</tr>
<tr>
<td>10.34270888</td>
<td>80</td>
<td>10.83105029</td>
<td>55</td>
</tr>
<tr>
<td>9.61636599</td>
<td>79</td>
<td>9.06603008</td>
<td>60</td>
</tr>
<tr>
<td>9.294892707</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.25013979</td>
<td>75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.567980411</td>
<td>105</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.39856086</td>
<td>73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.050582152</td>
<td>75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.637762433</td>
<td>73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.481027017</td>
<td>104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.172013331</td>
<td>73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.914392242</td>
<td>89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.34585705</td>
<td>86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.29452054</td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.03682351</td>
<td>95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.827036393</td>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.38984846</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.392229442</td>
<td>84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.19900756</td>
<td>89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.955773917</td>
<td>113</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.903411284</td>
<td>108</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.81987697</td>
<td>103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.80618738</td>
<td>73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.03271616</td>
<td>146</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.00709588</td>
<td>156</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Quantification of mRNA molecules at the cell and the transcription site of cell clones based on Cyclin D1 endogenous promoter

Quantitative RNA FISH experiments were imaged by 3D stacks of 76 slices. 3D stacks series were then subjected to deconvolution by the Huygens commercial software. Deconvolved 3D stacks are processed by the Imaris software providing a measured fluorescent intensity value for each individual mRNA. All measured values (of individual cell) are plotted as a frequency histogram to provide the most probable fluorescent intensity value of a single mRNA. The number of mRNAs located at the transcription site is calculated according to the fluorescent value of a single mRNA. Red values - repetitions we assume might become irrelevant after the completion of the analysis, due to atypical number of mRNAs at the transcription site. Blue values - repetitions we assume might become irrelevant after the completion of the analysis, due to the
atypical number of mRNAs at total volume of the cell. **Red & Blue values**- repetitions we assume to be irrelevant due to atypical number of mRNAs at both the transcription site and the total volume of the cell.

**Appendix 2.**

The study of SCP3, which has been submitted for publication to *Scientific Reports*.

**Engineered Promoters for Potent Transient Overexpression**

Dan Even*¹, Shani Basch- Barzilay*¹, Diana Ideses¹, Ravid Tikotzki¹, Orit Shefi² and Tamar Juven-Gershon¹

¹ The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, 5290002, Israel

²Faculty of Engineering and Bar-Ilan Institute of Nanotechnologies and Advanced Materials, Bar-Ilan University, Ramat-Gan 5290002, Israel

*These authors contributed equally to this work

Correspondence should be addressed to T. J-G. (tamar.gershon@biu.ac.il)

**Abstract**

Transcription initiation by RNA polymerase II is critical for gene expression, thus its regulation is of utmost importance to physiological and pathological processes. The core promoter, which is generally defined as the region to which RNA Polymerase II is recruited to initiate transcription, consists of different combinations of several short DNA sequences, termed core promoter elements or motifs (*e.g.* the TATA-box, initiator and downstream core promoter element (DPE)), which confer specific functional properties to each promoter. The
core promoter has previously been demonstrated to play a pivotal role in the regulation of gene expression. Earlier studies that examined the ability to modulate gene expression levels via the core promoter, led to the design of strong synthetic core promoters, termed super core promoters (SCPs), which combine different core elements into a single core promoter. Here, we engineered EGFP expression vectors driven by distinct core promoters. We use live cell imaging and FACS analyses to demonstrate the unusually strong long-term EGFP expression by SCPs, particularly by the novel SCP3. Our data provides further support for the importance of downstream core promoter elements in the regulation of transcription, and demonstrates for the first time the feasibility of long-term expression in transiently transfected cells.

**Introduction**

The transcription of protein-coding genes is a pivotal process underlying proper cellular function. The accurate initiation of transcription by RNA polymerase II (Pol II) is a critical step in the regulation of gene expression, in which Pol II is recruited to the core promoter via the basal transcription machinery (for a review, see'). The core promoter is typically 80 nucleotides long, encompassing from -40 to +40 relative to the transcription start site. In the past, it was presumed that all core promoters function via a single universal mechanism, however, this is now well established that core promoters differ in both structure and function’. The core promoter consists of several functional subregions, termed core elements or motifs. Some of the known core promoter elements are the TATA box, TFIIB recognition elements (BREu and BREd), DCE, initiator (Inr), TCT, MTE and DPE’. The TATA box, which is recognized and bound by the TBP subunit of the TFIID complex, is the best-known element’. The Inr motif is probably the most common element, and it encompasses the transcription start site which is designated as the A nucleotide in the Inr consensus (usually designated as position “+1”)’. Both the MTE and DPE motifs are located downstream of the Inr and serve as recognition sites for the TAF6 and TAF9 subunits of TFIID’. The MTE and DPE function in a cooperative manner with the Inr, and a precise
spacing between the Inr and each of these motifs is crucial for transcriptional activity\textsuperscript{18-21}.

There are no universal core promoter motifs and different combinations of motifs confer specific functional properties to the core promoter, \textit{e.g.} the ability to function in concert with specific enhancers\textsuperscript{23-26} or regulate developmental gene regulatory networks\textsuperscript{27,28}. Moreover, there are core promoters that lack any of the known core motifs, suggesting the existence of additional unknown core promoter elements that remain to be discovered.

Previous studies of the MTE motif revealed that although the MTE can function independently of the TATA box and DPE motifs, its collaboration with the TATA box, as well as with the DPE, results in strong synergy\textsuperscript{18}. This synergy raised the idea that by combining different core promoter elements into a single core promoter, it could be possible to design unusually strong synthetic core promoters. We have previously designed synthetic core promoters, designated Super Core Promoters (SCP1 and SCP2), which contain the TATA box, Inr, MTE and DPE elements that drive high levels of transcription both \textit{in vivo} and \textit{in vitro}\textsuperscript{29}. To date, no natural promoters that contain such a combination of core promoter elements in a single promoter have been identified.

In this study, we constructed \textit{EGFP} expression vectors driven by distinct core promoters, including the novel super core promoter (SCP3). We analyzed the expression driven by these core promoters both by live cell imaging and by FACS analysis, in two types of human cell lines over 8 days. Our results provide additional support for the importance of core promoter elements that are located downstream of the initiator for gene expression and demonstrate for the first time, robust and long-term expression of SCP-driven reporter gene expression in transiently transfected cells.

**Results**

**Generation of a new synthetic core promoter**

Synthetic core promoters, termed super core promoters (SCP\textsubscript{s}), which combine different promoter core elements into a single core promoter, have previously been shown to enable high levels of gene expression\textsuperscript{29}. The highest transcription activity observed in these studies
was by using the SCP2 in concert with the CMV enhancer. In order to explore the temporal dynamics of SCPs function in conjunction with the CMV enhancer, we designed a new super core promoter, based on the previously designed SCP2, termed SCP3. Similarly to SCP2, SCP3 contains the CMV TATA and Inr motifs, along with the *D.melanogaster Tollo* MTE\(^\text{18}\) and the human *Calm2* DPE\(^\text{30}\). Unlike SCP2, SCP3 does not contain the T7 promoter that is downstream of +45 relative to the +1 transcription start site (to position the *EGFP* reporter gene immediately downstream of the promoter). Additionally, SCP3 includes four nucleotide changes as compared with SCP2: to improve the Inr sequence (A to T in position +3 and T to C in position +4), improve the sequence adjacent to the MTE (G to C in position +16; based on single nucleotide analysis of the *Tollo* MTE\(^\text{19}\)) and improve the DPE (T to C in position +31; as in many DPE driven genes as well as in the MTE-dependent *Tollo* core promoter\(^\text{18,19,22}\) (Fig. 1). We designed and constructed four pRC/CMV-based expression vectors, each driving the expression of the *EGFP* reporter gene, in order to test the effects of the different core promoters on transcriptional activity: the incomplete CMV (pRC/CMV), the complete CMV (natural CMV), the SCP2\(^\text{29}\), and the novel SCP3.

**Live cell imaging reveals strong *EGFP* expression under the regulation of SCP3**

In a previous study\(^\text{29}\), a distinct increase in transcriptional activity was demonstrated using engineered core promoters, favoring the natural CMV and SCP2 core promoters in conjunction with the CMV enhancer, relative to the pRC/CMV core promoter. These studies utilized the chloramphenicol acetyltransferase (CAT) and luciferase reporter genes, and cells were assayed 24-48 h post-transfection for CAT and luciferase activities. Unlike these assays that require harvesting of the cells at a specific time point, here we used the *EGFP* reporter gene that enables us to qualitatively follow the activity of the various core promoters in the same population of transfected living cells over time.

To assess the transcription activity of the core promoters, we transiently transfected two types of human cell lines, HeLa S3 (human cervical carcinoma cells) and SH-SY5Y
(human neuroblastoma cells originating from bone marrow tissue), with the pRC/CMV, natural CMV, SCP2 or SCP3-driven EGFP-expression vectors, and imaged the fluorescence signals expressed by the transfected cell populations. Live cell imaging for short-term expression, performed during days 1-4 post-transfection, reveals that full-length core promoters have a substantial advantage over the commercial pRC/CMV core promoter, both in terms of fluorescence intensity and number of fluorescent cells (Fig. 2). During the 4 days after transfection, an increase in the fluorescence levels, which is indicative of the transcriptional activity, can be observed in all four promoters, though only slightly in the pRC/CMV promoter. Notably, SCP3 is consistently the most active core promoter.

To examine whether these promoters drive long-term expression, live cell imaging was performed during days 4-8 post-transfection. Remarkably, a distinct increase in activity was observed during days 4-8 post-transfection using the complete natural CMV, SCP2 and especially SCP3, as compared to the commercial pRC/CMV core promoter (Fig. 3). As expected of transient transfections, we observed a decrease in fluorescence intensity during the long term. However, in days 7 and 8, there are still relatively high fluorescence intensities driven by the full-length core promoters, and particularly the SCP3 core promoter. It is of note that transiently transfected cells express EGFP for longer time periods. Some SCP3-driven EGFP SH-SY5Y cells remain fluorescent even 3 weeks post-transfection (data not shown).

**Quantitative FACS analysis reveals strong EGFP expression under the regulation of SCP3**

To investigate the transcription activity of the various core promoters in a quantitative manner, we performed FACS analysis. HeLa S3 and SH-SY5Y cells were transiently transfected with the pRC/CMV, natural CMV, SCP2 or SCP3-driven EGFP-expression vectors, and analyzed by FACS on a daily basis for 8 days post-transfection (divided into short-term (d1-4) and long-term (d4-8) follow-up). To quantitatively measure the reporter
levels of transcription activity and the number of cells expressing the reporter under the
regulation of the four core promoters, we examined both the fluorescence intensity and the
number of cells expressing EGFP. A non-parametric Kruskal-Wallis test was performed to
examine the statistical differences between any pair of promoters. As the fluorescence
intensity and the number of EGFP-expressing cells differ between high fluorescence
expressors and low fluorescence expressors, we analyzed the cells that express high
fluorescence levels (designated “HIGH EXP”) in addition to analyzing the entire population of
fluorescent cells (designated “ALL EXP”). The FACS analyses presented in Figs. 4-5 depict
representative experiments, whereas the statistical significance was calculated for 5-6
individual experiments for each cell line for short- as well as long-term expression.

FACS analysis of short-term EGFP-expression reveals a constant advantage in favor of the
SCP3 promoter (Fig. 4). This advantage is reflected in both HeLa S3 and SH-SY5Y cell
lines, by the fluorescence intensity levels of all fluorescent cells. Analysis of the fluorescence
intensity of all HeLa S3 fluorescent cells reveals that SCP3 is significantly stronger than the
pRC/CMV promoter during all four days (n = 5, d1= p ≤ 0.022, d2= p ≤ 0.002, d3= p ≤ 0.013,
d4= p ≤ 0.003). Similarly, analysis of all EGFP-expressing SH-SY5Y cells, reveals that SCP3
is significantly stronger than pRC/CMV in days 1 and 2 post-transfection (n = 6, d1= p ≤ 0.000, d2= p ≤ 0.002) (Fig. 4a). Overall, we did not observe significant difference in the
fluorescence intensity driven by either promoter when we analyzed the high expressors in
both cell lines (Fig. 4b). The advantage of SCP3 is also reflected in the number of all EGFP-
expressing SH-SY5Y cells, where significant differences between SCP3 and pRC/CMV were
observed during days 1, 2 and 4 post-transfection (n = 6, d1= p ≤ 0.027, d2= p ≤ 0.007,d4=p
≤ 0.002) (Fig. 4c). The analysis of the number of high fluorescence intensity HeLa S3 and
SH-SY5Y cells, reveals significant advantage favoring SCP3 over the pRC/CMV promoter, in
all four examined days (HeLa S3: n = 5, d1= p ≤ 0.050, d2= p ≤ 0.026, d3= p ≤ 0.009, d4= p ≤ 0.011; SH-SY5Y: n = 6, d1= p ≤ 0.008, d2= p ≤ 0.001, d3= p ≤ 0.045, d4= p ≤ 0.011) (Fig. 4d).

In addition to the advantage of SCP3, short-term analysis also reveals a significant
advantage in favor of SCP2-driven transcription compared to the pRC/CMV promoter. This
advantage is observed by the fluorescence intensity levels of all fluorescent HeLa S3 cells during days 2 and 4 (n = 5, d2= p ≤ 0.015, d4= p ≤ 0.018) (Fig. 4a). In terms of the number of EGFP-expressing HeLa S3 cells, we observed temporal variations upon comparison of SCP2 to pRC/CMV, however in all EGFP-expressing SH-SY5Y cells, SCP2 expression is significantly stronger in days 2 and 4 (n = 6, d2= p ≤ 0.035, d4= p ≤ 0.035) (Fig. 4c,d).

Remarkably, FACS analysis of long-term expression (performed during days 4-8 post-transfection), reveals an advantage of SCP3 over the pRC/CMV promoter in both HeLa S3 and SH-SY5Y cells. SCP3 exhibits a significant advantage over the pRC/CMV promoter in the entire population of fluorescent HeLa S3 cells in days 4-7 (n = 6, d4= p ≤ 0.002, d5= p ≤ 0.002,d6= p ≤ 0.024,d7= p ≤ 0.003) (Fig. 5a). The fluorescence intensity of all EGFP-expressing SH-SY5Y cells also reveals stronger transcriptional activity of the SCP3 promoter as compared to pRC/CMV, in days 4, 5 and 8 (n = 5, d4= p ≤ 0.007,d5= p ≤ 0.004,d8= p ≤ 0.004) (Fig. 5a). Analysis of the fluorescence intensity of HeLa S3 and SH-SY5Ycells that express high levels of EGFP does not indicate constant significant differences among the promoters (Fig. 5b).

SCP3 directs the long-term expression of a larger number of HeLa S3 and SH-SY5Y cells that express EGFP as compared to the pRC/CMV promoter (Fig. 5c,d). This potent activity of SCP3 is observed in days 5-8 of all fluorescent HeLa S3 cells and in days 4, 6 and 8 of all fluorescent SH-SY5Y cells (HeLa S3: n = 6, d5= p ≤ 0.035,d6= p ≤ 0.024,d7= p ≤ 0.002,d8= p ≤ 0.001 ;SH-SY5Y:n = 5, d4= p ≤ 0.009,d6= p ≤ 0.031,d8= p ≤ 0.007) (Fig. 5c). The advantage of SCP3 over pRC/CMV is also observed during days 4, 5 and 7 of the high EGFP-expressing HeLa S3 cells, and days 4, 5, 6 and 8 of the high EGFP-expressing SH-SY5Y cells (HeLa S3: n = 6, d4= p ≤ 0.001,d5= p ≤ 0.001,d7= p ≤ 0.022 ;SH-SY5Y:n = 5, d4= p ≤ 0.003,d5= p ≤ 0.026,d6= p ≤ 0.036,d8= p ≤ 0.028) (Fig. 5d).

SCP2-driven transcription is significantly stronger as compared to the pRC/CMV promoter in terms of fluorescence intensity levels of all EGFP-expressing HeLa S3 cells during days 4, 6 and 7 (n = 6, d4= p ≤ 0.045, d6= p ≤ 0.014,d7= p ≤ 0.045), (Fig. 5a). The advantage of SCP2
is also reflected by the number of EGFP-expressing cells in days 5, 7 and 8 (n = 6, d5= p ≤ 0.018, d7= p ≤ 0.027, d8=p ≤ 0.035) (Fig. 5c).

Taken together, as SCP3 is stronger than the pRC/CMV, whereas the advantage of SCP2 is less pronounced, we suggest that SCP3 is generally more potent than SCP2. Importantly, using the newly designed synthetic promoter SCP3, and in some cases also SCP2, we achieved long-term EGFP expression, as reflected by the fluorescence intensity and the overall number of EGFP-expressing cells, in both HeLa S3 and SH-SY5Y cells.

**Discussion**

The use of SCPs, which combine different core promoter elements into a single core promoter, demonstrate that the core promoter is a regulatory module that can not only affect the initiation of transcription, but also achieve high levels of transcription\(^29\). In this study, we investigated and compared the transcription activity of four core promoters: the pRC/CMV, the natural CMV, SCP2 and the newly engineered SCP3. Qualitative analysis (using live cell imaging) and quantitative analysis (using flow cytometry) revealed a distinct increase in transcriptional activity in favor of core promoters containing both upstream and downstream elements, particularly the SCP3. This advantage is reflected both by high expression levels and by the high number of cells that express EGFP. In addition, by the use of the EGFP reporter gene, we were able to add a new, temporal dimension to the characterization of the SCPs. All experiments were performed using transient transfections and were analyzed during an eight days period following transfection, so that the earlier and the later four days represent short and long-term expression, respectively. Cells were transiently transfected and no antibiotic selection was used, suggesting that the transfected DNA did not integrate into the chromosome and its expression is limited by time, typically considered to be 24-96 h after transfection. It is of note that the fluorescence levels of all EGFP expressing cells substantially decline after 6 days, providing another indication that the plasmids did not stably integrate into the chromosomal DNA. Importantly, we discovered that under the regulation of the full-length
core promoters and particularly SCP2 and SCP3, gene expression is considerably prolonged. Thus, the SCPs are not only unusually strong core promoters with respect to transcription levels, they also provide a novel non-viral way for long-term gene expression in transiently transfected cells. This can be relevant in cases where long-term follow-up is important, such as the response to drugs and cytokines.

A modified core promoter has previously been employed to identify different enhancers and the expression patterns they drive in small subsets of neurons, for manipulation and dissection of neural circuitry\textsuperscript{31}. As enhancers may require specific core promoter elements in order to function properly, it was instrumental to generate transgenic \textit{Drosophila} lines by combining genomic fragments that contain putative enhancers, which are known to have brain-specific expression patterns, with a synthetic SCP-like promoter (DSCP) that is designed to work with a wide range of enhancers. This DSCP is an example to the versatility of designing new SCPs that can function with more than one enhancer. Hence, it is likely that new SCPs would be designed, both in conjunction with enhancers that can either provide a more specific or alternatively, permissive expression.

Interestingly, a recent structural study that used cryo electron microscopy, as well as DNA labeling and footprinting, discovered that the human basal transcription factor TFIID coexists in two distinct structural conformations, where the rearranged state is important for the assembly of the preinitiation complex\textsuperscript{32}. This structural transition is stimulated by the presence of TFIIA that facilitates the binding of TFIID to the core promoter, which occurs by independently interacting with different combinations of the TATA box, Inr, MTE and DPE core promoter motifs. The authors have shown that the core promoter architecture dictates TFIIA-dependent and -independent interactions of TFIID with core promoter DNA. Our results of high level and long-term transcription driven by the SCPs, provide additional support for the central role of the core promoter composition in the regulation of transcription initiation and gene expression.

As SCP3 differs from SCP2 in the absence of the T7 promoter as well as in substitutions of four nucleotides while the core promoter elements composition is the same,
the high transcription activity observed for SCP3-driven EGFP indicates that the entire core promoter sequence, and not just the presence of the core promoter elements themselves, is important for core promoter function. However, since the T7 promoter is absent in the SCP3-driven construct, we cannot rule out the possibility that the high transcriptional activity of SCP3 may result, at least in part, from the proximity of the core promoter to the EGFP reporter gene. Accordingly, optimization of existing SCPs could be utilized to further enhance gene expression. We have successfully used the SCPs in two different types of human cell lines and obtained varying EGFP expression levels in the two cell lines indicating that, as expected, the cell lines used may influence gene expression levels. Importantly, the engineering of core promoter sequences to enhance long-term gene expression in transiently transfected cells might provide a means for future gene-therapy applications.

**Methods**

**Core promoter sequences and plasmids**

EGFP expression plasmids containing the CMV enhancer were constructed using the pRC/CMV vector (Life Technologies) into which the EGFP reporter gene (Clonetech) was subcloned using restriction enzymes. The core promoter region of pRC/CMV, which contains the CMV TATA box and downstream vector sequences, was replaced with CMV, SCP2, or SCP3 sequences from –36 to +45 relative to the A+1 of the transcription start site. The pRC/CMV, natural CMV and SCP2 core promoters have been previously described29. The SCP3 core promoter sequence is:

AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGTCCGCCTGGAGACCTCGAGC
CGAGTGGTCGTGCCTCCATAGAA. Unlike the other three vectors, the SCP3 vector does not contain a T7 promoter between the promoter and the EGFP reporter gene.

**Cell culture and transfection**

HeLa S3 and SH-SY5Y cells were cultured in DMEM/F12 supplemented with 10% FBS and grown at 37°C with 5% CO₂. For flow cytometry analysis and live cell imaging, HeLa S3 or SH-SY5Y cells were plated in 24-well plates one day prior to transfection. Cells were
transfected with the various promoter–EGFP constructs by using the TransFast reagent (Promega) according to the manufacturer’s instructions. HeLa S3 cells were transfected with 1µg and SH-SY5Y cells were transfected with 0.75µg of each of the different promoter–EGFP constructs.

**Live cell imaging**

For live cell imaging, HeLa-S3 and SH-SY5Y cells were plated in 24-well plates one day prior to transfection. Cells were transfected with the different promoter–EGFP constructs as described, and incubated for 4 or 8 days. Cells were imaged every 24 h, starting 24 h after transfection for days 1-4, and 98 h after transfection for days 4-8 (imaging of days 1-4 and 4-8 was performed in separate experiments, in order to prevent cell overcrowding). Images were acquired using the Zeiss Observer Z1 inverted microscope with a 5X ECPlan-Neofluar objective, equipped with a 37°C 5% CO₂ incubation chamber. In order to image a whole well, images were acquired with Zeiss AxioVision software’s “titles” function, allowing stitching images corresponding to a field of a complete well. Resulting images were processed using Adobe Photoshop software.

**Flow cytometry**

FACS analyses for short- and long-term experiments were performed separately (1-4 and 4-8 days post-transfection, respectively, so that day 4 is common to both analyses), to prevent cell overcrowding. HeLa-S3 cells were plated at 2-7 x 10⁴ cells per well and SH-SY5Y cells were plated at 3-9 x 10⁴ cells per well (depending whether the cells were to be analyzed by FACS in days 1-4 or 4-8) in 24-well plates one day prior to transfection. Cells were transfected as described and incubated for 1-8 days. At the indicated time points, cells were harvested, centrifuged at 1000 rpm, resuspended in 0.3 ml PBS and subjected to flow cytometry analysis (Gallios, Beckman Coulter). The fluorescence of EGFP was measured in 10,000 cells per sample. The data was analyzed using the FlowJo software.
Statistical analysis

The fluorescence intensity and number of cells expressing EGFP for each promoter in each experiment, was compared to the pRC/CMV data obtained in the corresponding experiment (by division of the obtained absolute value by that of the pRC/CMV promoter). Statistical comparisons between the promoters were done using the non-parametric Kruskal-Wallis test for independent samples, with pairwise comparisons. The Kruskal-Wallis test was chosen since it is a non-parametric test, (which does not require the assumption that the samples are distributed normally) and since it can be applied to examine groups of unequal size. A p-value ≤ 0.05 was considered to be statistically significant.


Acknowledgments

We thank Anna Sloutskin, Yehuda Danino and Hila Shir for critical reading of the manuscript. We thank Dr. Rachel Levy-Drummer for invaluable assistance with statistical analysis. This research was supported in part by THE LEGACY HERITAGE BIO-MEDICAL PROGRAM OF THE ISRAEL SCIENCE FOUNDATION (grant No. 1447/13) to O.S and T.J-G.
Author contribution statement

D.E., S. B-B, O.S. and T.J-G have conceived the research, designed the study and analyzed data. S.B-B, D.I and R. T. have designed and generated reporter constructs, D.E. and T.J-G wrote the manuscript with input from all authors.

Competing financial interests

The authors declare competing financial interests

Figure caption

Figure 1. Schematic representation of the engineered core promoters.
The pRC/CMV vector (Life Technologies) contains the CMV enhancer and TATA box, but lacks any CMV sequences that are downstream of -16 relative to the +1 transcription start site (including the Inr element). Three variants of pRC/CMV were constructed in which the core promoter region (from -36 to +45) was replaced with either the natural CMV core promoter, which contains the CMV TATA and Inr elements, or with SCP2 or SCP3, which contains the CMV TATA and Inr, the Tollo MTE, and the Calm2 DPE. Single nucleotide changes in SCP3 (relative to SCP2) are marked by a red x. Each of these pRC/CMV-based constructs contains the EGFP reporter gene.

Figure 2. Live cell EGFP imaging of short-term expression of pRc/CMV-based constructs, in HeLa S3 and SH-SY5Y cells.
HeLa S3 and SH-SY5Y cells were transiently transfected with either the pRC/CMV, natural CMV, SCP2 or SCP3 vector expressing EGFP. The cells were imaged once a day during 1-4 days post-transfection (P.T.). Each circle displays the whole well image constructed by stitching individual microscopic fields. (a) HeLa S3 cells. (b) SH-SY5Y cells. Data shown are representative of 3 independent experiments for each cell type.
Figure 3. Live cell EGFP imaging of long-term expression of pRC/CMV-based constructs, in HeLa S3 and SH-SY5Y cells.

HeLa S3 and SH-SY5Y cells were transiently transfected with either the pRC/CMV, natural CMV, SCP2 or SCP3 vector expressing EGFP. The cells were imaged once a day during 4-8 days post-transfection (P.T.). Each circle displays the whole well image constructed by stitching individual microscopic fields. (a) HeLa S3 cells. (b) SH-SY5Y cells. Data shown are representative of 4 independent experiments for each cell type.

Figure 4. FACS analysis of short-term fluorescence intensity and number of fluorescent cells in HeLa S3 and SH-SY5Y cells.

HeLa S3 and SH-SY5Y cells were transiently transfected with pRC/CMV, natural CMV, SCP2 or SCP3 vector expressing EGFP. The cells were collected 1-4 days post-transfection (P.T.) for FACS analysis. (a) FACS analysis of fluorescence intensity of all HeLa S3 fluorescent cells and SH-SY5Y fluorescent cells. (b) FACS analysis of fluorescence intensity of high intensity HeLa S3 fluorescent cells and SH-SY5Y fluorescent cells. (c) FACS analysis of the number of all HeLa S3 fluorescent cells and SH-SY5Y fluorescent cells. (d) FACS analysis of the number of high intensity HeLa S3 fluorescent cells and fluorescent SH-SY5Y cells. Data shown are representative of 5 independent experiments using HeLa S3 cells and 6 independent experiments using SH-SY5Y cells. Statistical comparisons between the promoters were done using the Kruskal–Wallis test with pairwise comparisons. Significant p-values (p ≤0.05) are indicated in the results section.

Figure 5. FACS analysis of long-term fluorescence intensity and number of fluorescent cells in HeLa S3 and SH-SY5Y cells.

HeLa S3 and SH-SY5Y cells were transiently transfected with pRC/CMV, natural CMV, SCP2 or SCP3 vector expressing EGFP. The cells were collected 4-8 days post-transfection (P.T.) for FACS analysis. (a) FACS analysis of fluorescence intensity of all HeLa S3 fluorescent cells and SH-SY5Y fluorescent cells. (b) FACS analysis of fluorescence intensity of high intensity HeLa S3 fluorescent cells and SH-SY5Y fluorescent cells. (c) FACS analysis
of the number of all HeLa S3 fluorescent cells and SH-SY5Y fluorescent cells. (d) FACS analysis of the number of high intensity HeLa S3 fluorescent cells and SH-SY5Y fluorescent cells. Data shown are representative of 6 independent experiments using HeLa S3 cells and 5 independent experiments using SH-SY5Y cells. Statistical comparisons between the promoters were done using the Kruskal–Wallis test with pairwise comparisons. Significant p-values (p ≤0.05) are indicated in the results section are distributed normally) and since it can be applied to examine groups of unequal size. A p-value ≤ 0.05 was considered to be statistically significant.

**Figure 1**

![Diagram of promoter sequences]

96
Figure 4

Figure 5
The title of this work is expressed in its presentation and use of a system of simulation in vivo. This system is based on the system developed by "John Proff" of "Eliyahu Grunew". This system allows monitoring and comparison in real-time of the activity and kinetics of the gene in a single sample and in a specific site in the genome of a single cell.

The effectiveness of this system is based on the fact that its use prevents the variations (position effects) that exist in the systems in other systems that are a result of the location of the gene that goes into integration in the sample. In this way, all the changes observed in the activity and kinetics of the gene, are only caused by the known variable that we want to study.

In this work, I present the establishment of this system, where the known variable is the promoters that are modified. The use of modified core promoters allows us to define the importance and contribution of each core promoter element individually, and also the influences and relationships that exist between the motifs when they are found together in this core promoter. So far, the information obtained from this system provides significant differences between the promoters, which as mentioned can be due to the parameter studied alone, which points to a way of their individual and common activity of the core promoter elements.

In summary, the core promoter elements have a significant role in determining the nature of the promoter and the way it works, so the structure of the promoter is a crucial component in the control of the transfection process. Each core promoter element contributes in a defined manner to the activity and function of the promoter. Since promoters are composed of various motifs, there are motor relationships and influences between the motifs that affect and change the contribution of the motifs to the activity of the promoter. This characteristic of their activity and their individual and common contribution of the core promoter elements will contribute greatly to the existing knowledge about the transfection process, and will deepen our understanding of the nature of the strong promoters and the possible applications of these promoters, such as strong and long-range expression of genes, etc.

transient transfection
תהליכים רבים כגון גדילה, התפתחות, התימיות ושימירה על המאוספטיות, החוסיות לקיום של כל אורגניזם חיוניים לקיומו של כל אורגניזם. חיוניים לתהליך השעתוק ולביצועᾑה לעוניה, התהליך השעתוק המ디יקט על ידי RNA polymerase II (Pol II) hüוה שלב קריטי בביוכי גז מתודית, תלבוש. בשל זה מרוכזים מטמון וצורה core promoter המובילים ליצירת RNA המשגרת של השעתוק.- רצף -ה מוזאך Anat לרצף הדני" אומיום, הדורשה להתחלת השעתוק. ברמה הברדילית, מתארים את התהליך המ디יקט של השעתוק.- רצף -ה ויצור צורה core promoter- ה وفي מחזור נובע קריטי בביטוי גנים מקודדי חלבון.- Pol II. שלב זה מורכב מאירועים רבים המוביל זה ליצירתו של Pol II, על ידי מערכת השעתוק הבזאילית, המובילה החלקול, על ידי פעילותו ותפקודו של RNA core promoter של מוחק של 80bp (40) - (40) +יתו לאות התחלתי השעתוק. רצף -ה מבורכש(core promoter elements / motifs) (כרומט) (DPE - TATA-box), çevir ששיים ו��ויות של פורמותורו.

מחזור קודימי מהכן שלב של מטמון של הרמגורומורו בדיד יין לתוך הובטח core promoters סינטטיים (super core promoters, SCPs), אשר רמות השעתוק המדוברים בין גזים המבקרים על-ידי, גבורה בשמורות. פורמותורים מודificados שאך קארים core promoter מפוחפס על פורמותורים אנדונימיים (modified core promoters) יואליאים, בין ב鍋ינור אקראייה ס" שกระทביל ברבר-ה. ביתสะפוחת סופק לא סינסטר core promoter -רבע-ה, סולקה, וכלל-ה SSCP -_sshאר סופק super core promoter.Bכל מצוין או בטוי הגנה המתחבלת חום הפרוטומורו המודificados הווה. הת῟ואות השקתבל מדרושים את השימוע המפריע ל"downstream core promoter" (רצף-ה) ב השייח השעתוק, יומן ממציא👩‍💻‍💻 משך בטוי אורך ורמות השעתוק גבורה של המתחבלת חום הגנה ש دمشיכו פורמהוטו שבעבר מערכת.

בגזרה לפיעוץ את המגנות המאbestos בטוי לחם מומשך של גני הנتأكد חום בקור ותתאה הפרוטומורו המודificados. ל"ח, בדקנו את מדברת המאפרור של פורמותורו אנדונימיות של בקטריה שעתוק אולימ. הגנו על תשעיל תכלית הבחריל השעתוק, ונראה כי לזרבע הפרוטומורו בועקר רצף ברוםزهرת תכלית השעתוקי (downstream sequence), יש תמייק ששבה בחריות האופי של הפרוטומורו רבודה היה תלוי או אלה של בקטריה שעתוק מוסים, ועם בטובות דרכ העדשה של הפרוטומורו מתנגת ליגריוו שניה.
עבודה זו נועשתה בהדרכת של ד"ר תמר יוב-גרושו,
מן הפקולטה למדעי החיה על שם מינה ואברדר ופודמן של אוניברסיטת בר אילן,

ומתמקדת באיתיותה של נ見える היצורים על שם פודמן ופודמן של אוניברסיטת בר אילן.
CHARACTERIZATION OF GENE EXPRESSION AND TRANSCRIPTION KINETICS UNDER THE REGULATION OF MODIFIED CORE PROMOTERS

 espa
とっても美しい花びらが大きく広がっています。