Biochemistry and Biophysics Reports 39 (2024) 101795

Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

journal homepage: www.elsevier.com/locate/bbrep

Engineered mouse H1 promoter mutants with superior RNA polymerase III activity

Jiaying Wu [a], Yufei Zhou [b], Di Zhao [b], Ran Xu [b,c], Jienan Wang [b], Hong Lin [b], Zhiwen Ding [b,d,*], Yunzeng Zou [a,b,d,**]

[a] Institutes of Biomedical Sciences, Fudan University, Shanghai, 200032, China
[b] Shanghai Institute of Cardiovascular Diseases, Zhongshan Hospital, Shanghai, 200032, China
[c] Shanghai Geriatric Medical Center, Shanghai, 201104, China
[d] Departments of Cardiology, Qingpu Branch of Zhongshan Hospital, Fudan University, Shanghai, 218120, China

ARTICLE INFO

Keywords:
Mouse H1 promoter
Pol III activity
Transcription

ABSTRACT

Vectors incorporating the human H1 (hH1) promoter are being applied for RNA interference (RNAi) experiments and genome editing. Although extensive studies have been conducted on the hH1 promoter, our understanding of the mouse H1 promoter remains limited. In this study, we predicted the 163 bp mouse H1 (mH1) promoter and 84 bp mouse H1 core (mH1 core) promoter through global alignment and detected its RNA polymerase II (Pol II) and III activities through the expression of the EGFP and the abundance of artificial sequence, which were generally slightly weaker than those of the hH1 promoter. Furthermore, to boost its Pol III activity, we engineered various promoter mutants by introducing mutations or systematically swapping elements. Surprisingly, the Pol II activity of mH1 core mut5 with AT stretch was at least 2-fold greater than that of the wild type, making it a potential candidate for target protein expression purposes. Fortunately, the Pol III activities of mH1 mut1 and mH1 core mut5 were at least 1.5 times stronger than those of the parental promoters in human and mouse cell lines on account of AT stretch, as did the mH1 mut4 with AT stretch and proximal sequence element (PSE) and TATA box insertion mutations. We highly recommend these three promoters as valuable supplements to the type 3 Pol III promoter toolbox.

1. Introduction

RNA interference and CRISPR-mediated gene editing techniques have been extensively utilized for gene functional analysis as well as therapeutic interventions. The abundance of expressed RNA significantly affects gene knockout or editing efficiency. Therefore, it is crucial to carefully select the appropriate promoter when constructing vectors. Type 3 RNA polymerase III (Pol III) promoters, such as U6, 7SK, and H1, are extensively used for the expression of small noncoding RNAs, including small interfering RNA (siRNA) or short hairpin RNA (shRNA) for RNA interference (RNAi) purposes and single guide RNA (sgRNA) or engineered prime editing guide RNA (epegRNA) for CRISPR-mediated genome editing platforms [1–3]. These Pol III promoters can also load Pol II and transcribe lengthy translation-competent mRNAs that differ in Pol II strength (H1 ≫ U6 > 7SK) [4]. Researchers generally select the U6 promoter in adenovirus and lentivirus interference vectors because its Pol III activity is the strongest (U6 > H1 > 7SK) [4]. However, it is sometimes better to choose the H1 promoter because of the limited packaging capacity of viral vectors or to avoid long stretches of homology that can lead to recombination [5].

A single H1 promoter can drive both guide RNA and endonuclease expressions in the CRISPR-Cas9 system [6]. Moreover, the human H1 (hH1) promoter and H1 core promoter are within a more compact region than others. Consequently, the H1 promoter holds significant potential for viral vector applications. The full-length H1 promoter can be divided into two somewhat modular components, a core promoter region and upstream activating sequences. Different regions have different regulatory effects on transcriptional initiation. The H1 core promoter is composed of a distal sequence element that enhances transcription and a basal region directing basal transcription. The distal sequence element contains a staf binding site and an octamer motif. The basal region consists of a proximal sequence element (PSE) and a TATA box, which

* Corresponding author. Shanghai Institute of Cardiovascular Diseases, Zhongshan Hospital, Fudan University, Shanghai, 200032, China.
** Corresponding author. Institutes of Biomedical Sciences, Fudan University, Shanghai, 200032, China.
E-mail addresses: zhiwen_d@fudan.edu.cn (Z. Ding), zou.yunzeng@zs-hospital.sh.cn (Y. Zou).

https://doi.org/10.1016/j.bbrep.2024.101795
Received 5 May 2024; Received in revised form 11 July 2024; Accepted 18 July 2024

are separated by a spacer [7].

While extensive studies have been conducted on the human H1 promoter, comprehensive investigations of the mouse H1 promoter remain limited. Here, we identified and cloned the 163 bp mouse H1 (mH1) promoter and the 84 bp mouse H1 core (mH1 core) promoter. In parallel, we analyzed their Pol III and Pol II activities and designed engineered promoter mutants with enhanced Pol III and/or Pol II activity by introducing mutations or swapping elements.

2. Materials and methods

2.1. Vectors construction

The 163 bp mH1 promoter, 84 bp mH1 core promoter, and 363 bp Fs + mH1 promoter sequences were amplified with PCR from mouse genomic DNA. The vectors pEGFP-C1 and pLV-H1-shRNA were used as sources of EGFP and human H1 promoter and human H1 core promoter sequences. We ordered a gene fragment encoding artificial sequence at OBiO Tech. Promoters and EGFP sequence or artificial sequence were spliced with overlap extension PCR. The promoter-EGFP fragments or promoter-artificial sequence fragments were ligated into the pLV-puro-basic vector using the *Cla* I and *Sma* I restriction enzyme sites by Gibson cloning, which constituted the backbone construct pLV-puro-promoter-EGFP and pLV-puro-promoter-artificial sequence. We obtained promoter mutant fragments by overlap extension PCR and connected these fragments to pLV-puro-promoter-EGFP and pLV-puro-promoter-artificial sequence vectors through Gibson assembly. The sequences of all plasmid constructs were confirmed by DNA sequencing at AZENTA.

2.2. Cell cultures

HEK293T, AC16 and MC38 cells were grown as monolayers in DMEM (Gibco) medium supplemented with 10 % fetal bovine serum, penicillin (100 U/ml) and streptomycin (100 U/ml) at 37 °C and 5 % carbon dioxide. AC16 is a human cardiomyocyte cell line, and the MC38 cell line is derived from C57BL6 murine colon adenocarcinoma cells.

2.3. Lentivirus production and transduction

HEK293T cells were trypsinized and seeded in a 6-well plate 1 day before transfection to achieve a confluency of approximately 80 % on the day of transduction. The culture medium was replaced 1 h before transfection. Then, 2 μg of pLV-puro-promoter-EGFP/artificial sequence lentiviral constructs were transfected together with 2 μg of the packaging plasmids pMDL, pREV, and pVSVG via calcium phosphate. The medium was replaced 13 h after transfection with fresh complete medium. After an additional 48 h, the supernatant containing the virus was harvested and filtered through a 0.45 μm pore size membrane to remove cell debris, and the titer of the lentivector stock was assessed. Moreover, cells transfected with pLV-puro-promoter-EGFP were collected for western blotting and flow cytometry at 48 h post-transfection. Cells transfected with the pLV-puro-promoter-artificial sequence were collected for quantitative PCR.

2.4. Infection

To establish stably transfected cells, one day before infection, AC16 or MC38 cells were seeded in a 12-well plate. The cells were incubated at 37 °C overnight, and the cell density was approximately 40 % on day 2. The culture medium was aspirated from the cell culture, the mixture (virus + medium + polybrene) was added to the cells, and the mixture was shaken gently. The plate was centrifuged at 1500×*g* for 30 min and incubated at 37 °C. The virus-containing medium from the cell culture medium was replaced, and fresh medium was added to the cells at 24 h post-infection. Puromycin (AC16: 1 μg/ml, MC38: 3 μg/ml) can be added at 48 h post-infection for screening. At 72 h post-infection, the cells were collected as needed to detect the expression of artificial sequence or EGFP.

2.5. RNA extraction and quantitative PCR

To explore the artificial sequence abundance at the transcriptional level, quantitative polymerase chain reaction (qPCR) was performed. Total RNA was extracted from cultured cells using TRIZOL reagent (Vazyme, Cat# R401-01), reverse-transcribed into complementary DNA and amplified by using the HiScript III RT SuperMix for qPCR (+gDNA wiper) Kit (Vazyme, Cat# R323-01). qPCR was conducted in triplicate using ChamQ Universal SYBR qPCR Master Mix (Vazyme, Cat# Q711-02) with the appropriate forward and reverse primers. Assays were performed according to the manufacturer's suggestions. The PCR primers used were synthesized by AZENTA (Suzhou, China) and are shown in Supplementary Table S1. The results of the qPCR were analyzed by the $2^{-\Delta\Delta CT}$ method. The values were normalized to the β-actin levels.

2.6. Western blot analysis

Total proteins were extracted from the indicated cells with RIPA lysis buffer. The concentration of each sample was measured using the BCA Protein Assay Kit (Beyotime, Cat# P0012S). Proteins were subjected to sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) with a Tris-glycine system at 80 V for 30 min and 110 V for 60 min and transferred to polyvinylidene fluoride membranes (Millipore, USA) at 200 mA for 90 min. The membranes were then blocked at room temperature for 20 min by using Fast Blocking Western Mix (YEASEN, Cat# 36122ES60). After incubation with horseradish peroxidase (HRP)-conjugated primary antibodies at 4 °C for one night, the membranes were washed three times. An electrochemiluminescence detection kit from Share-bio Biotechnology (Cat# SB-WB001) was used for blot chemiluminescence. Image Lab 3.0 software (Bio-Rad, GA, USA) was used to obtain and analyze the signals. The target proteins were normalized against the loading control β-actin. The primary antibodies used for the western blots were as follows: GFP (1: 4000 dilution) was acquired from Thermo (Cat# 600-103-215), and β-actin (1: 5000 dilution) was purchased from Proteintech (Cat# HRP-66009).

2.7. Flow cytometry

The Pol II activities of the promoters were monitored by flow cytometry. Harvested cells that expressed EGFP were washed with fluorescence-activated cell sorting buffer (PBS with 2 % BSA), and data acquisition was performed by a MoFlo XDP High Speed Cell Sorter and Analyzer (BECKMAN). The mean fluorescence intensity of EGFP was analyzed by FlowJo-v10 software.

2.8. Statistical analysis

For all the quantitative PCR samples, analysis was performed using two-way ANOVA with replication, and a p value < 0.05 was considered to indicate statistical significance.

3. Results

3.1. Characterization of the mouse H1 promoter

The mouse H1 promoter was identified by comparing the mouse H1 genomic DNA sequence with that of the human H1 promoter. Global alignment analysis revealed that the mouse H1 promoter contained a 163 bp region with 52.2 % identical to the human H1 promoter sequence and elements similar to those in the hH1 promoter (Fig. 1a). The sequence of their core promoter, which included an octamer, a staf

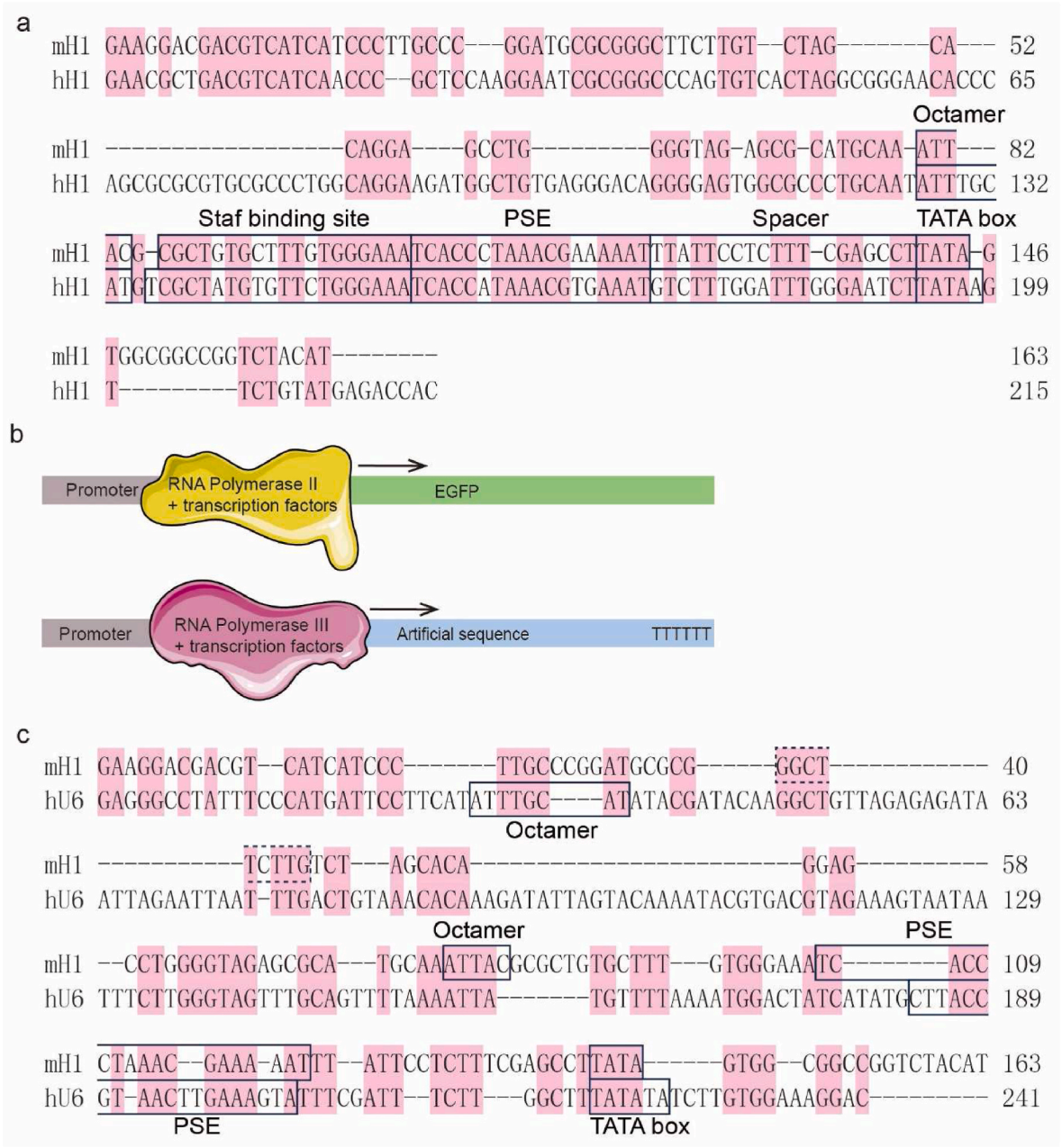

Fig. 1. Global alignment of the mouse H1 promoter with others. **(a)** The Needleman–Wunsch algorithm was used to calculate a global alignment of the mouse H1 (mH1) promoter and human H1 promoter (hH1) sequences. Promoter elements, including the octamer, staf binding site, PSE, spacer, and TATA box, are indicated by black rectangles. **(b)** Schematic of the promoter constructs used for measuring Pol II or III activity. Pol II activity of the respective promoters was quantified based on the measured EGFP expression. Pol III activity is reflected by the abundance of artificial sequence, a noncoding RNA terminated at the T6 signal. **(c)** The Needleman–Wunsch algorithm was used to calculate a global alignment of the mH1 promoter and human U6 promoter (hU6) sequences. The similarity between the two sequences is 39.2 %. The solid black rectangles outline the PSE, octamer, and TATA box. The dotted rectangle outlines the AT stretch site. PSE, proximal sequence element. Pol II/III, RNA polymerase II/III.

binding site, a PSE, a spacer, a TATA box, and an initiator, was the minimal DNA sequence required for transcription initiation. We successfully cloned the 163 bp mH1 promoter, the 84 bp mH1 core promoter and the 363 bp Fs + mH1 promoter obtained by adding 100 bp sequences on both sides of the mH1 promoter. The hH1 promoter and the hH1 core promoter were utilized as controls.

3.2. *Assessing the Pol III activity of the mH1 promoter in multiple cell types*

To analyze the Pol III activity of the mH1 promoter, we constructed a plasmid in which the respective promoters transcribed an artificial sequence followed by the efficient Pol III termination signal TTTTTT (Fig. 1b). Transient transfection of equal molar amounts of vectors containing different promoters into HEK293T cells resulted in a cell viability of ~95 % at 48 h post-transfection (as observed by fluorescence microscopy using a plasmid with EGFP harboring a CMV promoter). Total RNA was extracted, reverse-transcribed into cDNA, and the abundance of artificial sequence was analyzed by qPCR. Preliminary experiments confirmed that mH1 promoter, mH1 core promoter, and Fs + mH1 promoter almost were in similar Pol III strength but weaker than hH1 and hH1 core promoter (Fig. 2a). In consideration of the

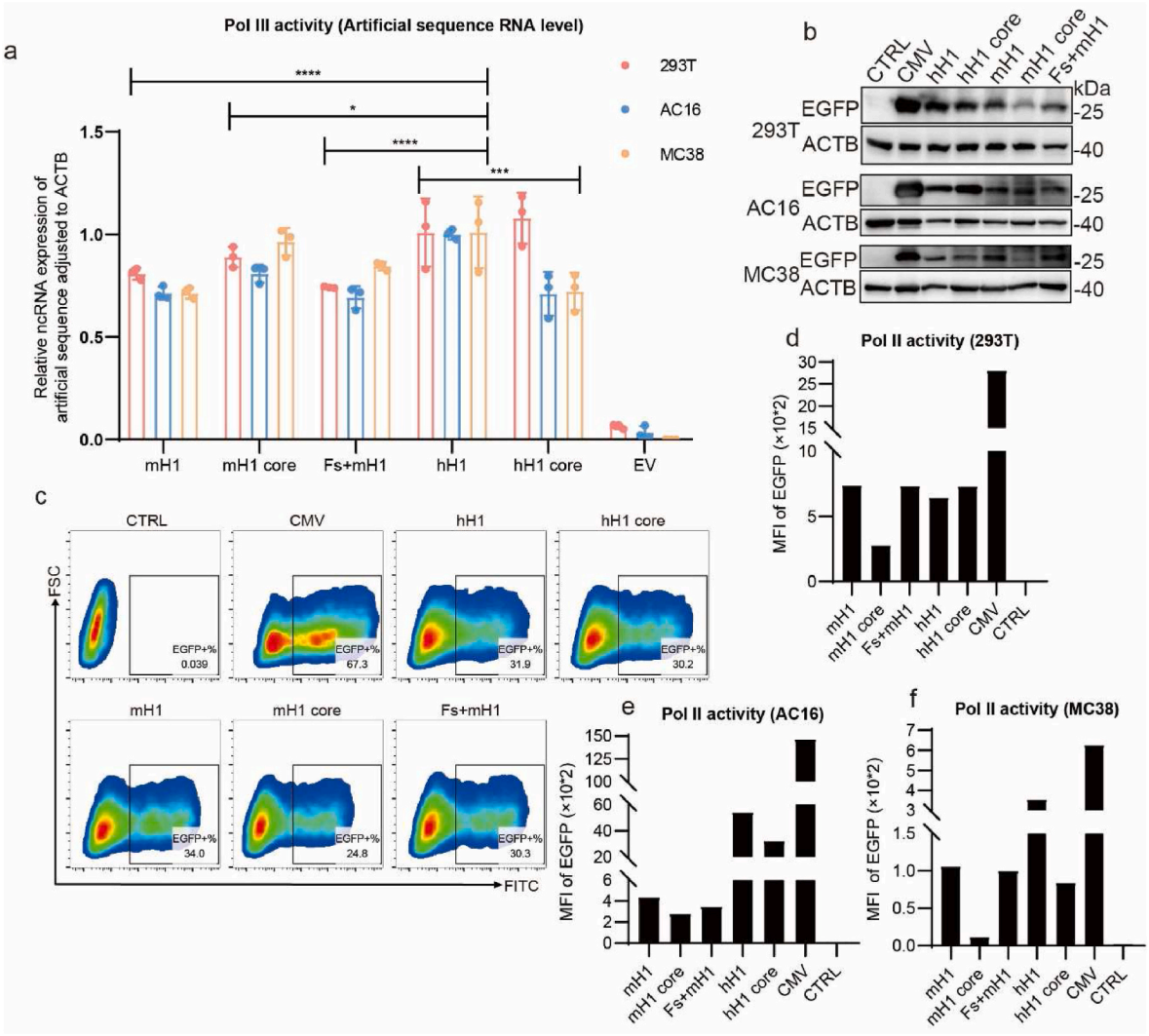

Fig. 2. Pol III/II activity of the mH1 promoter versus that of the hH1 promoter in multiple cell types. **(a)** Quantitative PCR for the artificial sequence transcript was conducted for transfected HEK293T, infected AC16, and MC38 cells. The promoter-less EV construct was used as a negative control, while the positive control was the hH1 promoter construct. **(b)** Western blot detection of EGFP expression. ACTB served as the internal control. **(c)** HEK293T cells were transfected with the EGFP construct driven by the promoters. After 2 days, the cells were analyzed by flow cytometry. **(d–f)** The mean fluorescence intensity of EGFP in three cell types was determined by flow cytometry. 293T cells were collected 48 h after transient transfection; AC16 and MC38 cells were collected 72 h after lentivirus infection. The data are presented as the mean ± SD (n = 3). *$p < 0.05$; ***$p < 0.001$; ****$p < 0.0001$. Pol II/III, RNA polymerase II/III; MFI, mean fluorescence intensity; AC16, human cardiomyocyte cell line; MC38, murine colon adenocarcinoma cell line.

transfection mode and speciation specificity, AC16 cells (a human cardiomyocyte cell line) and MC38 cells (a murine colon adenocarcinoma cell line) were infected with lentivirus. Quantification of the RNA indicated similar Pol III activity for the four promoters, except for the hH1 promoter, which was slightly stronger than the others (Fig. 2a).

3.3. *Assessing the Pol II activity of the mH1 promoter in multiple cell types*

It has been demonstrated that the hH1 promoter, a type 3 RNA Pol III promoter, can recruit RNA polymerase II for transcribing extended messenger transcripts (*4*). We hypothesized that the mH1 promoter would have Pol III and Pol II activity. In parallel, to measure the Pol II activity profile of the mH1 promoter, HEK293T cells transfected with the pLV-puro-promoter-EGFP plasmid were collected for western blotting and flow cytometry at 48 h post-transfection. The MFI of CMV promoter-activated EGFP was much greater than that of H1 promoters (both human and mouse) (Fig. 2d–f). Among all the promoters, the MFI of EGFP driven by the mH1 core promoter was found to be the lowest (Fig. 2d–f). Comparing the MFI of EGFP revealed that Pol II activity was similar for the mH1 promoter, Fs + mH1 promoter, hH1 promoter, and hH1 core promoter (Fig. 2d–f). These findings were further confirmed through Western blot analysis showing the protein expression of EGFP (Fig. 2b). Interestingly, upon infecting AC16 cells with lentivirus, we observed higher MFIs for the hH1 promoter and hH1 core promoter compared to all three mouse H1 promoters (Fig. 2d–f). Additionally, after infecting MC38 cells with lentivirus, there was a reduction in the MFI ratio between the mouse H1 promoter and the human or CMV promoter (Fig. 2f). Based on these observations, we hypothesized that Pol II activity of the H1 promoter might be species specific. Furthermore, our survey indicated that core promoters consistently exhibited weaker Pol II activity than full-length promoters in both mouse and human systems, possibly due to regulatory elements present within additional sequences involved in transcriptional regulation processes. Notably, the similarity in Pol II and III activities between the mH1 promoter and Fs + mH1 promoter suggested the reliability of global alignment results and that the mH1 promoter sequence was capable of performing complete promoter functions (Fig. 2).

3.4. Engineering of the mH1 promoter with improved Pol III activity

The H1 promoter is particularly attractive for capacity-limited viral vector systems because it has a short sequence and strong Pol III activity. Thus, the H1 promoter is frequently applied in RNAi experiments and gene editing. However, gene editing efficiency is closely related to the expression of the sgRNA or epegRNA. The higher the expression level is, the better the efficiency. Therefore, we hope to strengthen the Pol III activity of the mH1 promoter through promoter engineering.

The sequences of the human U6 (hU6) promoter and the mH1 promoter were aligned (Fig. 1c), as we sought to harness the superior Pol III activity exhibited by the hU6 promoter compared to that of the mH1 promoter. Furthermore, we meticulously classified crucial elements of the hU6 promoter based on extensive prior research (Fig. 1c) [8,9]. We

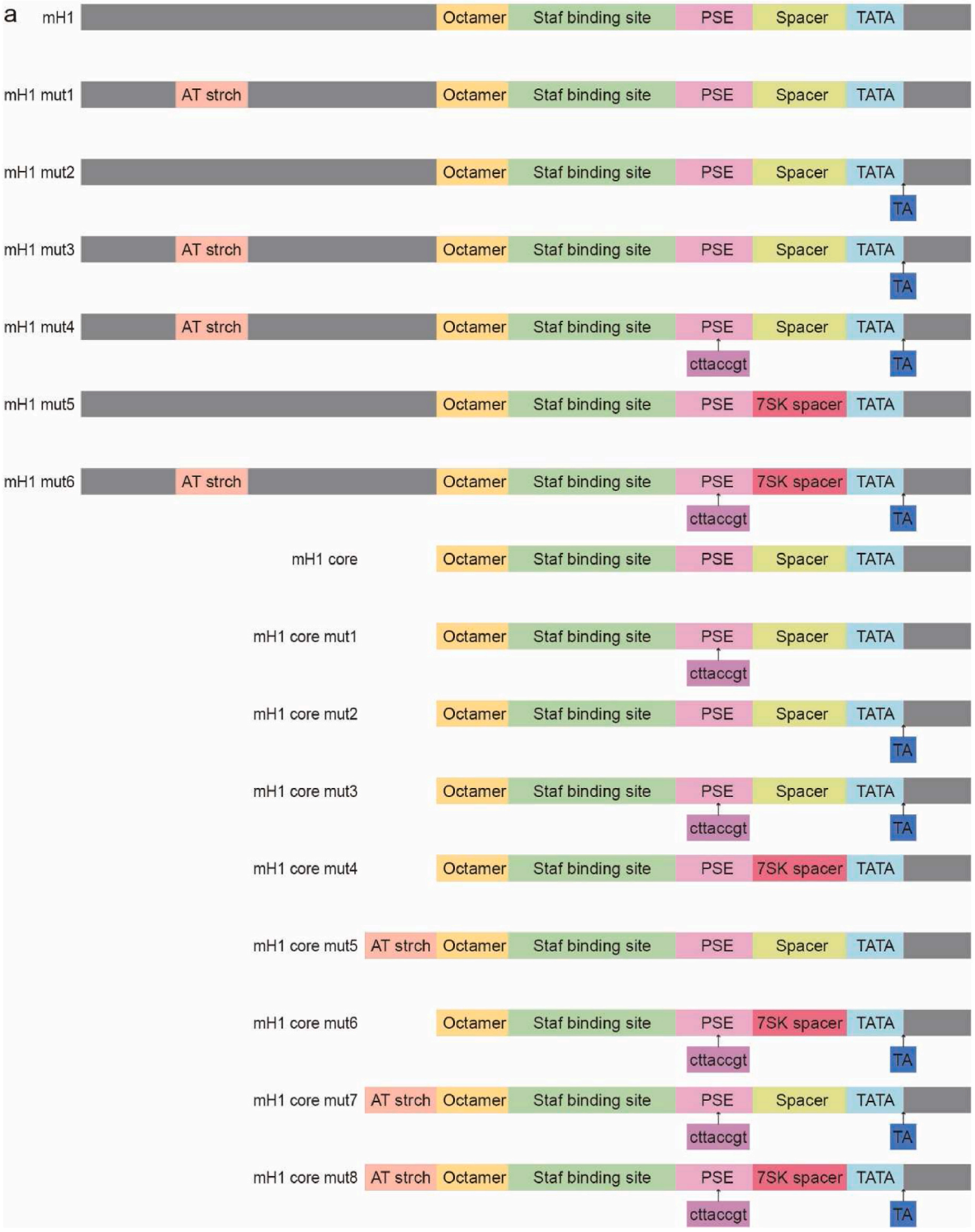

Fig. 3. Introducing mutations or other promoter elements in the mH1 and mH1 core promoters were introduced. **(a)** We designed six mutants for the mH1 promoter and eight mutants for the mH1 core promoter. The promoter elements included the octamer, staf binding site, PSE, spacer, and TATA box. The "AT strch" stands for the AT stretch. The "cttaccgt" and "TA" were inserted into the PSE and TATA box, respectively.

found that the mismatched sequences in the upstream of the mH1 promoter and the hU6 promoter were rich in AT bases, so we replaced the mH1 promoter sequence "GGCTTCTTG" shown by the dotted line with "ATATTTGCA", called the AT stretch, resulting in the birth of the mH1 mut1 promoter (Fig. 3a). Additionally, we inserted the sequence "ATATTTGCATATAA" into the front part of the mH1 core promoter, also called AT stretch, thereby obtaining the mH1 core promoter mutant 5 (Fig. 3a). Fortunately, the two mutants significantly increased Pol III activity by at least 1.5 times (Figs. 4a and 5a). Interestingly, the two mutants differed in Pol II activity. mH1 mut1 tended to be weaker than the wild type (Fig. 4b–f), while mH1 core mut5 showed an ascending tendency (Fig. 5b–f). Comparing mH1 core mut3 and mH1 core mut7, the latter had only one more AT stretch than the former (Fig. 3a), and its Pol II and III activities were significantly enhanced (Fig. 5). In addition, although there was no difference in Pol III activity between mH1 core mut6 and mH1 core mut8, the Pol II activity of the latter also exhibited a trend of enhancement (Fig. 5). Overall, AT stretch enhanced Pol II and III activities. We speculated that changes in nucleotide sequence, possibly introducing an enhancer element, or a high AT content might reduce DNA methylation or loosen chromatin structure and lower nucleosome occupancy, thereby promoting the discovery and binding of transcription factors and RNA polymerases to promoters.

Furthermore, the TATA box is recognized by the TATA-binding protein (TBP), which is a central part of the preinitiation complex [10]. Only a minority of mammalian core promoters have a clear TATA box [11]. This is probably because general transcription factors are complexes of multiple interchangeable proteins with various sequence preferences [12], allowing flexible and degenerate core promoter

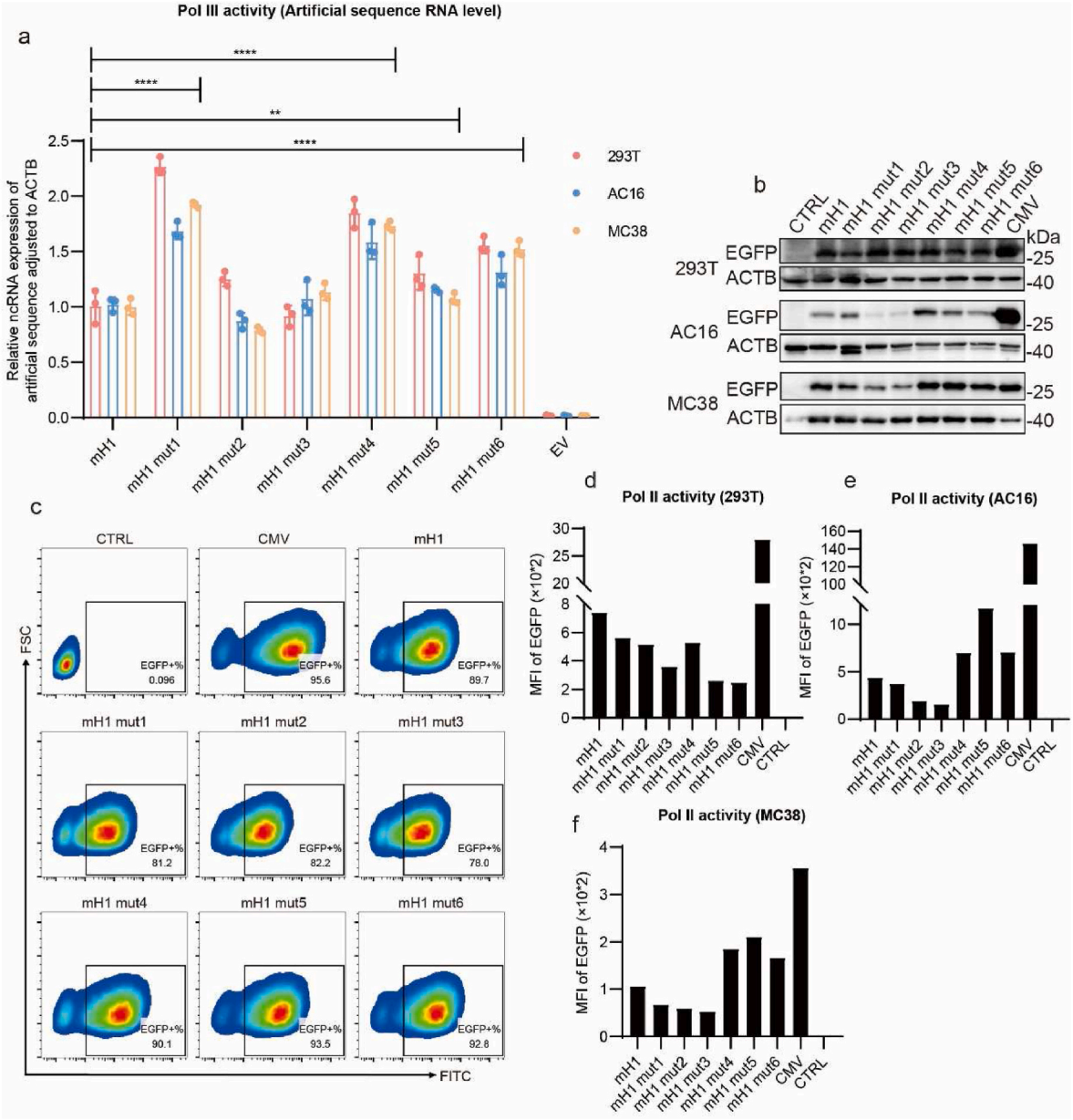

Fig. 4. Pol III/II activity of mH1 promoter mutants in multiple cell types. **(a)** Total cellular RNA from transfected HEK293T cells, infected AC16 cells, and MC38 cells were subjected to qPCR to quantify the artificial sequence RNA level. **(b)** Western blot detection of EGFP expression driven by the mH1 promoter mutants in three cell types. **(c)** Flow cytometry detection of EGFP expression in MC38 cells at 72 h post-infection. Cells infected with lentiviral empty vectors served as a negative control, while the CMV promoter construct was utilized as a positive control. **(d–f)** The mean fluorescence intensity of EGFP in three cell types was determined by flow cytometry. The data represent the mean (±S.D.) of three biological replicates. $^{**}p < 0.01$; $^{****}p < 0.0001$. Pol II/III, RNA polymerase II/III; MFI, mean fluorescence intensity; AC16, human cardiomyocyte cell line; MC38, murine colon adenocarcinoma cell line.

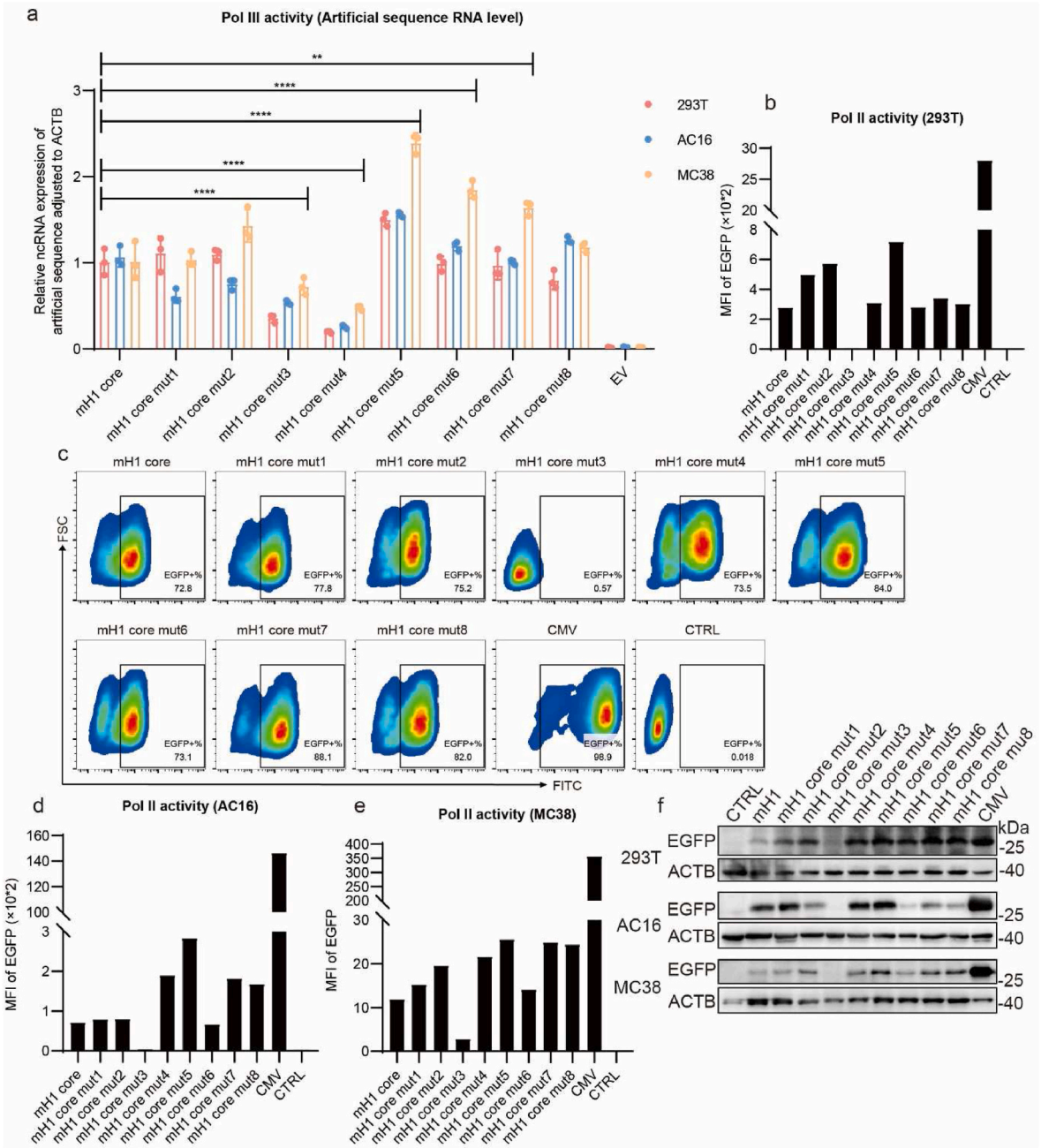

Fig. 5. Pol III/II activity of mH1 core promoter mutants in multiple cell types. **(a)** Artificial sequence transcript detection was performed by qPCR experiment. Total cellular RNA was harvested after 2 days post-transfection or 3 days post-infection. The empty vector that lacks a promoter was used as a negative control. **(b)** Equimolar amounts of DNA constructs were transfected into HEK293T cells, and the mean fluorescence intensity of EGFP reflecting Pol II activity was measured by flow cytometry. **(c–d)** AC16 cells infected with lentivirus were analyzed by flow cytometry, and the mean fluorescence intensity was determined. **(e)** Flow cytometry for the mean fluorescence intensity of EGFP was conducted for infected MC38 cells. **(f)** Western blot detection of EGFP expression in three cell types. The data are presented as the mean ± SD (n = 3). $^{**}p < 0.01$; $^{****}p < 0.0001$. Pol II/III, RNA polymerase II/III; MFI, mean fluorescence intensity; AC16, human cardiomyocyte cell line; MC38, murine colon adenocarcinoma cell line.

recognition. Thus, we endeavored to insert "TA" into the TATA box of the mH1 promoter or mH1 core promoter to improve its binding affinity for TBP, thereby influencing Pol III or II activity (Fig. 3a). The designed mutants, mH1 mut2 (Fig. 4) and mH1 core mut2 (Fig. 5), exhibited negligible disparity in Pol III activity compared to their wild type counterparts, yet the mutation had increased the Pol II activity of the mH1 core mut2 in the three cell lines. Simultaneously, we combined the AT stretch with the mutated sequence of the TATA box to generate the mH1 mut3 promoter (Fig. 3a). While there was no alteration observed in Pol III activity for mH1 mut3, but its Pol II activity was weakened, which was consistent with that of mH1 mut2 with solely underwent a TATA box mutation (Fig. 4). This discrepancy highlighted the distinct effects of TATA box mutations on both full-length promoters and core promoters, indicating that there was a certain interplay between the upstream sequences and the TATA box. Furthermore, it implied that the impact of the TATA box on Pol II activity was substantial and emphasized the complexity of transcriptional regulation.

In a previous study, replacing miniature human H1 PSE with that of 7SK led to the mutant exhibiting 4-fold increased Pol III specificity over the parental H1 promoter [13]. We also discovered that the PSE of the

hU6 promoter differed from that of the mH1 promoter upon sequence alignment, prompting us to incorporate a portion of the hU6 PSE sequence into the mH1 promoter PSE. Consequently, this led to the creation of mH1 core mut1 (Fig. 3a). This alteration resulted in an augmentation of Pol II activity while leaving Pol III activity unaffected (Fig. 5). Chromatin immunoprecipitation sequencing experiments revealed that the TATA-free PSE-containing promoter predominantly recruited RNA polymerase II [14]. This might be attributed to the fact that the longer PSE sequence bound more readily and/or tightly with RNA polymerase II. On another note, mH1 core mut3 combined both mutations in its PSE and TATA box by incorporating elements from both mH1 core mut1 and mH1 core mut2. Strangely enough, there was a significant reduction in activities for both Pol II and Pol III in mH1 core mut3 (Fig. 5). It was a pity that we lacked a satisfactory explanation for this phenomenon, other than to conjecture that changes in sequence might alter chromatin structure so that hindering efficient access of the transcription initiation complex to the promoter.

Previous studies have demonstrated that miniature hH1 core mutants harboring the 7SK spacer exhibit remarkably diminished Pol II activity [13]. The 7SK spacer displays modest Pol II activity and represents an excellent choice in terms of Pol III specificity [4]. Consequently, we opted to substitute the mH1 spacer with the 7SK spacer, resulting in the generation of the mutants mH1 mut5 and mH1 core mut4. While the Pol III activity of mH1 mut5 experienced a slight increase (Fig. 4a), that of mH1 core mut4 underwent a decrease (Fig. 5a). In both cases, Pol II activity was either reduced or remained unchanged in 293T cells (Figs. 4d and 5b) but was enhanced in AC16 (Figs. 4e and 5c-d) and MC38 cells (Figs. 4f and 5e). This observation suggested that the impact of the 7SK spacer on Pol III activity within the context of the mH1 promoter was regulated by upstream sequences and might be influenced by the transfection methods employed. It was well established that the H1 promoter exhibited exceptional characteristics, as it recruited both Pol II and III complexes. We hypothesized that distinct binding patterns or competition mechanisms between Pol II and Pol III on plasmid DNA versus chromatin DNA contributed to this outcome.

4. Discussion

The mH1 promoter and mH1 core promoter were predicted in this study through optimal global alignment, with their Pol II and III activities assessed by the expression of the EGFP and the abundance of artificial sequence, which were generally slightly weaker than those of the hH1 promoter. Furthermore, to boost its Pol III activity, we engineered various hybrid promoter mutants by introducing mutations or systematic swapping elements. Fortunately, the Pol III activities of mH1 mut1 and mH1 core mut5 exhibit at least 1.5 times stronger than that of the parental promoter due to the AT stretch, as does the mH1 mut4 with multiple mutations. These three promoters are valuable supplements to the Pol III promoter toolbox that we highly recommend.

Interestingly, we observed that full-length promoter mutants and core promoter mutants sometimes displayed different effects on Pol III activity despite having identical mutations. For example, replacing mH1 mut5 with a 7SK spacer increased Pol III activity while substituting mH1 core mut4 decreased it (Figs. 4a and 5a). We attempted to combine multiple mutations to obtain mutants with enhanced Pol III or II activity, such as mH1 mut6, mH1 core mut7 and mH1 core mut8, but did not achieve satisfactory results. In addition, the Pol II activity of mH1 core mut5 showed an enhancement of at least 2-fold compared to that of the wild type across the three cell lines, making it a potential candidate for target protein expression purposes (Fig. 5b–f). Taken all findings together, these all highlighted the complexity involved in transcription initiation.

We adopted two strategies for promoter engineering, insertional mutagenesis and hybrid construction. Although the desired promoters with enhanced Pol II or III activity were obtained, the whole process was inefficient, time-consuming and laborious. Another strategy for developing novel promoters is nucleotide diversification by using error-prone PCR [15]. Promoter engineering by nucleotide diversification requires creating tens or even millions of variations and then analyzing the effects of the different sequences on promoter function [10]. There is no doubt that this is an enormous amount of work. Exhilaratingly, the prediction of promoter transcription strength has proven amenable to machine learning techniques [16–18] that can also enable the de novo design of promoters and the identification or design of their parts [19]. The behavior of promoters is too complex to follow a clear set of rules. Thus, the combination of machine learning and hybrid promoter engineering will be an efficient way to perform research and more likely to succeed in finding potential promoters for applications in areas such as synthetic biology and cell therapeutics.

Funding

This research was funded by the National Natural Science Foundation of China (No. 81900245 to ZW.D.). and No. 81730009, and 81941002 to YZ.Z.).

Consent for publication

Not applicable.

CRediT authorship contribution statement

Jiaying Wu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Yufei Zhou:** Writing – review & editing, Visualization, Investigation, Formal analysis, Data curation. **Di Zhao:** Writing – review & editing, Visualization, Investigation, Formal analysis, Data curation. **Ran Xu:** Data curation. **Jienan Wang:** Data curation. **Hong Lin:** Data curation. **Zhiwen Ding:** Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Yunzeng Zou:** Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors thank Mr. Guoping Zhang for his assistance in the flow cytometry experiments and daily ordering of supplies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrep.2024.101795.

References

[1] J.W. Nelson, et al., Engineered pegRNAs improve prime editing efficiency, Nat. Biotechnol. 40 (2022) 402. +.
[2] L. Cong, et al., Multiplex genome engineering using CRISPR/cas systems, Science 339 (2013) 819–823.
[3] O. ter Brake, et al., Lentiviral vector design for multiple shRNA expression and durable HIV-1 inhibition, Mol. Ther. 16 (2008) 557–564.

[4] Z.L. Gao, E. Herrera-Carrillo, B. Berkhout, RNA polymerase II activity of type 3 Pol III promoters, Mol. Ther. Nucleic Acids 12 (2018) 135–145.
[5] J.R. Davis, et al., Efficient prime editing in mouse brain, liver and heart with dual AAVs, Nat. Biotechnol. 42 (2024) 253–264.
[6] Z.L. Gao, E. Herrera-Carrillo, B. Berkhout, A single H1 promoter can drive both guide RNA and endonuclease expression in the CRISPR-cas9 system, Mol. Ther. Nucleic Acids 14 (2019) 32–40.
[7] L. Schramm, N. Hernandez, Recruitment of RNA polymerase III to its target promoters, Gene Dev. 16 (2002) 2593–2620.
[8] R. Preece, et al., 'Mini' U6 Pol III promoter exhibits nucleosome redundancy and supports multiplexed coupling of CRISPR/Cas9 effects, Gene Ther. 27 (2020) 451–458.
[9] Q. Li, et al., Identification and validation of a Schistosoma japonicum U6 promoter, Parasites Vectors 10 (2017).
[10] A.P. Cazier, J. Blazeck, Advances in promoter engineering: novel applications and predefined transcriptional control, Biotechnol. J. 16 (2021).
[11] P. Carninci, et al., Genome-wide analysis of mammalian promoter architecture and evolution, Nat. Genet. 38 (2006) 626–635.
[12] G. Orphanides, T. Lagrange, D. Reinberg, The general transcription factors of RNA polymerase II, Gene Dev. 10 (1996) 2657–2683.
[13] Z. Gao, et al., Engineered miniature H1 promoters with dedicated RNA polymerase II or III activity, J. Biol. Chem. 296 (2021).
[14] M. Snyder, et al., Genomic study of RNA polymerase II and III SNAPc-bound promoters reveals a gene transcribed by both enzymes and a broad use of common activators, PLoS Genet. 8 (2012).
[15] M. Levo, E. Segal, In pursuit of design principles of regulatory sequences, Nat. Rev. Genet. 15 (2014) 453–468.
[16] A. Hossain, et al., Automated design of thousands of nonrepetitive parts for engineering stable genetic systems, Nat. Biotechnol. 38 (2020).
[17] R.P.H. de Jongh, A.D.J. van Dijk, M.K. Julsing, P.J. Schaap, D. de Ridder, Designing eukaryotic gene expression regulation using machine learning, Trends Biotechnol. 38 (2020) 191–201.
[18] C.G. de Boer, et al., Deciphering eukaryotic gene-regulatory logic with 100 million random promoters (vol 38, pg 56, 2020), Nat. Biotechnol. 38 (2020), 1211-1211.
[19] Y. Wang, et al., Synthetic promoter design in Escherichia coli based on a deep generative network, Nucleic Acids Res. 48 (2020) 6403–6412.

Abbreviations

The following abbreviations are used in this manuscript

RNAi: RNA interference
siRNA: small interfering RNA
shRNA: short hairpin RNA
sgRNA: single guide RNA
epegRNA: engineered prime editing guide RNA
hH1 promoter: human H1 promoter
hH1 core promoter: human H1 core promoter
mH1 promoter: mouse H1 promoter
mH1 core promoter: mouse H1 core promoter
Fs + mH1: adding 100 bp sequences on both sides of the mH1 promoter
Pol II: RNA polymerase II
Pol III: RNA polymerase III
PSE: proximal sequence element
DSE: distal sequence element
BSA: Bovine serum albumin
MFI: mean fluorescence intensity