10906–10919 *Nucleic Acids Research, 2017, Vol. 45, No. 18* *Published online 30 August 2017*
doi: 10.1093/nar/gkx768

In silico design of context-responsive mammalian promoters with user-defined functionality

Adam J. Brown[1,*], Suzanne J. Gibson[2], Diane Hatton[2] and David C. James[1,*]

[1]Department of Chemical and Biological Engineering, University of Sheffield, Mappin St., Sheffield S1 3JD, UK and [2]Biopharmaceutical Development, MedImmune, Cambridge CB21 6GH, UK

Received May 23, 2017; Revised July 25, 2017; Editorial Decision August 20, 2017; Accepted August 22, 2017

ABSTRACT

Comprehensive *de novo*-design of complex mammalian promoters is restricted by unpredictable combinatorial interactions between constituent transcription factor regulatory elements (TFREs). In this study, we show that modular binding sites that do not function cooperatively can be identified by analyzing host cell transcription factor expression profiles, and subsequently testing cognate TFRE activities in varying homotypic and heterotypic promoter architectures. TFREs that displayed position-insensitive, additive function within a specific expression context could be rationally combined together *in silico* to create promoters with highly predictable activities. As TFRE order and spacing did not affect the performance of these TFRE-combinations, compositions could be specifically arranged to preclude the formation of undesirable sequence features. This facilitated simple *in silico*-design of promoters with context-required, user-defined functionalities. To demonstrate this, we *de novo*-created promoters for biopharmaceutical production in CHO cells that exhibited precisely designed activity dynamics and long-term expression-stability, without causing observable retroactive effects on cellular performance. The design process described can be utilized for applications requiring context-responsive, customizable promoter function, particularly where co-expression of synthetic TFs is not suitable. Although the synthetic promoter structure utilized does not closely resemble native mammalian architectures, our findings also provide additional support for a flexible billboard model of promoter regulation.

INTRODUCTION

Context-specific promoter-performance is a function of promoter activity dynamics, long-term gene expression cassette behavior (e.g. propensity for silencing, compatibility with other genetic components (1)), and promoter–cell interactions (e.g. off-target effects on cellular processes (2,3)). Given the inherent limitations of naturally-evolved sequences, for example undesirable sizes and unpredictable expression dynamics, synthetic promoters are typically preferred for most applications. The most common method of synthetic promoter construction is design of entirely synthetic systems, comprising synthetic transcription factors (TFs; i.e. zinc finger (4), transcription activator-like effector (5), chimeric (6) and CRISPR-TFs (7)) that transactivate elements containing their cognate binding sites (transcription factor regulatory elements (TFREs)). Utilization of synthetic and non-mammalian DNA-binding domains facilitates construction of promoters that impose minimal host cell interactions, and enables transcriptional outputs that are beyond the natural limitations of mammalian TF–TFRE pairs (e.g. fine-tuned, trigger-inducible transcriptional control (6,8,9)). Moreover, as these orthogonal systems are not designed to function in specific cell types or expression conditions, a single device can be used in diverse contexts. However, the associated metabolic burden and introduction of other exogenous recombinant protein(s) may limit the attractiveness of these systems in some applications, such as gene therapy and biopharmaceutical production.

When co-expression of synthetic TFs is not suitable, artificial promoters can be designed to interact with a host cell's existing repertoire of TFs. As different mammalian cell-types express unique and varying complements of TFs, the use of such promoters is typically limited to specific cell-types and/ or conditions (10,11). Further, as these promoters harness cellular TFs, they impose synthetic promoter-endogenous TF-endogenous promoter interactomes that can negatively affect the activity dynamics of both synthetic and endogenous promoters (2,12,13). For example, exogenous promoters harbouring copies of native-TFREs have been shown to compete for available TFs and titrate them away from endogenous genes, inducing changes in the host cell transcriptome (2). Elements causing such retroactivity effects (14) are incompatible with gene therapy and

*To whom correspondence should be addressed. Tel: +44 114 222 7505; Email: d.c.james@sheffield.ac.uk
Correspondence may also be addressed to Adam J. Brown. Tel: +44 114 222 7594; Email: adam.brown@sheffield.ac.uk

biopharmaceutical production processes, where promoter function must be coordinated with other desirable cellular functionalities (e.g. proliferation and cell survival). Promoters intended for these applications must also display long-term expression stability, which can be compromised by the presence of sequence features such as CpG motifs (methylation-mediated silencing (15,16)) and repeat elements (homologous recombination-mediated silencing (17–19)). However, using currently available promoter construction techniques, it is difficult to optimize promoter, expression cassette and host-cell performance simultaneously. Indeed, *de novo*-creation of mammalian promoters exhibiting precisely-designed activities in *vitro*/*vivo* has not previously been demonstrated.

Ideally, promoters could be designed to exhibit any user-defined set of functionalities by simply selecting and arranging an appropriate composition of TFREs, regardless of whether they interact with endogenous or synthetic TFs. Such comprehensive *in silico* design is apparently limited by the complex rules that govern promoter activity, including the orientation, spatial positioning, and order of composite TFREs, and the function, expression level, and activity of their cognate TFs (20,21). However, Smith *et al.* recently systematically tested the rules of TFRE organization in thousands of artificial sequences, and found that promoter activities were predominantly simply a function of relative TFRE copy numbers (22). These sequences therefore primarily functioned according to the billboard model of promoter regulation, where TFREs act as independent blocks with flexible positioning (23,24). When conforming to this regulatory model, constituent TFREs can be re-arranged in any configuration without affecting promoter activity dynamics, in the same way that symbols on a billboard can be arranged in any order without altering the total sum of information present. Alternatively, promoters can adhere to an enhancesome model of regulation, whereby TFREs must be strictly and specifically positioned in order to enable formation of protein interfaces (25–27). In reality, many promoters contain a mixture of position-flexible additive/ subtractive TFREs and position-sensitive cooperative TFREs (28). Indeed, in the promoters constructed by Smith et al. activities *were* impacted by combinatorial interactions between various TFRE-pairs (22). If composite TFREs are capable of synergistic interactions the complexity of *in silico* promoter design is substantially increased. However, studies have shown that combinatorial interactions between TFs are relatively uncommon (29). Moreover, Smith *et al.* only identified eight (out of a possible sixty six) combinatorial TFRE interactions among the twelve TFREs that were used for promoter construction (22). Therefore, it should be possible to specifically select combinations of modular TFREs that do not function cooperatively, and utilize them to build promoters *in silico* according to relatively simple design rules.

Given that TF concentration levels correlate well with cognate binding site activities (30,31), we hypothesized that *de novo*-design of context-specific, customizable promoters could be achieved by (i) profiling TF expression in the host cell, (ii) identifying TFREs that do not function cooperatively, and iii) determining the relative transcriptional activity of TF–TFRE interactions within heterotypic elements (i.e. the contribution to overall promoter activity provided by a single copy of each TFRE). Following this measure-model-manipulate paradigm we identified a pool of TFREs that displayed intra-promoter position-independent function within CHO cells, the preferred production host for therapeutic proteins. Accordingly, we were able to, for the first time, specifically select and arrange heterotypic TFRE-combinations *in silico* in order to create promoters with context-required, user-defined functionalities. To demonstrate this, we have *de novo*-designed promoters for use in the context of biopharmaceutical production that are protected from silencing, exhibit precisely designed activity dynamics, and impose no off-target effects on cell growth or viability. By showing that customizable 'billboard elements' can be designed to interact with host-cell machinery without causing retroactive effects, this study facilitates *de novo*-creation of context-responsive, optimized promoters for applications where the use of completely orthogonal synthetic TF–TFRE pairs is not suitable.

MATERIALS AND METHODS

Analysis of host cell TF expression dynamics

Total RNA was extracted from three CHO-K1-derived cell lines (parental host cell line, and host expressing either glutamine synthetase (GS) or GS and an IgG antibody) during exponential and stationary phases of growth using RNAeasy mini kits (Qiagen, Crawley, UK). RNA purity and integrity were confirmed using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Paisley, UK) and 2100 Bioanalyzer (Agilent Technologies, Wokingham, UK). RNA-seq libraries were prepared using the TruSeq RNA library preparation kit (Illumina, Essex, UK) and sequenced using an Illumina HiSeq 2000 system (Illumina). Sequence reads were mapped to the CHO-K1 reference genome using Tophat (32,33), and the relative abundance of each transcript was calculated using Cufflinks (34). A curated database of experimentally validated mouse TFs was obtained from TFcheckpoint (35,36). The mean transcript abundance of each TF gene across all six experimental conditions was determined, and genes with expression levels above the 70th percentile were selected for further analysis. Gene expression stability was measured by calculating the maximum fold change (MFC) in transcript abundance across all transcriptomes. Cognate binding sites of stably expressed TFs (MFC < 1.5) were obtained from previously published studies and online databases ((37); see supplementary information, Table S1).

Due to confidentiality restrictions, RNA-seq data from proprietary CHO cell lines cannot be deposited in public databases. However, the datasets can be obtained from the authors for non-commercial research purposes upon acceptance of a material transfer agreement.

In vitro-construction of homotypic and heterotypic promoters

A minimal CMV core promoter (see supplementary information, Figure S1) was synthesized (Sigma, Poole, UK) and inserted upstream of the secreted alkaline phosphatase (SEAP) gene in a previously described promoterless reporter vector (38). To construct homotypic TFRE-

reporters, synthetic oligonucleotides containing six repeat copies of a specific TFRE in series (see supplementary information, Table S1) were inserted upstream of the CMV core promoter. To create libraries of heterotypic promoters, TFRE building blocks containing a single copy of a discrete TF binding sequence were constructed as previously described (39), and ligated together in varying combinations with T4 DNA ligase (Thermo Fisher Scientific). A 'cloning-block' containing KpnI and XhoI restriction endonuclease sites was included in ligation mixes at a 1:20 molar ratio to TFRE blocks. Random TFRE block-assemblies were digested with KpnI and XhoI (Promega, Southampton, UK), gel extracted (Qiaquick gel extraction kit, Qiagen), and inserted upstream of the CMV core promoter in SEAP-reporter vectors. Plasmids were sequenced to determine the TFRE-composition of each *in vitro*-constructed synthetic promoter.

Modeling of heterotypic promoter activities

In vitro-constructed heterotypic promoter activities were modeled as a function of constituent TF binding site copy numbers. A comparison of different modeling approaches (linear regression, generalized linear, generalized additive, Gaussian process) determined that all models had equivalent predictive power. Accordingly, to minimize complexity, we used a multiple linear regression model $\hat{Y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_{12} x_{12}$ where $\hat{Y}$ represents promoter activity, and x_1–x_{12} are the copy numbers of 12 discrete TFRE blocks. Regression coefficients (β_1–β_{12}; calculated using least-squares estimation, $\hat{\beta} = (X^T X)^{-1} X^T y$) were analyzed to determine the relative transcriptional activity of a single copy of each TFRE block within heterotypic promoter architectures. The predictive ability of the model, the possibility of overfitting, and model robustness were assessed using leave-one-out and five-fold cross-validations. Overfitting was further investigated by calculating the coefficient of variation for each regression weight across all cross-validation models. Multicollinearity between predictor variables was assessed by determining variance inflation factors.

In silico design of heterotypic promoters

Every possible 1–14 block combination of twelve discrete TF binding sites ($n = 9\ 657\ 699$) was generated using the 'combinations' function in R. The relative transcriptional activity of each TFRE-combination was determined using our model of *in vitro*-constructed heterotypic promoter activities. TFRE-combinations with desired design criteria were selected from the library by applying successive filtration steps (as described in Results and Discussion). Constituent TFREs were arranged to minimize the occurrence of CpG dinucleotides, repeat sequences, and restriction endonuclease sites. To aid this process, binding sites were separated with specifically designed 6 bp spacer sequences. Designed promoter sequences were analysed for the presence of repeat sequences and endonuclease sites using FAIR (http://bioserver1.physics.iisc.ernet.in/fair/) and Webcutter (http://rna.lundberg.gu.se/cutter2/) (40). To confirm that unintended, additional TF binding sites had not been created at TFRE-spacer junctions, promoters were analyzed using MatInspector (https://www.genomatix.de/matinspector.html) and Transcription Affinity Prediction tool (TRAP: http://trap.molgen.mpg.de/cgi-bin/trap_form.cgi) (41,42). Designed sequences were synthesized (GeneArt, Regensburg, Germany) and cloned upstream of the minimal CMV core promoter in SEAP-reporter vectors.

CHO cell culture and transfection

Chinese hamster ovary (CHO) cells (CHO-K1-derived) were routinely cultured in CD-CHO medium (Thermo Fisher Scientific) at 37°C in 5% (v/v) CO_2 in vented Erlenmeyer flasks (Corning, UK), shaking at 140 rpm, and subcultured every 3–4 days at a seeding density of 2×10^5 cells/ml. Cell concentration and viability were determined by an automated Trypan Blue exclusion assay using a Vi-Cell cell viability analyser (Beckman-Coulter, High Wycombe, UK). Exponential and stationary phases of culture were determined by measuring viable cell concentrations every 24 h and calculating specific cell growth rates (d^{-1}). Two hours prior to transient transfections, 2×10^5 cells from a mid-exponential phase culture were seeded into individual wells of a 24-well plate (Nunc, Stafford, UK). Cells were transfected with DNA-lipid complexes comprising DNA and Lipofectamine (Thermo Fisher Scientific), prepared according to the manufacturer's instructions. Internal controls (hCMV-IE1-SEAP, SV40-SEAP, NFkB-RE-SEAP) were included in each plate to confirm reproducible transfection performance and normalize synthetic promoter activities. Transfected cells were incubated for 24 h prior to quantification of SEAP protein expression using the Sensolyte pNPP SEAP colorimetric reporter gene assay kit (Cambridge Biosciences, Cambridge, UK). To confirm that SEAP activities in cell culture supernatants were correlated with SEAP mRNA levels, total RNA was extracted from selected transfected cells and analysed by qPCR.

To construct stable pools, synthetic promoter-SEAP reporter plasmids (5 μg) coexpressing a glutamine synthetase selection marker gene were transfected into CHO cells (1×10^7; triplicate transfections) by electroporation using the Amaxa Nucleofector system (Lonza, Slough, UK; program U024). Stable transfectants were selected in 50 μM methionine sulfoximine (Sigma). For batch-production processes, 6×10^6 cells from a mid-exponential phase culture were inoculated into 30 ml CD-CHO medium in vented Erlenmeyer flasks. Cell concentration, culture viability and SEAP expression (at mRNA and protein levels) were measured during exponential (day 4) and stationary (day 7) growth phases. To validate long-term expression stability, 7-day batch-production processes were repeated after high and low producer stable pools had been subcultured in MSX-containing medium for 8 weeks (60 cell generations).

RNA extraction, reverse transcription and qPCR analysis

Total RNA was extracted from cells using RNeasy mini kits (Qiagen, UK). RNA purity and integrity were confirmed using a NanoDrop spectrophotometer (Thermo Fisher Scientific) and 2100 Bioanalyzer (Agilent Technologies). 800

ng of extracted RNA was reverse transcribed using the Quantitect reverse transcription kit (Qiagen), according to manufacturer's instructions (genomic DNA was eliminated during this procedure). cDNA was diluted 1:10 in nuclease free water prior to qPCR analysis using a 7500 fast real-time PCR system (Applied Biosystems, Cheshire, UK). Reaction mixtures containing 12.5 μl QuantiFast SYBR green PCR master mix (Qiagen), 2 μl cDNA, 2.5 μl primer mix (final concentration of 200 nM per primer), and 8 μl nuclease free water were prepared in MicroAmp fast optical 96-well plates (Applied Biosystems). Amplification conditions were as follows: 95°C for 5 min, followed by 40 cycles at 95°C for 15 s and 60°C for 60 s. Melting curve analysis was performed from 60 to 95°C. Reaction mixtures containing no template, or products from reverse transcription reactions performed in the absence of reverse transcriptase, were used as negative controls. All samples were run in triplicate and mean Ct (cycle threshold) values were used for further analysis. Relative SEAP mRNA levels were calculated using the $2^{-\Delta\Delta Ct}$ method (43), where Gnb1 and Fkbp1a were utilized as internal control reference genes as they exhibit highly stable levels of expression across the experimental conditions used in this study (data not shown). Primer amplification efficiencies were determined from standard curves (10-fold serial dilutions of pooled cDNA samples) using the equation $E = 10^{(-1/slope)}$. All primers had amplification efficiencies between 98% and 100% ($r^2 > 0.998$; primer sequences are listed in Supplementary Table S2).

RESULTS AND DISCUSSION

Analysis of host-cell TF expression dynamics facilitates identification of binding sites with context-specific functionalities

To demonstrate our process of custom promoter design we created sequences for use in CHO cells, the predominant host for the production of biopharmaceuticals. While transcriptional control has previously been demonstrated in CHO cells, *in vitro* construction methods have not enabled customizable specification of sequence features in order to prevent promoter silencing and minimize off-target effects on key cellular processes that underpin protein production (39,44). To profile the TF repertoire of CHO cells, we analyzed TF expression levels in six different experimental conditions, comprising three discrete CHO cell lines (CHO-K1 derived parental host cell line, and host expressing either glutamine synthetase (GS), or GS and an IgG antibody), sampled at exponential and stationary phases of culture. Given the inherent genetic instability of transformed mammalian cell lines, CHO cells that are subjected to cloning, selection and adaptation procedures typically display significant genetic/functional divergence (45–47). This was true for our transgenic derivatives, where in both cell lines over 1000 genes were differentially expressed (fold change > 1.5) compared to the parental host during exponential phase growth (data not shown). Given the difficulty of directly measuring effective TF concentrations (i.e. TFs that are appropriately modified and localized in the nucleus), we determined TF expression at the mRNA level. While this does not allow precise quantification of active TF levels, it does provide information on general TF expression patterns (e.g. no/low/high/differential expression), enabling identification of cognate TFREs with corresponding activity dynamics (30,31,48). Moreover, this method is easily applicable to promoter design for most mammalian cell types, for which transcriptomic datasets are typically available (49).

While it is estimated that mammalian genomes contain ∼2000 TF-encoding genes, only a fraction of these have been experimentally-verified as DNA-binding TFs (11,50). Accordingly, we restricted our analysis to the 774 TFs that have been shown to both exhibit sequence-specific DNA binding and regulate RNA polymerase II-dependent transcription (35). The mean expression level of each TF across six experimental conditions was determined. As shown in Figure 1, 388/774 TFs are expressed in CHO cells, where expression levels span over three orders of magnitude. Depending on required functionalities, synthetic promoters can be designed to interact with any combination of available host-cell TF-parts. For example, cell type-specificity could be achieved by designing promoters to bind TFs that are preferentially upregulated in the intended host cell, and specifically downregulated in cell types where off-target activity is undesirable. In this example, we aimed to create promoters that would have minimal impact on the CHO cell processes that underpin protein production, such as proliferation and cell survival. Therefore, we targeted TFs that are relatively highly expressed in CHO cells (ranked in the top 30% of TF mRNA expression levels), rationalizing that heterologous promoters can interact with these abundant cellular components without affecting the host cell transcriptome (i.e. these TFs are unlikely to become limiting if the nuclear copies of their cognate binding sites are moderately increased) (2,51). Further, we wanted designed promoters to exhibit stable activities in the context of different CHO cell lines and growth phases. We therefore focused our search on TFs that displayed high expression stability across all six CHO cell transcriptomes, measured as a maximum fold change (ratio between the highest and lowest expression level) of <1.5. Finally, to minimize the risk of silencing, we did not want promoters to interact with TFs that primarily function as repressors (52,53). Accordingly, we discounted TFs that have only been experimentally shown to negatively regulate transcription from RNA polymerase II promoters ((50); TFs shown to function as both transrepressors and transactivators were not discarded). Application of these selection criteria identified 67 CHO cell TFs with requisite expression profiles and functionalities (Figure 1A).

As shown in Figure 1B, due to binding site redundancy (overlap), the 67 identified TFs theoretically interact with 32 discrete regulatory elements. Amongst the TFRE-TFRE combinatorial partnerships recently identified by Smith *et al.*, none involved two TFREs that were both active in the context of homotypic promoters (22). We hypothesized that TFREs that are active in homotypic elements, and therefore do not require TFRE-partners to drive transcription, may be less likely to function cooperatively with each other when combined together in heterotypic architectures. Accordingly, to identify TFREs with potential position-insensitive function, we created secreted alkaline phosphatase (SEAP) reporter constructs that each contained six repeat copies of a specific TFRE in series, upstream of a minimal mam-

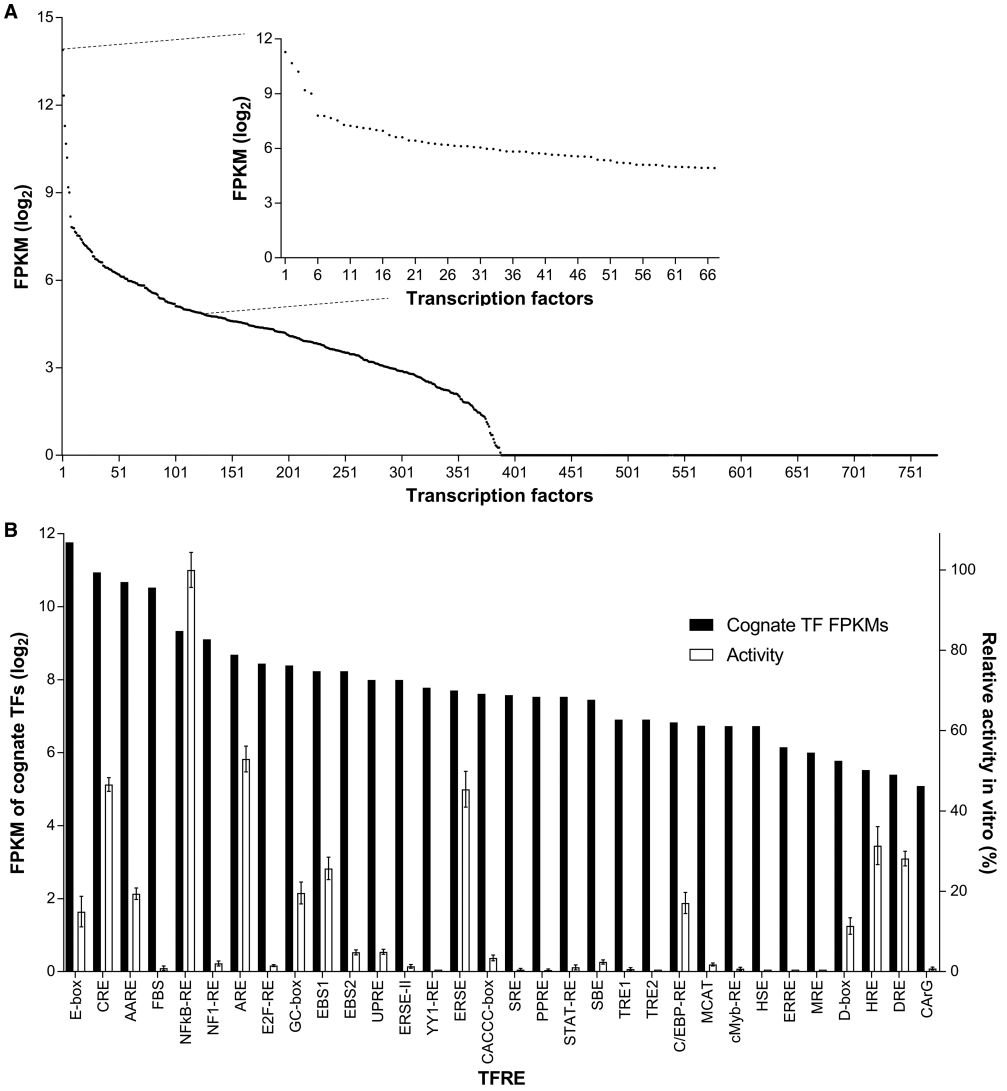

Figure 1. Identification of transcriptionally active transcription factor regulatory elements (TFREs) that bind relatively abundant host-cell components. (**A**) RNA-seq analysis of CHO cell transcriptomes determined the relative expression level of host-cell transcription factors (TFs). Points represent the average expression level of each TF in three discrete CHO cell lines, sampled at exponential and stationary phases of culture ($n = 6$). Inset graph shows the expression level of 67 TFs that exhibit high (ranked in the top 30% of TF mRNA expression levels) and stable expression across different CHO cell lines and growth phases (maximum fold change < 1.5). FPKM = fragments per kilobase of transcript per million fragments mapped. (**B**) Cognate binding sites of TFs with appropriate expression dynamics were identified and cloned in series (6× copies) upstream of a minimal CMV core promoter in secreted alkaline phosphatase (SEAP)-reporter vectors. CHO cells were transiently transfected with each homotypic TFRE-reporter and SEAP activity was measured 24 h post-transfection. Data are expressed as a percentage of the production exhibited by the strongest homotypic promoter. Bars represent the mean ± SD of three independent experiments ($n = 3$, each performed in triplicate). (**C**) An example homotypic promoter nucleotide sequence is shown. AARE sites are in bold, 6 bp spacer sequences are underlined, and the CMV core promoter is italicized.

malian core promoter (hCMV-IE1 core containing a TATA box and an initiator element; hCMV-IE1 core promoter and TFRE consensus sequences are shown in supplementary information, Figure S1 and Table S1, respectively). Measurement of SEAP reporter production after transient transfection of CHO cells with each homotypic TFRE-reporter plasmid showed that 12/32 TFREs could independently mediate activation of recombinant gene transcription (E-box, CRE, AARE, NFkB-RE, ARE, GC-box, EBS1, ERSE, C/EBP-RE, D-box, HRE, DRE). As depicted in Figure 1B, relative TFRE activities were not proportional to cognate TF expression levels. This may be explained by a lack of correlation between mRNA expression levels and effective TF concentrations. Additionally, the function of some TFs may be dependent on interactions with co-activators that are not sufficiently expressed in CHO cells. Moreover, some TFREs may require interactions with nearby enhancer elements or adjacent TFRE-

partners in order to drive transcription. TFRE activities may also be affected by the specific promoter structure utilized. For example, it has been shown that the transcriptional activity of a discrete TFRE in homotypic elements can vary between inactive, active, and highly-active, depending on the regulatory elements present in the downstream core promoter (54). We therefore concluded that the activity of individual TFREs in homotypic promoter architectures is highly context-specific and cannot be precisely predicted *in silico* from the mRNA abundance of cognate TFs. However, profiling cellular TFs at the mRNA level does facilitate identification of TF–TFRE pairs that have user-required functionalities within a specific expression-context.

TFREs that are active in homotypic architectures display position-insensitive function in heterotypic promoters

To test the hypothesis that TFREs that are transcriptionally active in homotypic promoters are likely to display position-insensitive function when combined together in the context of heterotypic architectures, we constructed promoters with varying TFRE-compositions. For each of the 12 TFREs identified as active in homotypic promoters, we synthesized oligonucleotide building blocks containing a single copy of the TF binding sequence. TFRE blocks were ligated to assemble random strings of TF binding sites, which were inserted upstream of the minimal CMV core promoter in SEAP reporter plasmids (see Methods, 'in-vitro construction of heterotypic promoters'). Composite-TFREs within each promoter were separated by the same 6 bp spacer sequence that was used in homotypic promoter construction (see Figure 1C). As shown in Figure 2A, seven distinct promoter libraries were constructed by mixing varying combinations of TFRE blocks. Library TFRE-compositions were designed to assess whether the function of individual TFREs was additive (i.e. adding additional copies of a discrete TFRE to a heterotypic element will increase overall promoter strength by a predictable, fixed amount), and unaffected by either relative spatial positioning (i.e. distance to the transcriptional start site) or the identify of neighbouring TFREs (i.e. cooperative interactions). Accordingly, to guarantee that each TFRE would occur in diverse contexts and copy numbers, individual TFREs were included in multiple libraries with varying accompanying elements. While we did not expect relative differences in TFRE activities to be maintained between homotypic and heterotypic architectures, we assumed that TFREs with higher activity in the former may be likely to drive stronger transactivation in the latter. Accordingly, in an attempt to test TFRE function over a wide range of transcriptional outputs, we used relative TFRE activities in homotypic elements to design library compositions such that each TFRE would occur within the context of variable-strength promoters. Unlike massively parallel reporter assays, this multi-library approach does not enable an exhaustive, systematic evaluation of TFRE functionality (55–57). However, it is an efficient method for this specific application, where we aimed to rigorously assess the position-sensitivity of TFRE function using a minimal number of test-promoters.

Measurement of SEAP production after transfection of CHO cells with 140 discrete synthetic promoter-reporter plasmids is shown in Figure 2B. These data show that promoter activities spanned two orders of magnitude, where the most active promoter exhibited a 2.3-fold increase in SEAP production over that deriving from a control vector containing the potent human cytomegalovirus immediate early 1 promoter (hCMV-IE1; GenBank accession number M60321.1, nucleotides 517–1193). With the exception of library 5, promoter activities within each library varied by at least an order of magnitude, where the mean activity of each library ranged from 5.2 to 61.5 relative promoter units (RPU). To check that SEAP enzymatic activity was an accurate proxy for transcriptional activity, SEAP mRNA levels in transfected cells were measured by qPCR. This analysis confirmed that within our transient expression system the SEAP production from each plasmid was directly proportional to relative promoter activities (supplementary information, Figure S2). Accordingly, to increase promoter screening throughput, we did not routinely perform qPCR analysis for experiments that utilized this expression system.

Given that TFREs were specifically selected for their putative position-insensitive function in heterotypic architectures, we hypothesized that promoter activities would simply be a function of the relative transcriptional activity contributed by each TF binding site, independent of TFRE spacing or order. To test this assumption, and determine the activity of each TFRE in heterotypic architectures, we sequenced promoters to reveal their TFRE-compositions, and modeled promoter activities as a function of TFRE copy numbers (promoters varied in length between 4 and 18 TF binding sites, mean = 9, and therefore discrete TFREs could occur in multiple copies per promoter)). The resulting linear regression model had high predictive power, where observed and predicted values for promoter activity were highly correlated (leave-one-out cross-validation $r^2 = 0.90$). To evaluate the robustness of the model, and assess the possibility of overfitting, we analyzed the model using five-fold cross-validation. The correlation between observed and expected promoter activities was similarly high ($r^2 = 0.87$; 80% of residuals < 5 RPU), validating the model's predictive ability. The regression coefficient of each predictor variable (i.e. TFRE copy numbers) did not significantly vary between cross-validation models (coefficient of variation < 10%), confirming that the model was not overfitted. Further, calculation of regression coefficient variance inflation factors (VIF) validated that multicollinearity was not an issue (all VIFs < 3 (58)). The data therefore show that promoter activities were predominantly a function of the type and quantity of constituent TFREs, and constructed sequences functioned as 'billboard promoters', where the relative organization of composite TF binding sites had minimal influence on promoter activity (24,59). Further, they confirm that the multi-library approach used in this study enables position-insensitive TFRE functions to be determined by testing a relatively small number of TFRE-combinations (we note that a sample size of 140 promoters adheres to the general rule that linear regression analysis requires at least ten observations per predictor variable (60)). We assume that the simplified, robust model of promoter regulation described will be generalizable to similar contexts where heterotypic mammalian promoters com-

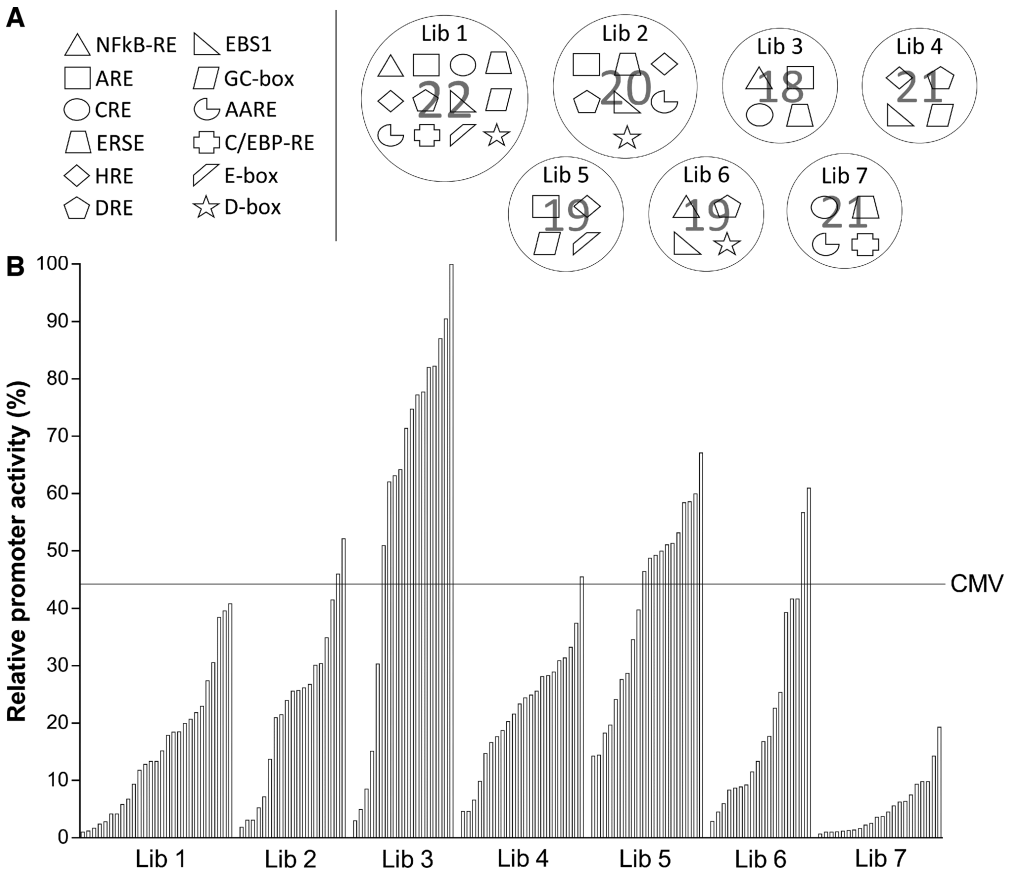

Figure 2. Heterotypic assemblies of modular transcription factor regulatory element (TFRE)-blocks exhibit transcriptional activities spanning over two orders of magnitude. (**A**) TFREs that were transcriptionally active in homotypic architectures (see Figure 1) were combined together in varying combinations to construct libraries of heterotypic promoters. Multiple constructs were created within each library, where the order, orientation, spatial positioning and copy number of composite-TFREs was varied (the total number of discrete promoters created within each library is shown). (**B**) Heterotypic elements were inserted upstream of a minimal CMV core promoter in secreted alkaline phosphatase (SEAP)-reporter vectors and transiently transfected into CHO cells. SEAP expression was quantified 24 h post-transfection. Data are expressed as a percentage of the production exhibited by the strongest heterotypic promoter. SEAP production from the control hCMV-IE1-SEAP reporter is shown as the black line. Each bar represents the mean of two transfections; for each promoter, <10% variation in SEAP production was observed. qPCR analysis of SEAP transcript abundance confirmed that relative protein activities in cell culture supernatants were linearly correlated with SEAP mRNA levels (see Supplementary Figure S2).

prise TFREs that do not function cooperatively within a specific cellular environment.

As the only predictor variables in our model are the number of copies of each TF binding site, the model coefficients represent the contribution a single copy of each TFRE makes to overall promoter activities. Analysis of model coefficients identified that within the context of heterotypic promoters only 7/12 TFREs were transcriptionally active (NFkB-RE, ARE, DRE, ERSE, GC-box, C/EBP-RE, EBS1; $P \leq 0.01$), where the remaining five were either transcriptionally inactive (AARE, HRE, E-box), or transcriptionally repressive (D-box, CRE; $P \leq 0.01$). As shown in Figure 3C, TFRE transcriptional activities in heterotypic architectures ranged from –35 to 100 relative TFRE-activity units (normalized TFRE activity values = regression coefficients × 8.9). Accordingly, activity in the context of homotypic promoters was not predictive of TFRE activity in heterotypic promoters. It is well known that TFREs can exhibit differential function dependent on the quantity of concatenated binding sites (20,22,61,62). Indeed, we did not use data from our homotypic promoter experiments to train the model of heterotypic promoter activities as we assumed that TFRE function may vary between the two organizational structures used (i.e. discrete binding sites occurred in 6× concatenated copies vs. predominantly single copies neighboured by differing TFREs). There are multiple potential mechanisms that may explain why AARE, HRE and D-box were active in homotypic clusters but inactive as individual copies within heterotypic elements. For example, cognate TFs for these TFREs may require multiple adjacent sites to facilitate stable binding via protein–protein interactions (62–64). Alternatively, TFRE clustering may increase localized cognate TF concentrations and alter their binding/ unbinding dynamics (63,65).

The finding that two TFREs not only lacked transactivation function but repressed transcription in heterotopic promoters indicates that cognate TFs for these sites can differentially function as activators or repressors depending on promoter context. Their repressor function in heterotypic architectures may be explained by bound TFs at these sites preventing TFs binding at active TFREs via steric hindrance or by altering DNA structures (e.g. bend-

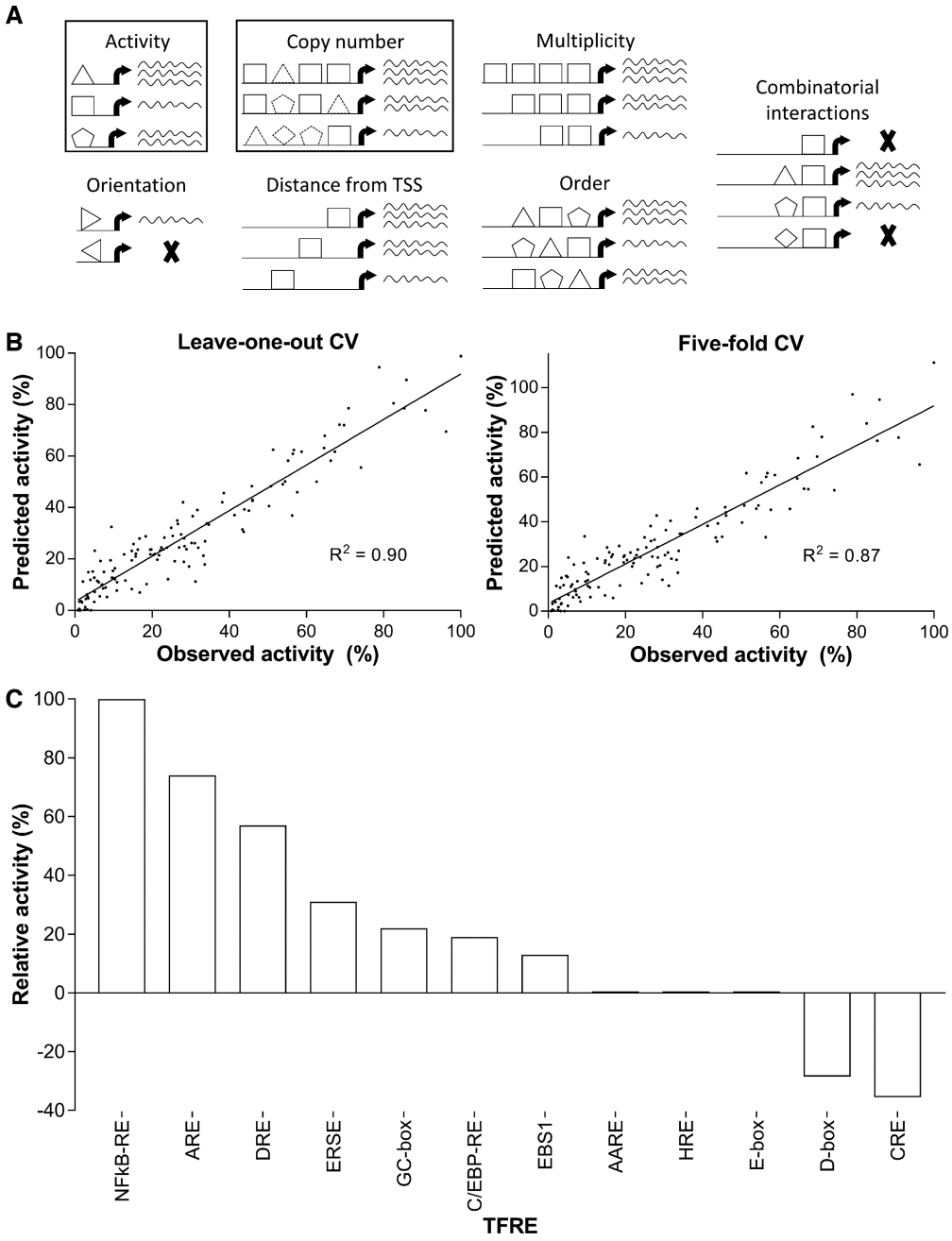

Figure 3. The function of modular transcription factor regulatory element (TFRE)-blocks in heterotypic promoter architectures is independent of binding site order and spacing. (**A**) The function of discrete TFREs within heterotypic elements can be influenced by multiple 'rules'. We modelled heterotypic promoter activities (see Figure 2) using TFRE copy numbers as the only predictor variable. (**B**) The linear regression model's predictive power was analyzed using leave-one-out and five-fold cross validations (CV). (**C**) The relative transcriptional activity of a single copy of each modular TFRE-block within heterotypic promoters was determined by analyzing the model coefficients. TFRE regression coefficients were multiplied by 8.9 to obtain normalized TFRE activities.

ing, stretching, supercoiling) (66,67). Although given then their function was largely position-independent according to our model, where a single copy of each TFRE contributed a fixed amount of transrepression to overall promoter activity, it may be more likely that TFs bound at these sites functioned via active, rather than passive, mechanisms (e.g. by interacting with components of the general transcriptional machinery (67)). While further experimental work (e.g. TFRE mutation studies) may have clarified the mechanisms by which neutral and repressive TFREs functioned, this was not required for *in silico* promoter design. Indeed, as the function of repressive TFREs was predominantly position-insensitive, albeit subtractive, they could still theoretically be utilized as modular building blocks in synthetic promoter construction.

The identification of seven TFREs that display the objective additive, position-insensitive functionality facilitates simple *in silico* promoter design. Before utilizing these

TFREs for *de novo* promoter creation, to confirm that we had rigorously tested possible position-dependent determinants of their function, we analyzed heterotypic promoter sequences to determine (i) the frequency at which each TFRE occurred in varying promoter positions (i.e. first TFRE upstream of the core promoter = position one), and (ii) the number of times each possible heterotypic TFRE-TFRE pair appeared. This analysis showed that all TFREs were represented in positions 1–9 at least five times (mean = 11; supporting information Table S3) and all TFRE pairs occurred on at least four occasions (mean = 11; supporting information Table S4). Due to the limitations of the promoter construction method used (i.e. randomly assembled TFRE strings, as opposed to specifically designed compositions (22)), positional features were not equally represented across the libraries. For example, the TFRE pairs DRE-EBS and EBS-C/EBP-RE were present in 35 and 4 copies respectively. Accordingly, the potential impact of under-represented features may have been hidden. However, as all features were tested in multiple varying contexts, and the model explaining heterotypic element activities had high predictive ability, we concluded that testing additional promoter variants was unnecessary.

In conclusion, our data indicate that while TFREs that were active in homotypic cluster promoters did display position-insensitive behavior in heterotypic elements, their function could be either additive, neutral or subtractive. For future studies it may therefore be preferable to initially assess TFRE activities in single site-copy homotypic promoters. However, the transcriptional output from a single TF binding site is rarely sufficient to drive detectable levels of recombinant gene expression (20,22). Accordingly, for identification of modular, additive TFREs that are transcriptionally active in single-copy repeats, we recommend the two-step screening method developed here, whereby, (i) multisite-copy homotypic promoters are created to identify TFREs that are unlikely to interact combinatorially, and (ii) multi TFRE-synthetic elements are constructed to determine the relative activity and function of each TFRE in heterotypic promoters.

In silico design of promoters with context-specific functionalities

Given that our model of promoter activities had both high predictive power and simple explanatory variables (i.e. TFRE copy numbers), we hypothesized that we could, for the first time, demonstrate the *in silico* design of mammalian promoters that exhibit predictable activities *in vitro*. To test this, we designed seven promoters containing a single copy of each TFRE that was transcriptionally active in heterotypic architectures (NFkB-RE, ARE, DRE, ERSE, GC-box, C/EBP-RE, EBS1), where constituent TF binding sites were arranged in completely different orders within each construct (Figure 4A). Previous studies have shown that promoters containing the same combination of TFREs can exhibit significantly different activities depending on the relative organisation of constituent binding sites (22,25,68). However, according to our model the function of our specifically selected TFREs is position-independent, and therefore promoters with identical TFRE-compositions should display the same level of activity (model-predicted activity of our seven 'test promoters' = 37.3 RPU). To confirm this, synthetic promoters were chemically synthesized and inserted upstream of the minimal CMV core promoter in SEAP reporter vectors. Measurement of SEAP production after transient transfection of CHO cells with each reporter plasmid showed that promoter activities ranged from 30.2 to 43.3 RPU (coefficient of variation = 9.8%). Therefore, the *in vitro* activity of all designed promoters was within 7 RPU of the predicted activity, corresponding to an error range of ±18%, where there was only a 1.4-fold difference in expression between the strongest and weakest promoter. These data therefore confirmed that TFRE order and spacing had minimal effect on promoter activities. Accordingly, we concluded that promoters with precisely-defined activities could be created *in silico* by simply selecting appropriate TFRE combinations and organizing them in any configuration.

Utilizing our twelve modular TFRE-parts, 9 657 699 TFRE-combinations can be created with total binding site copy numbers ranging from 1 to 14. Each of these combinations was constructed and tested *in silico,* using our model of heterotypic promoter activities to determine relative synthetic promoter strengths. This library shows a range of promoter activity between 2 and 157 RPUs and could be in used in a wide range of applications. As a proof of concept of the utility of this library, we selected combinations for use in biopharmaceutical production according to the context-specific promoter design criteria required. For example, to minimize the risk of promoter silencing, we discounted any combination containing a TFRE that was shown to exhibit transcriptional repressor function in heterotypic promoters (CRE, D-box; Figure 3C) (52,53). Further, given that recombinant protein overexpression in CHO cells can induce the unfolded protein response (69), we also disregarded all promoters containing the ER stress-response element (ERSE) in order to prevent formation of an ER-stress-recombinant gene expression positive feedback loop that could inhibit restoration of proteostasis (70,71). Although TFREs were specifically selected according to high expression of their cognate TFs in CHO cells, we assumed there could still be maximum threshold site copy numbers above which adding further copies may lead to levels of TF sequestration sufficient to cause changes in endogenous expression profiles (2,3,12,13). Accordingly, to prevent the possibility of synthetic elements causing off-target effects on key CHO cell processes, we (i) discounted promoters containing TFREs that were inactive in heterotypic architectures (i.e. prevented unnecessary, non-functional interactions with host TFs; AARE, HRE, E-box), (ii) limited the maximum copies of each TFRE per promoter to a relatively small number (≤ 5), and iii) selected combinations where the copy number of the most abundant constituent-TFRE was minimized (e.g. a promoter containing one copy of four different TFREs was preferred to a construct containing two copies of two different TFREs). We rationalized that this dual approach of minimizing TFRE quantities and targeting abundant cellular components would effectively eliminate the risk of affecting the host cell's native transcriptome. Finally, as the optimal rate of transcription varies for each recombinant gene, dependent on polypeptide-specific

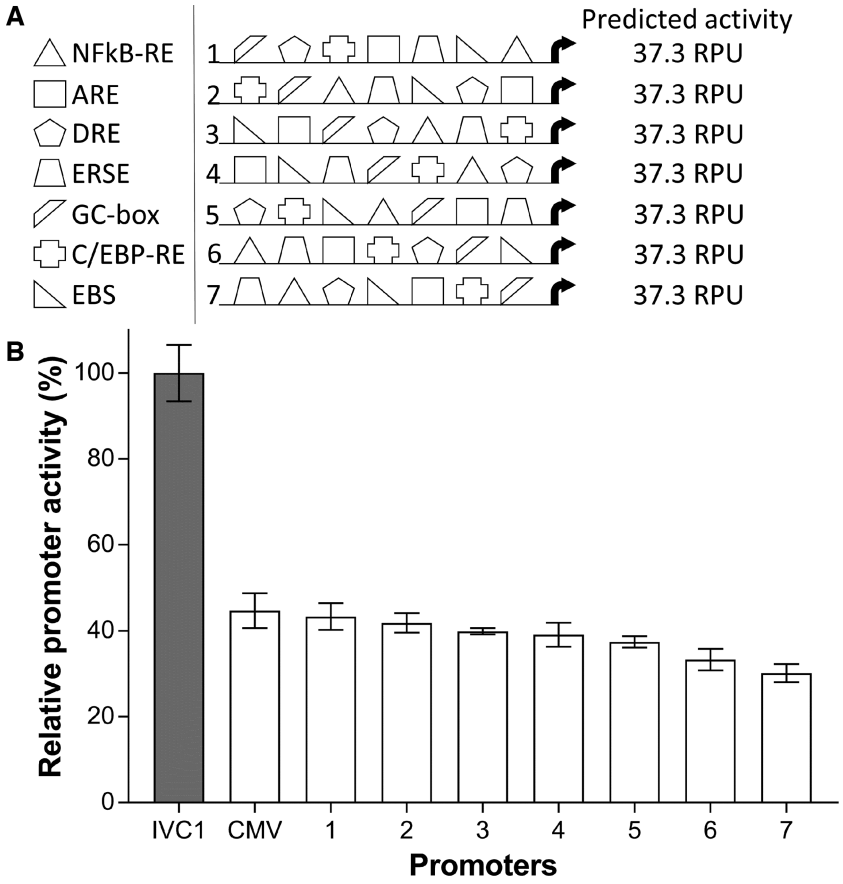

Figure 4. Synthetically designed promoters exhibit predictable activities *in vitro*. (**A**) Promoters with identical transcription factor regulatory element (TFRE)-compositions, but varying TFRE-orders, were designed *in silico*. The *in vitro* activity of each synthetic element was predicted using our model of heterotypic promoter activities (see Figure 3). (**B**) Synthetic promoters were chemically synthesized, inserted upstream of a minimal CMV core element in secreted alkaline phosphatase (SEAP)-reporter vectors, and transiently transfected into CHO cells. SEAP expression was quantified 24 h post-transfection. Data are expressed as a percentage of the production exhibited by the strongest *in vitro*-constructed heterotypic promoter, IVC1 (equivalent to 100 relative promoter units (RPU); see Figure 2). Values represent the mean ± SD of three independent experiments ($n = 3$, each performed in triplicate).

folding and assembly rates, we selected multiple discrete TFRE-combinations in order to enable a wide range of different transcriptional outputs (5, 10, 20, 40, 60, 80, and 100 RPU). As an example, a promoter with the TFRE-composition 2xARE: 1xC/EBP-RE: 1xGC-box: 1xEBS1: 1xDRE: 1xNFkB-RE was selected as it met all requisite design criteria and had a predicted activity of 40.2 RPU.

When TFRE organization has minimal effect on promoter activity, constituent TF binding sites can be optimally arranged to minimize the occurrence of undesirable sequence features. Accordingly, to maximise recombinant gene expression stability, we organized TFREs in configurations which prevented the formation of features associated with promoter silencing. Firstly, given that promoter methylation-mediated epigenetic silencing has been shown to cause production instability in CHO cells, we minimized the number of CpG dinucleotides within each construct (15,16). Further, as gene silencing can also be caused by deletion of DNA segments via homologous recombination (17,72), we specifically prevented the occurrence of repeat sequences. Given that eukaryotic machinery can recombine identical sequences longer than 40 bp, we reasoned that preventing intra-promoter repeats larger than 20 bp, and avoiding the repetition of any two-string TFRE block (e.g. ARE-DRE), would provide robust protection against homologous recombination-mediated silencing (18). Finally, to protect against recombination-mediated gene deletion when multiple promoters are used in conjunction (e.g. expression of monoclonal antibodies), inter-promoter repeat sequences larger than 35 bp were specifically precluded (19,73,74). Exclusion of undesirable features was further facilitated by the ability to specifically design the spacer sequences separating composite TFREs within each promoter (to maintain a consistent promoter structure between *in vitro*-constructed and *in silico*-designed elements, all spacers were designed to be 6 bp in length). Utilizing *in silico* arrangement of TFREs and spacers, our synthetically designed promoters contain an average of 5.2 CpG dinucleotides and 0 intra-promoter repeats >20 bp. In comparison our *in vitro*-constructed heterotypic promoters (mean per promoter = 20.7 CpG dinucleotides and 3.8 intra-promoter repeats > 20 bp), and the hCMV-IE1 promoter (34 CpG dinucleotides and 1 repeat > 20 bp), contain significantly higher quantities of silencing-associated sequence features. To facilitate cloning into diverse expression vectors, promoters were also designed to minimize the

occurrence of restriction endonuclease sites (263/308 analyzed restriction sites do not occur in any promoter). Lastly, to prevent improper regulation of promoter activity, all sequences were designed to ensure that additional, 'accidental TF binding sites' were not created at TFRE-spacer junctions.

Custom-designed sequences exhibit predictable functionalities *in vitro*

For each desired promoter activity level we designed two synthetic sequences with different TFRE-compositions (supplementary information, Table S5). Synthetic promoters were chemically synthesized and inserted upstream of the minimal CMV core promoter in SEAP reporter vectors. Measurement of SEAP production after transient transfection of CHO cells with each reporter plasmid showed that designed and observed activities were highly correlated ($r^2 = 0.92$; Figure 5B). At very low–high levels of transcription (5–60 RPU), the *in vitro* activity of all promoters was within 5 RPU of predicted activities. However, when transcriptional output was very high (80–100 RPU), the difference between observed and predicted activities varied by 10–22 RPU (11–22%). It has previously been shown that transcriptional noise increases concomitantly with both promoter activity and TF binding site copy number (75,76). This may explain why the four strongest promoters (with the largest TFRE copy numbers) exhibited the greatest deviation between observed and predicted activities. However, as all promoters with designed activities ≥80 exhibited activities ≥78 *in vitro*, we concluded that very strong promoters *can* be routinely created *in silico*, but that at this level of expression very precise control of transcription may be intractable.

To test the function of our designed sequences in an intragenomic context, we stably transfected CHO cells with synthetic promoter-reporter plasmids coexpressing a glutamine synthetase selection marker gene (77). In order to analyze the full range of transcriptional control, we evaluated promoters with activities of 5, 10, 20, 80 and 100 RPU. To ensure that gene expression variability was directly linked to differences in promoter activity, rather than site-specific integration effects, we employed a stable pool system where recombinant vectors randomly integrate into varying chromosomal locations to create heterogeneous cell populations (each pool was generated using a single synthetic promoter-reporter plasmid). Stably transfected CHO cell pools were selected in medium containing methionine sulphoximine and re-adapted to suspension culture. To evaluate promoter performance in an industrially-relevant bioproduction context, promoter activities were measured during a 7-day batch-production process. As shown in Figure 5C, the promoter activities observed in transient expression systems were maintained in chromosomal contexts. qPCR analysis of SEAP mRNA abundance revealed that the ratio of relative promoter strengths in stable expression systems (100:72:28:16:8) was highly similar to our original designed ratio of promoter activities (100:80:20:10:5). Moreover, relative promoter activities were maintained between exponential (day 4) and stationary (day 7) phases of growth. These data confirm our assumption that promoter activity dynamics can be specifically tailored by designing sequences to bind TFs with synchronous expression profiles. Further, no synthetic promoter had a significant effect on cell growth or viability (viable cell concentration and culture viability varied by <20% between all cell pools at days 4 and 7; data not shown), validating our selection of TFRE-combinations that were specifically designed to minimize off-target effects on cellular performance. Finally, to assess gene expression stability, high and low producer stable pools were subcultured in MSX-containing medium for sixty generations. As shown in Figure 5D SEAP production was not significantly reduced following long-term culture, confirming that synthetic sequences had been successfully designed to prevent promoter silencing.

CONCLUSION

In conclusion, we have, for the first time, demonstrated *in silico* construction of mammalian promoters that exhibit precisely designed activities *in vitro*. By analysing both host cell TF levels and cognate TFRE activities, we were able to build promoters using TF–TFRE pairs that exhibited desired activity dynamics without imposing retroactive effects on cellular performance. Over-activity of these TFs/TFREs is generally associated with cancerous phenotypes, and we therefore assume that their high levels of activity in CHO cells are required to maintain uncontrolled proliferation and suppression of apoptosis (78–83). While we did not experimentally verify which TFs interacted with each binding site, given that synthetic elements functioned as designed, we assume that unpredicted TF–TFRE interactions were either uncommon or did not negatively impact expected TFRE/cell functions. By identifying TFREs that function via positon-independent mechanisms within a specific expression-context, we were able to select and organize TFRE-compositions *in silico* in order to simultaneously optimize promoter, expression cassette and host cell performance.

Utilization of TFREs with position-insensitive function facilitated construction of a simple model explaining heterotypic element activities that has higher predictive power than any previously published model of mammalian promoter activity. The synthetic promoter architecture used is significantly different from native eukaryotic structures, where TFRE-flanking sites, interactions with enhancers, and complex regulatory mechanisms (e.g. inducibility) can effect promoter activity dynamics. However, our findings do provide further support that mammalian promoters can function according to the billboard model of regulation. Given the simplified model of promoter regulation described, it may soon be possible to design promoters entirely *in silico* from 'OMICS' datasets, obviating the requirement for *in vitro* screening. However, this will likely require a detailed understanding of how many discrete TF–TFRE interactions function within the context of varying promoter architectures and expression contexts. The design process described in this study can be applied to *de novo*-create optimized context-responsive promoter sequences for any specific host cell-type or expression context where expression of exogenous synthetic TFs is not suitable. While our designed sequences access endogenous TFs, the described method

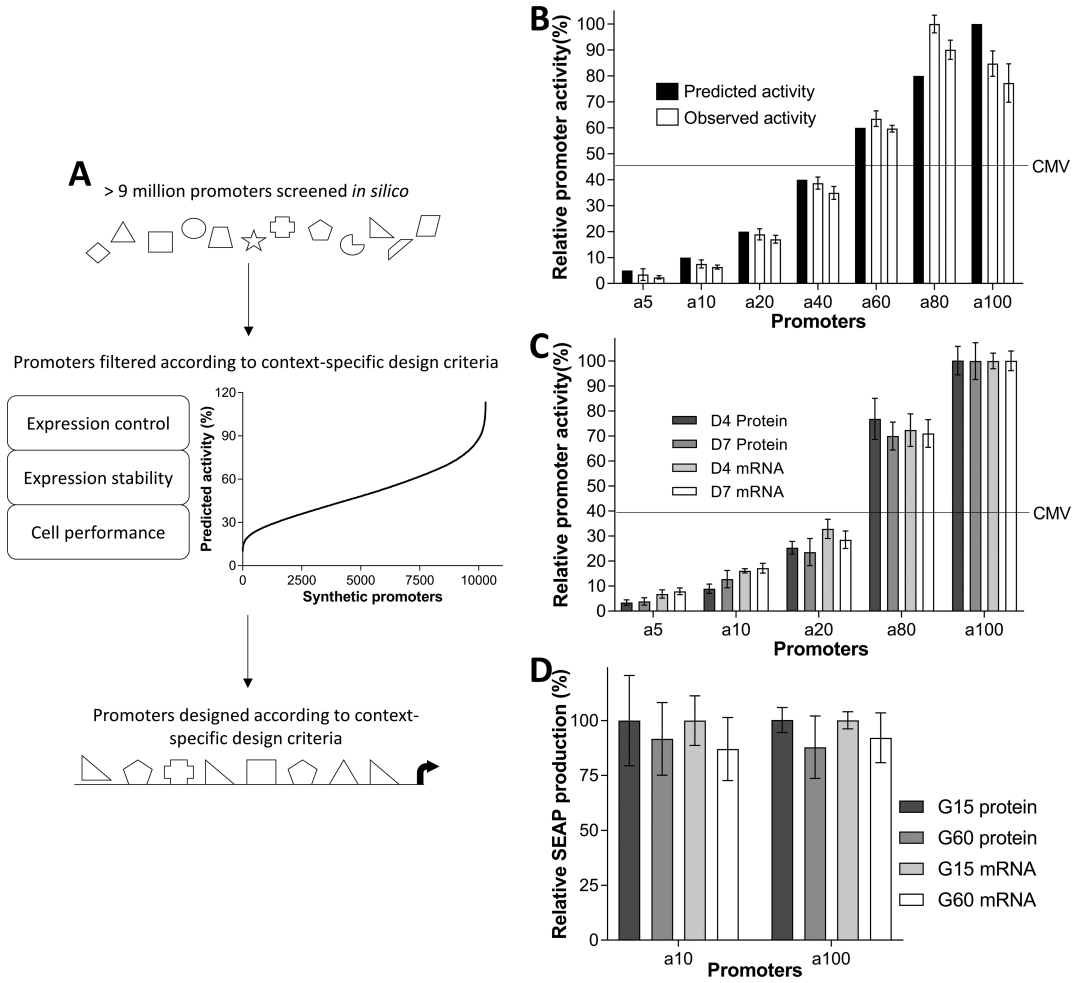

Figure 5. *In silico* designed sequences exhibit custom-defined functionalities *in vitro.* (**A**) Millions of transcription factor regulatory element (TFRE)-combinations were constructed and tested *in silico* using our model of heterotypic element activities (see Figure 3). Selection criteria were applied to identify combinations optimal for the context of biopharmaceutical production in CHO cells. Constituent TFREs within each promoter were then specifically arranged to prevent occurrences of sequence features that can contribute to promoter silencing. (**B**) Synthetic promoters with varying designed activities were chemically synthesized, inserted upstream of a minimal CMV core element in secreted alkaline phosphatase (SEAP)-reporter vectors, and transiently transfected into CHO cells. SEAP expression was quantified 24 h post-transfection. Data are expressed as a percentage of the production exhibited by the strongest promoter. (**C**) CHO cells were stably transfected with synthetic promoter-reporter plasmids coexpressing a glutamine synthetase selection marker gene. Three distinct stable pools were created for each reporter-plasmid, where recombinant vectors randomly integrate into varying chromosomal locations to create heterogeneous cell populations. Following selection in medium containing methionine sulfoximine, promoter activities were measured during a 7-day batch-production process. SEAP titer and mRNA abundance were determined in mid-exponential and stationary phases of growth. Data are expressed as a percentage of the expression exhibited by the strongest promoter. (**D**) Stable pools were subcultured in selective medium for sixty generations. SEAP expression was quantified at the end of a 7-day batch-production process. Data are expressed as a percentage of the production exhibited by each pool at generation fifteen. Values represent the mean ± SD of three independent experiments ($n = 3$, each performed in triplicate).

of modular-regulatory element identification and assembly may also aid the construction of genetic circuits that utilize multiple completely synthetic TF–TFRE pairs.

DATA AVAILABILITY

The sequences of *in silico* designed synthetic promoters have been deposited with the DNA Data Bank of Japan (accession numbers LC270626–LC270639).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Dr Tarik Senussi for construction of recombinant cell lines, Dr Darren Geoghegan for construction of RNA-seq libraries, and Yash Patel for assistance with generation of stable pools.

FUNDING

Funding for open access charge: MedImmune.
Conflict of interest statement. The authors have a patent application filed based on the work in this paper.

REFERENCES

1. Mutalik,V.K., Guimaraes,J.C., Cambray,G., Lam,C., Christoffersen,M.J., Mai,Q.-A., Tran,A.B., Paull,M., Keasling,J.D. and Arkin,A.P. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods*, **10**, 354–360.
2. Brewster,R.C., Weinert,F.M., Garcia,H.G., Song,D., Rydenfelt,M. and Phillips,R. (2014) The transcription factor titration effect dictates level of gene expression. *Cell*, **156**, 1312–1323.
3. Karreth,F.A., Tay,Y. and Pandolfi,P.P. (2014) Target competition: transcription factors enter the limelight. *Genome Biol.*, **15**, 114.
4. Gaj,T., Gersbach,C.A. and Barbas,C.F. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.*, **31**, 397–405.
5. Perez-Pinera,P., Ousterout,D.G., Brunger,J.M., Farin,A.M., Glass,K.A., Guilak,F., Crawford,G.E., Hartemink,A.J. and Gersbach,C.A. (2013) Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat. Methods*, **10**, 239–242.
6. Rössger,K., Charpin-El-Hamri,G. and Fussenegger,M. (2014) Bile acid-controlled transgene expression in mammalian cells and mice. *Metab. Eng.*, **21**, 81–90.
7. Chavez,A., Scheiman,J., Vora,S., Pruitt,B.W., Tuttle,M., Iyer,E.P., Lin,S., Kiani,S., Guzman,C.D. and Wiegand,D.J. (2015) Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods*, **12**, 326–328.
8. Gitzinger,M., Kemmer,C., Fluri,D.A., El-Baba,M.D., Weber,W. and Fussenegger,M. (2011) The food additive vanillic acid controls transgene expression in mammalian cells and mice. *Nucleic Acids Res.*, **40**, e37.
9. Müller,K., Engesser,R., Schulz,S., Steinberg,T., Tomakidi,P., Weber,C.C., Ulm,R., Timmer,J., Zurbriggen,M.D. and Weber,W. (2013) Multi-chromatic control of mammalian gene expression and signaling. *Nucleic Acids Res.*, **41**, e124–e124.
10. Schlabach,M.R., Hu,J.K., Li,M. and Elledge,S.J. (2010) Synthetic design of strong promoters. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 2538–2543.
11. Vaquerizas,J.M., Kummerfeld,S.K., Teichmann,S.A. and Luscombe,N.M. (2009) A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.*, **10**, 252–263.
12. Hansen,A.S. and O'Shea,E.K. (2013) Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression. *Mol. Syst. Biol.*, **9**, 704.
13. Borkowski,O., Ceroni,F., Stan,G.-B. and Ellis,T. (2016) Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Curr. Opin. Microbiol.*, **33**, 123–130.
14. Del Vecchio,D., Ninfa,A.J. and Sontag,E.D. (2008) Modular cell biology: retroactivity and insulation. *Mol. Syst. Biol.*, **4**, 161.
15. Kim,M., O'Callaghan,P.M., Droms,K.A. and James,D.C. (2011) A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. *Biotechnol. Bioeng.*, **108**, 2434–2446.
16. Yang,Y., Chusainow,J. and Yap,M.G. (2010) DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. *J. Biotechnol.*, **147**, 180–185.
17. Jasin,M. and Rothstein,R. (2013) Repair of strand breaks by homologous recombination. *Cold Spring Harb. Perspect. Biol.*, **5**, a012740.
18. Baudin,A., Ozier-Kalogeropoulos,O., Denouel,A., Lacroute,F. and Cullin,C. (1993) A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. *Nucleic Acids Res.*, **21**, 3329–3330.
19. Sleight,S.C., Bartley,B.A., Lieviant,J.A. and Sauro,H.M. (2010) Designing and engineering evolutionary robust genetic circuits. *J. Biol. Eng.*, **4**, 12.
20. Sharon,E., Kalma,Y., Sharp,A., Raveh-Sadka,T., Levo,M., Zeevi,D., Keren,L., Yakhini,Z., Weinberger,A. and Segal,E. (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.*, **30**, 521–530.
21. Weingarten-Gabbay,S. and Segal,E. (2014) The grammar of transcriptional regulation. *Hum. Genet.*, **133**, 701–711.
22. Smith,R.P., Taher,L., Patwardhan,R.P., Kim,M.J., Inoue,F., Shendure,J., Ovcharenko,I. and Ahituv,N. (2013) Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nat. Genet.*, **45**, 1021–1028.
23. Kulkarni,M.M. and Arnosti,D.N. (2003) Information display by transcriptional enhancers. *Development*, **130**, 6569–6575.
24. Arnosti,D.N. and Kulkarni,M.M. (2005) Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J. Cell. Biochem.*, **94**, 890–898.
25. Swanson,C.I., Evans,N.C. and Barolo,S. (2010) Structural rules and complex regulatory circuitry constrain expression of a Notch-and EGFR-regulated eye enhancer. *Dev. Cell*, **18**, 359–370.
26. Thanos,D. and Maniatis,T. (1995) Virus induction of human IFNβ gene expression requires the assembly of an enhanceosome. *Cell*, **83**, 1091–1100.
27. Merika,M. and Thanos,D. (2001) Enhanceosomes. *Curr. Opin. Genet. Dev.*, **11**, 205–208.
28. Spitz,F. and Furlong,E.E. (2012) Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.*, **13**, 613–626.
29. Ravasi,T., Suzuki,H., Cannistraci,C.V., Katayama,S., Bajic,V.B., Tan,K., Akalin,A., Schmeier,S., Kanamori-Katayama,M. and Bertin,N. (2010) An atlas of combinatorial transcriptional regulation in mouse and man. *Cell*, **140**, 744–752.
30. Gertz,J. and Cohen,B.A. (2009) Environment-specific combinatorial cis-regulation in synthetic promoters. *Mol. Syst. Biol.*, **5**, 244.
31. Segal,E., Raveh-Sadka,T., Schroeder,M., Unnerstall,U. and Gaul,U. (2008) Predicting expression patterns from regulatory sequence in Drosophila segmentation. *Nature*, **451**, 535–540.
32. Trapnell,C., Pachter,L. and Salzberg,S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.
33. Xu,X., Nagarajan,H., Lewis,N.E., Pan,S., Cai,Z., Liu,X., Chen,W., Xie,M., Wang,W. and Hammond,S. (2011) The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.*, **29**, 735–741.
34. Trapnell,C., Williams,B.A., Pertea,G., Mortazavi,A., Kwan,G., Van Baren,M.J., Salzberg,S.L., Wold,B.J. and Pachter,L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, **28**, 511–515.
35. Chawla,K., Tripathi,S., Thommesen,L., Lægreid,A. and Kuiper,M. (2013) TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. *Bioinformatics*, **29**, 2519–2520.
36. Tripathi,S., Vercruysse,S., Chawla,K., Christie,K.R., Blake,J.A., Huntley,R.P., Orchard,S., Hermjakob,H., Thommesen,L. and Lægreid,A. (2016) Gene regulation knowledge commons: community action takes care of DNA binding transcription factors. *Database*, **2016**, baw088.
37. Mathelier,A., Zhao,X., Zhang,A.W., Parcy,F., Worsley-Hunt,R., Arenillas,D.J., Buchman,S., Chen,C.-y., Chou,A. and Ienasescu,H. (2013) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **42**, D142–D147.
38. Brown,A.J., Mainwaring,D.O., Sweeney,B. and James,D.C. (2013) Block decoys: transcription-factor decoys designed for in vitro gene regulation studies. *Anal. Biochem.*, **443**, 205–210.
39. Brown,A.J., Sweeney,B., Mainwaring,D.O. and James,D.C. (2014) Synthetic promoters for CHO cell engineering. *Biotechnol. Bioeng.*, **111**, 1638–1647.
40. Senthilkumar,R., Sabarinathan,R., Hameed,B.S., Banerjee,N., Chidambarathanu,N., Karthik,R. and Sekar,K. (2010) FAIR: A server for internal sequence repeats. *Bioinformation*, **4**, 271.
41. Manke,T., Roider,H.G. and Vingron,M. (2008) Statistical modeling of transcription factor binding affinities predicts regulatory interactions. *PLoS Comput. Biol.*, **4**, e1000039.
42. Cartharius,K., Frech,K., Grote,K., Klocke,B., Haltmeier,M., Klingenhoff,A., Frisch,M., Bayerlein,M. and Werner,T. (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*, **21**, 2933–2942.
43. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*, **25**, 402–408.
44. Tornøe,J., Kusk,P., Johansen,T.E. and Jensen,P.R. (2002) Generation of a synthetic mammalian promoter library by modification of

sequences spacing transcription factor binding sites. *Gene*, **297**, 21–32.
45. Davies,S.L., Lovelady,C.S., Grainger,R.K., Racher,A.J., Young,R.J. and James,D.C. (2013) Functional heterogeneity and heritability in CHO cell populations. *Biotechnol. Bioeng.*, **110**, 260–274.
46. Korke,R., Rink,A., Seow,T.K., Chung,M.C., Beattie,C.W. and Hu,W.-S. (2002) Genomic and proteomic perspectives in cell culture engineering. *J. Biotechnol.*, **94**, 73–92.
47. Derouazi,M., Martinet,D., Schmutz,N.B., Flaction,R., Wicht,M., Bertschinger,M., Hacker,D., Beckmann,J. and Wurm,F. (2006) Genetic characterization of CHO production host DG44 and derivative recombinant cell lines. *Biochem. Biophys. Res. Commun.*, **340**, 1069–1077.
48. Gertz,J., Siggia,E.D. and Cohen,B.A. (2009) Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature*, **457**, 215–218.
49. Sheng,X., Wu,J., Sun,Q., Li,X., Xian,F., Sun,M., Fang,W., Chen,M., Yu,J. and Xiao,J. (2016) MTD: a mammalian transcriptomic database to explore gene expression and regulation. *Brief. Bioinform.*, **18**, 28–36.
50. Tripathi,S., Christie,K.R., Balakrishnan,R., Huntley,R., Hill,D.P., Thommesen,L., Blake,J.A., Kuiper,M. and Lægreid,A. (2013) Gene Ontology annotation of sequence-specific DNA binding transcription factors: setting the stage for a large-scale curation effort. *Database*, **2013**, bat062.
51. van Dijk,D., Sharon,E., Lotan-Pompan,M., Weinberger,A., Segal,E. and Carey,L.B. (2017) Large-scale mapping of gene regulatory logic reveals context-dependent repression by transcriptional activators. *Genome Res.*, **27**, 87–94.
52. Wajapeyee,N., Malonia,S.K., Palakurthy,R.K. and Green,M.R. (2013) Oncogenic RAS directs silencing of tumor suppressor genes through ordered recruitment of transcriptional repressors. *Genes Dev.*, **27**, 2221–2226.
53. Smith,Z.D. and Meissner,A. (2013) DNA methylation: roles in mammalian development. *Nat. Rev. Genet.*, **14**, 204–220.
54. Juven-Gershon,T. and Kadonaga,J.T. (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev. Biol.*, **339**, 225–229.
55. Kheradpour,P., Ernst,J., Melnikov,A., Rogov,P., Wang,L., Zhang,X., Alston,J., Mikkelsen,T.S. and Kellis,M. (2013) Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res.*, **23**, 800–811.
56. Melnikov,A., Murugan,A., Zhang,X., Tesileanu,T., Wang,L., Rogov,P., Feizi,S., Gnirke,A., Callan,C.G. Jr and Kinney,J.B. (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.*, **30**, 271–277.
57. Mogno,I., Kwasnieski,J.C. and Cohen,B.A. (2013) Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants. *Genome Res.*, **23**, 1908–1915.
58. Hair,J.F., Black,W.C., Babin,B.J., Anderson,R.E. and Tatham,R.L. (1998) *Multivariate Data Analysis*. Prentice Hall, New Jersey.
59. Rastegar,S., Hess,I., Dickmeis,T., Nicod,J.C., Ertzer,R., Hadzhiev,Y., Thies,W.-G., Scherer,G. and Strähle,U. (2008) The words of the regulatory code are arranged in a variable manner in highly conserved enhancers. *Dev. Biol.*, **318**, 366–377.
60. Harrell,F. (2015) *Regression Modeling Strategies: with Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer, NY.
61. Grskovic,M., Chaivorapol,C., Gaspar-Maia,A., Li,H. and Ramalho-Santos,M. (2007) Systematic identification of cis-regulatory sequences active in mouse and human embryonic stem cells. *PLoS Genet.*, **3**, e145.
62. Giniger,E. and Ptashne,M. (1988) Cooperative DNA binding of the yeast transcriptional activator GAL4. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 382–386.
63. Ezer,D., Zabet,N.R. and Adryan,B. (2014) Homotypic clusters of transcription factor binding sites: A model system for understanding the physical mechanics of gene expression. *Comput. Struct. Biotechnol. J.*, **10**, 63–69.
64. Chu,D., Zabet,N.R. and Mitavskiy,B. (2009) Models of transcription factor binding: sensitivity of activation functions to model assumptions. *J. Theor. Biol.*, **257**, 419–429.
65. Ezer,D., Zabet,N.R. and Adryan,B. (2014) Physical constraints determine the logic of bacterial promoter architectures. *Nucleic Acids Res.*, **42**, 4196–4207.
66. Struhl,K. (1989) Molecular mechanisms of transcriptional regulation in yeast. *Annu. Rev. Biochem.*, **58**, 1051–1077.
67. Gaston,K. and Jayaraman,P.-S. (2003) Transcriptional repression in eukaryotes: repressors and repression mechanisms. *Cell. Mol. Life Sci.*, **60**, 721–741.
68. Senger,K., Armstrong,G.W., Rowell,W.J., Kwan,J.M., Markstein,M. and Levine,M. (2004) Immunity regulatory DNAs share common organizational features in Drosophila. *Mol. Cell*, **13**, 19–32.
69. Hussain,H., Maldonado-Agurto,R. and Dickson,A.J. (2014) The endoplasmic reticulum and unfolded protein response in the control of mammalian recombinant protein production. *Biotechnol. Lett.*, **36**, 1581–1593.
70. Gorman,A.M., Healy,S.J., Jäger,R. and Samali,A. (2012) Stress management at the ER: regulators of ER stress-induced apoptosis. *Pharmacol. Ther.*, **134**, 306–316.
71. Sano,R. and Reed,J.C. (2013) ER stress-induced cell death mechanisms. *BBA-Mol. Cell Res.*, **1833**, 3460–3470.
72. Moynahan,M.E. and Jasin,M. (2010) Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat. Rev. Mol. Cell Biol.*, **11**, 196–207.
73. Lambert,S., Saintigny,Y., Delacote,F., Amiot,F., Chaput,B., Lecomte,M., Huck,S., Bertrand,P. and Lopez,B. (1999) Analysis of intrachromosomal homologous recombination in mammalian cell, using tandem repeat sequences. *Mutat. Res., DNA Repair*, **433**, 159–168.
74. Read,L.R., Raynard,S.J., Rukść,A. and Baker,M.D. (2004) Gene repeat expansion and contraction by spontaneous intrachromosomal homologous recombination in mammalian cells. *Nucleic Acids Res.*, **32**, 1184–1196.
75. Sharon,E., van Dijk,D., Kalma,Y., Keren,L., Manor,O., Yakhini,Z. and Segal,E. (2014) Probing the effect of promoters on noise in gene expression using thousands of designed sequences. *Genome Res.*, **24**, 1698–1706.
76. Murphy,K.F., Balázsi,G. and Collins,J.J. (2007) Combinatorial promoter design for engineering noisy gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 12726–12731.
77. Cockett,M., Bebbington,C. and Yarranton,G. (1990) High level expression of tissue inhibitor of metalloproteinases in Chinese hamster ovary cells using glutamine synthetase gene amplification. *Nat. Biotechnol.*, **8**, 662–667.
78. Beishline,K. and Azizkhan-Clifford,J. (2015) Sp1 and the 'hallmarks of cancer'. *FEBS J.*, **282**, 224–258.
79. Dolcet,X., Llobet,D., Pallares,J. and Matias-Guiu,X. (2005) NF-kB in development and progression of human cancer. *Virchows Arch.*, **446**, 475–482.
80. Kansanen,E., Kuosmanen,S.M., Leinonen,H. and Levonen,A.-L. (2013) The Keap1-Nrf2 pathway: mechanisms of activation and dysregulation in cancer. *Redox Biol.*, **1**, 45–49.
81. Romero-Ramirez,L., Cao,H., Nelson,D., Hammond,E., Lee,A.-H., Yoshida,H., Mori,K., Glimcher,L.H., Denko,N.C. and Giaccia,A.J. (2004) XBP1 is essential for survival under hypoxic conditions and is required for tumor growth. *Cancer Res.*, **64**, 5943–5947.
82. Kar,A. and Gutierrez-Hartmann,A. (2013) Molecular mechanisms of ETS transcription factor-mediated tumorigenesis. *Crit. Rev. Biochem. Mol. Biol.*, **48**, 522–543.
83. Nerlov,C. (2007) The C/EBP family of transcription factors: a paradigm for interaction between gene expression and proliferation control. *Trends Cell Biol.*, **17**, 318–324.