J Mol Med (2012) 90:1487–1496
DOI 10.1007/s00109-012-0928-6

ORIGINAL ARTICLE

Rapid identification of novel functional promoters for gene therapy

Ian A. Pringle · Deborah R. Gill · Mary M. Connolly · Anna E. Lawton · Anne-Marie Hewitt · Graciela Nunez-Alonso · Seng H. Cheng · Ronald K. Scheule · Lee A. Davies · Stephen C. Hyde

Received: 23 March 2012 / Revised: 21 May 2012 / Accepted: 5 June 2012 / Published online: 6 July 2012

Abstract Transcriptional control of transgene expression is crucial to successful gene therapy, yet few promoter/enhancer combinations have been tested in clinical trials. We created a simple, desktop computer database and populated it with promoter sequences from publicly available sources. From this database, we rapidly identified novel CpG-free promoter sequences suitable for use in non-inflammatory, non-viral in vivo gene transfer. In a simple model of lung gene transfer, five of the six promoter elements selected, chosen without prior knowledge of their transcriptional activities, directed significant transgene expression. Each of the five novel promoters directed transgene expression for at least 14 days post-delivery, greatly exceeding the duration achieved with the commonly used CpG-rich viral enhancer/promoters. Novel promoter activity was also evaluated in a more clinically relevant model of aerosol-mediated lung gene transfer and in the liver following delivery via high-pressure tail vein injection. In each case, the novel CpG-free promoters exhibited higher and/or more sustained transgene expression than commonly used CpG-rich enhancer/promoter sequences. This study demonstrates that novel CpG-free promoters can be readily identified and that they can direct significant levels of transgene expression. Furthermore, the database search criteria can be quickly adjusted to identify other novel promoter elements for a variety of transgene expression applications.

Keywords Gene therapy · Promoter · Database · Transgene expression · CpG

Ian A. Pringle and Deborah R. Gill contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00109-012-0928-6) contains supplementary material, which is available to authorized users.

I. A. Pringle · D. R. Gill · M. M. Connolly · A. E. Lawton · A.-M. Hewitt · G. Nunez-Alonso · L. A. Davies · S. C. Hyde (✉)
GeneMedicine Research Group, NDCLS, John Radcliffe Hospital, University of Oxford,
OX3 9DU Oxford, UK
e-mail: steve.hyde@ndcls.ox.ac.uk

S. H. Cheng · R. K. Scheule
Genzyme Corporation,
Framingham, MA, USA

I. A. Pringle · D. R. Gill · M. M. Connolly · A. E. Lawton · A.-M. Hewitt · G. Nunez-Alonso · L. A. Davies · S. C. Hyde
The UK Cystic Fibrosis Gene Therapy Consortium,
Oxford, UK

Introduction

An ideal *in vivo* gene transfer vector is non-inflammatory and directs transgene expression only in the desired target tissue. Despite this, vectors employed in the majority of preclinical and clinical studies to date have utilised only a modest number of widely permissive, often viral, enhancer/promoter transcriptional control elements. Current non-viral gene transfer vectors, typically comprising a cationic polymer or liposome complexed with plasmid DNA (pDNA) containing a transgene expression cassette, frequently elicit a mild but appreciable host inflammatory reaction when administered in vivo [1, 2]. Such host inflammatory reactions appear to be initiated by innate host defence mechanisms that recognise unmethylated CG dinucleotides (CpG motifs) present within the pDNA component. Systematic mutation of all CpG motifs to alternative non-stimulating dinucleotide sequences dramatically reduces the host inflammatory reaction to both systemic and lung delivery of non-viral formulations [3]. As retention of even a single CpG within a pDNA molecule has been

shown under some circumstances to provoke a host inflammatory response, the ideal non-viral gene transfer vector arguably contains a CpG-free, tissue-specific transgene expression cassette coupled to a CpG-free pDNA backbone. One synthetic approach to the creation of the necessary CpG-free enhancer/promoter elements is mutation of commonly used CpG-rich sequences [4]. However, as CpG-motifs are often associated with transcription factor binding sites, this synthetic approach often renders enhancer/promoter elements non-functional. Consequently, only a very small number of functional CpG-free enhancer/promoter elements have been identified by this method [3, 4].

To extend the utility of CpG-free expression cassette design and offer the possibility of tissue-specific transgene expression, it would be advantageous to identify and evaluate a wide range of naturally occurring CpG-free enhancer/promoter elements. However, few, naturally CpG-free enhancer/promoter elements have been characterised and none to date, as far as we are aware, have been evaluated in the context of in vivo gene therapy. It was hypothesised that within large mammalian sequence databases currently available, promoters should exist that have regions around the transcription start site (TSS) that are CpG-free. Regrettably, none of the commonly available database mining tools we evaluated were suitable for the identification of such sequences. Therefore, a desktop database was created to allow existing enhancer/promoter sequence data to be interrogated. Using a simple search strategy, we could rapidly screen this database for the presence of not just CpGs, but any transcription factor binding site or other nucleotide sequence for which we chose to screen. This has led to the identification of many previously uncharacterised CpG-free promoters. A number of these promoters were isolated and inserted into CpG-free transgene expression vectors to allow their promoter activity to be evaluated in several models of in vivo non-viral gene transfer.

Materials and methods

Creation of a promoter database using FileMaker Pro

The PromoSer 3.0 database [5, 6], a collection of predicted promoter sequences containing 33,692 human, 22,549 mouse and 5,705 rat entries (−2,000 to +100 relative to the transcription start site), and the Eukaryotic Promoter Database (EPD) [7, 8], a collection of ~4,700 experimentally verified promoter sequences, were downloaded as text files. Simple application of find and replace functions parsed these files into a comma separated values (csv) files containing two columns of header information and promoter sequence. A FileMaker Pro (FileMaker, Inc., Santa Clara, CA, USA) database was created with two tables for sequence data and search strings. The sequence data table contained fields for the header and sequence information from the promoter databases as well as further fields to extract the gene name and species from the header data. The sequences of 4,746 transcription factor binding sites were obtained from MacVector 9.5 (MacVector Inc, Carey, NC, USA) and parsed to create a four-column csv file with columns for transcription factor name, sequence, reverse sequence and a PubMed reference. These data were imported into the search string table of the FileMaker Pro database with matching field names. The search string table was used to create two values lists for the forward and reverse orientations of the transcription factor binding sites. In the promoter table, a number of extra fields were added to allow the user to specify a search region within the promoters, and global fields were created to accommodate the value lists passed from the search string tables. Finally, multiple calculation fields were created to count the number of instances of the specified transcription factor binding sites in the specified area of the promoters. A manual entry for CpG (CG) was created to search for the presence of CpGs. This process was repeated (with some modification) to provide extra layouts to count for up to 16 different transcription factor binding sites (or other arbitrary user-defined DNA sequence) or to count the frequency of up to 100 members of a family of binding sites. Finally, the PromoSer and Eukaryotic Promoter Database csv files were imported into separate versions of the FileMaker Pro database. The finished FileMaker Pro databases and csv files containing the promoter sequences and transcription factor binding sites are available as Supplementary Information.

Construction of plasmids containing the CpG-free promoters

BAC or PAC clones containing the promoter sequences were identified in GenBank and purchased (Invitrogen, Paisley, UK), and the DNA was purified using the Large-Construct Kit (Qiagen, Crawley, UK). The promoter sequences were amplified by PCR using primers designed to add *Bgl*II and *Hin*dIII restriction sites at the 5′ and 3′ end of the PCR fragment, respectively, for five of the six promoters (human regenerating islet-derived 1β (hREG1B), human carboxypeptidase B1 (hCBOX), human tryptophan 2,3 dioxygenase (hTDOX), human apolipoprotein A-II (hAPOA2) and murine fatty acid binding protein 4 (mFABP)). For the sixth promoter (hTSHB), 5′ *Bgl*II and 3′ *Nhe*I sites were added. The promoter fragments were then inserted into the CpG-free pDNA backbone (Fig. 2a). A *CFTR* encoding version of the enhanced mFABP plasmid was created by replacing the CpG-free Lux transgene with the CpG-free form of the CFTR cDNA from pGM169 [3]. Plasmids were propagated in the *Escherichia coli* strain GT115 (Invivogen, Toulouse, France) and purified using Qiagen Endofree Mega kits (Qiagen). CpG-rich plasmids containing the luciferase (Lux) transgene under the

transcriptional control of the CMV, RSV, SV40 and UbC promoters have been described previously [9].

Procedures involving mice

Female BALB/c mice aged 6–8 weeks were used throughout the current study. Mice were housed and treated in accordance with UK Home Office ethical and welfare guidelines and fed on standard chow and water ad libitum.

Plasmid DNA delivery

For the delivery of pDNA to the lungs by instillation, pDNA was complexed with liposomes of GL67A at a charge ratio of 1:4 as described previously [10]. Mice received 100 μL of the indicated pDNA/GL67A complexes containing 80 μg pDNA by intranasal insufflation [10]. For the delivery of pDNA to the lungs by aerosol, 2 mg of pDNA was complexed with 25 kDa branched polyethylenimine (PEI) (Sigma Aldrich, Poole, UK) at an N/P ratio of 10:1 and aerosolised for ~30 min to groups of mice as described previously [11]. For the delivery of pDNA to the liver, pDNA was diluted to 100 μg/ml in D-PBS (Invitrogen), and a volume of diluted pDNA equivalent to 10 % bodyweight (approximately 1.8 to 2.0 ml) was injected into the tail vein over a period of approximately 5 s [12, 13].

Luciferase reporter assay

Mouse lung and liver samples were homogenised in 1× RLB, and luciferase activity was quantified using the Promega luciferase assay reagent system (Promega) as described previously [9]. The level of luciferase activity, expressed in arbitrary relative light units (RLU), for each sample was normalised against total protein concentration determined using a detergent-compatible protein assay (BioRad, Larne, UK). To aid in comparison with other studies, standards of recombinant luciferase protein (Promega) were also routinely analysed. For comparative purposes, 100 RLU/mg total lung protein corresponds to 2,467 ng recombinant luciferase (Promega) per mg total lung protein.

Real-time quantitative TaqMan RT-PCR

Total mRNA was purified from mouse lungs using the RNeasy Mini kit (Qiagen). Absolute levels of vector derived mRNA and murine *Cftr* mRNA were determined by TaqMan qRT-PCR as described previously [3, 14].

Lung inflammation

Levels of TNF-α in bronchoalveolar lavage fluid (BALF) were determined using Bio-Plex cytokine assays (Bio-Rad Ltd, Hemel Hempstead, UK) as described previously [15, 16].

Statistical analysis

Statistical analyses were performed using GraphPad Prism v5.0c for Mac OSX (GraphPad Software Inc, La Jolla, CA, USA). Power analysis to aid in the selection of group sizes was performed using G*Power v3.1.2 [17]. Differences in luciferase activity were determined using ANOVA on log-normalised values, using Tukey's multiple comparison post hoc tests where appropriate. Data for mRNA levels could not be successfully log-normalised and were thus analysed by non-parametric Kruskal–Wallis and Mann–Whitney *U* (MW) tests where appropriate. Differences in dinucleotide frequencies were determined using Friedman's test, using Dunn's multiple comparison post hoc test where appropriate. A probability value (*p* value) of less than 0.05 was considered statistically significantly. In figures, bars represent mean± standard error of the mean; the symbols NS, *, ** and *** indicate a *p* value of >0.05, <0.05, <0.01 and <0.001, respectively, compared with the unlabelled comparator group.

Results

Identification of CpG-free promoters

Initially the size of the available pool of CpG-free promoters was determined. A FileMaker Pro database (Supplementary Information) was created to house the PromoSer 3.0 dataset of predicted promoter sequences containing 33,692 human, 22,549 mouse and 5,705 rat entries (−2,000 to +100 bp relative to the TSS) [6]. This was first used to calculate the frequency of all dinucleotides in the region −2,000 to the TSS (Fig. 1a). Most dinucleotides exhibited a normal distribution with peaks around 100–150 dinucleotides per 2,000 bp. However, as previously described for mammalian genomes in general [18], the frequency of CG dinucleotides (CpGs) in the promoter regions represented in the database was markedly skewed, with a peak at ~15 CpGs per 2,000 bp (Fig. 1a). Notably, ~20 % of promoters in this large dataset contained ≤15 CpGs in the selected region. To establish if a *usable* population of CpG-free promoters was available for use in gene therapy models, the distance from the TSS to the first occurrence of each of dinucleotide was calculated for each promoter (Fig. 1b). In the first 150 bp, 86 % of the promoter sequences contained a CpG, and there was no apparent difference between the occurrence of CpGs and the other dinucleotides ($p>0.99$; Friedman). However, as the distance from the TSS increased, a bias in favour of CpG-free promoter sequences appeared, such that there were 1,339 promoters containing ≥500 bp of CpG-free sequence 5′ of the TSS, while only 158 promoters, at most, remained free of any other dinucleotide. This pattern is readily apparent in 50-bp windows between 300 and

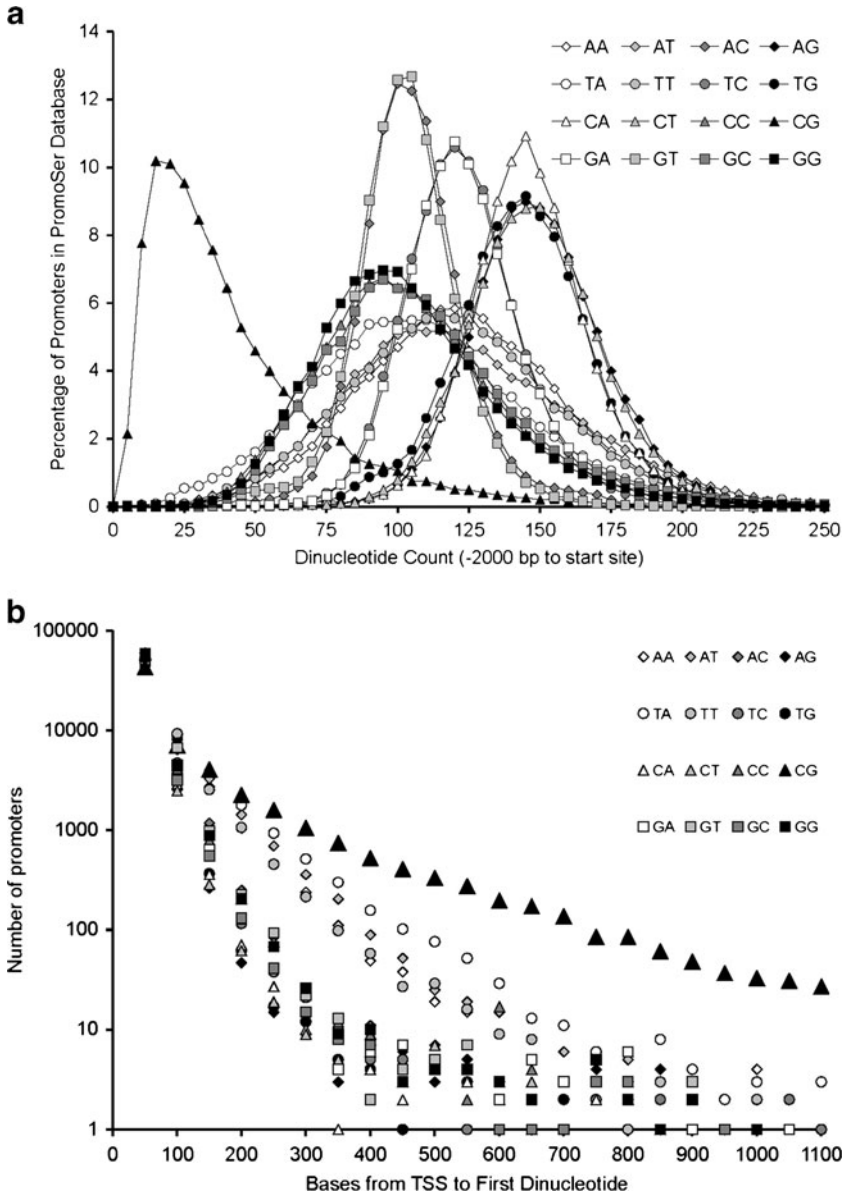

Fig. 1 Promoter dinucleotide frequencies. **a** Percentage distribution of all 16 DNA dinucleotides in the PromoSer database (−2,000 bp to the TSS). The number of each of the 16 dinucleotides in the 61,946 promoters in the PromoSer 3.0 dataset was determined and expressed as a percentage count (**b**). Occurrence rate of each dinucleotide in PromoSer database. The first occurrence of each dinucleotide relative to the TSS was collated into 50-bp bins for representation. CG occurrence rate (*black triangles*), along with rates for all other dinucleotides, is shown. See also Supplementary Fig. 1

1,000 bp from the TSS where 4–11-fold more CpG-free promoters were found than for any other dinucleotide ($p<$ 0.001; Friedman/Dunn's multiple comparison post hoc) (Fig. 1b and Supplementary Fig. 1).

From these observations, it can be concluded that CpG-free promoters are readily identifiable within mammalian genomes and that they represent an under-utilised experimental resource. This conclusion was strengthened when a second FileMaker Pro database was created to interrogate an alternative dataset, the EPD, which contains experimentally determined promoter sequences (1,869 humans, 195 mice, 117 rats and 2,911 others). Unique sequences within the EPD (−2,000 to +100 bp relative to the transcription start site) were screened for the presence of CpGs (in the region −500 to +10 bp), and 18 promoters were identified as being CpG-free (Table 1). A further 55 promoters contained only one CpG and a further 60 contained two. The 18 CpG-free promoters identified from the two databases were further analysed for their suitability, and five human promoters and one mouse promoter were selected (based on suitability of restriction sites to allow rapid cloning) for further study (Table 2). Five of the six promoters (hREG1B, hCBOX, hTDOX, hAPOA2 and mFABP) have >500-bp CpG-free DNA upstream of the transcription start, but the sequence between the TSS and the associated coding sequence initiation codon (ATG) was not CpG-free. For these promoters, the upstream CpG-free region and 10 bases downstream of the TSS was isolated. These promoter sequences were inserted into a common CpG-free pDNA backbone [3] containing a CpG-free CMV enhancer, a 5′ UTR region comprising two CpG-free exons and a synthetic CpG-free intron and a CpG-free luciferase reporter gene (Fig. 2a). The

Table 1 CpG-free promoters identified from initial screen of database

Species	Gene	EPD accession #
Bovine	Alpha s1-casein, CSN1S1 gene	EP11090
Bovine	Thyroid-stimulating hormone alpha	EP11126
Chicken	Myosin adult fast-white (myosin heavy chain) 1	EP11080
Human	Apolipoprotein A-II	EP11088
Human	Thyroid stimulating hormone-β	EP30071
Human	Regenerating islet-derived 1β	EP73427
Human	Guanylate binding protein 2, interferon-inducible	EP73490
Human	Carboxypeptidase B1	EP73648
Human	KIAA0174 gene product	EP73814
Human	Tryptophan 2,3 dioxygenase	EP74093
Murine	Fatty acid binding protein 4	EP24033
Murine	Insulin I	EP17068
Murine	Interferon β	EP23039
Murine	Immunoglobulin variable region 1	EP07116
Murine	Immunoglobulin (heavy chain) variable region: specific for 4-hydroxyl-3-nitrophenacetyl	EP14073
Murine	Immunoglobulin (heavy chain) variable region: specific for 4-hydroxyl-3-nitrophenacetyl	EP14071
Rat	Prostatic steroid binding protein C2, C2 gene	EP16047
Rat	Insulin II	EP17070

remaining promoter, hTSHB, had an entirely CpG-free sequence spanning from −750 bp relative to the TSS to the initiation codon at +350. This sequence included a region predicted to be the naturally occurring first intron of the hTSHB gene [19, 20]. Consequently, this entire region was used and the hTSHB construct lacked the synthetic 5′ UTR region included in the other promoter constructs.

CpG-free promoter activity following in vivo lung gene transfer

Plasmid DNA for each of the six selected promoters was complexed with GL67A liposomes (a leading non-viral gene transfer agent [21]) and transcriptional activity in vivo was assessed in a mouse lung gene transfer model [3]. Lux activity in the lungs of mice dosed by instillation with the novel plasmids was determined at 1 and 14 days post-dosing and compared with constructs containing the commonly used CpG-rich CMV, RSV, SV40, UbC enhancer/promoters and CpG-free hCEFI enhancer/promoter all previously found to be effective in this system [3, 13]; untreated naïve mice acted as negative controls (Fig. 2b, c). The commonly used CMV, RSV, SV40, UbC and hCEFI enhancer promoters all directed substantial levels of Lux activity compared with naïve controls ($p<0.001$; ANOVA/Tukey's post hoc (A/T)). Mean Lux activity directed by the novel mFABP promoter was similar to that observed with the commonly used CMV, SV40 and hCEFI enhancer/promoters ($p>0.05$; A/T). Encouragingly, four of the other five novel CpG-free promoters were transcriptionally active, directing Lux activity at levels significantly greater than observed in naïve animals ($p>0.05$ for hAPOA2, $p<0.001$ for all other promoters; A/T), though at levels that were lower than those observed with the commonly used CMV, SV40 and hCEFI enhancer/promoters ($p<0.05$; A/D).

By day 14, Lux activity from the CpG-rich CMV, RSV and SV40 viral promoter had, as expected [9], fallen to the level of untreated naïve animals ($p>0.05$; A/T). In contrast, 5/6 novel CpG-free promoters and the commonly used UbC and hCEFI enhancer/promoters maintained significant Lux activity (hAPOA2 $p>0.05$, others $p<0.001$; A/T). Encouragingly, Lux activity directed by the mFABP promoter was indistinguishable from that directed by the UbC and hCEFI enhancer/promoters ($p>0.05$; A/T). Interestingly, for the novel CpG-free promoters, the rank order of transgene expression level observed at day 1 was maintained at day 14, with the mFABP directing the highest level and hAPOA2 directing the lowest level of expression at both time points. None of the five novel CpG-free promoters that directed significant transgene expression appeared to suffer from the striking attenuation of promoter activity (>4-logs) observed with the CpG-rich viral enhancer/promoters in the lung (Fig. 2b, c). Importantly, a range of transcriptional activities was observed, allowing for the selection of desired expression level by judicious choice of promoter sequence.

Subsequently, the inflammatory response generated by each of the six novel CpG-free promoters was compared with the CpG-rich CMV, RSV and SV40 enhancer/promoters and the CpG-free hCEFI enhancer/promoter. The pDNAs were complexed with GL67A, instilled into the mouse lung, and at 1 day post-

Table 2 CpG-free promoters identified and used in this study

Gene	CpG-free region	Cloned region	BAC/PAC ID #
Human thyroid stimulating hormone-β	−750:+350 (ATG)	−750:+350	RP4-666F24
Human regenerating islet-derived 1β	−625:+70	−625:+10	RP11-88C6
Human carboxypeptidase B1	−630:+16	−630:+10	RP11-505J9
Human tryptophan 2,3 dioxygenase	−670:+36	−670:+10	RP11-401G24
Human apolipoprotein A-II	−513:+26	−513:+10	RP11-401G24
Murine fatty acid binding protein 4	−623:+52	−623:+10	RP23-436F15

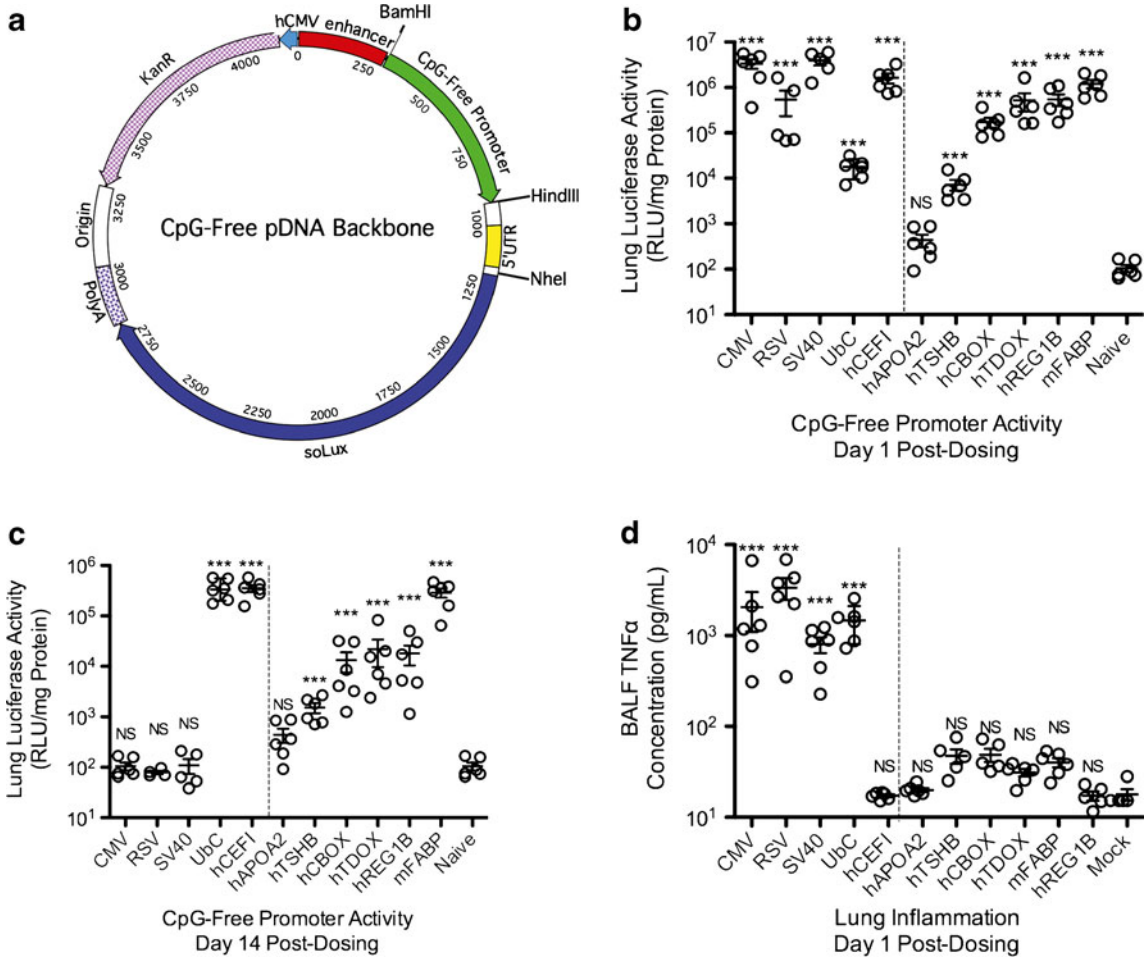

Fig. 2 CpG-free promoter activity following pDNA/GL67A instillation to the mouse lung. **a** Structure of pDNA expression vectors used in the study. The selected promoters were inserted into a common plasmid backbone, entirely free of CpGs, containing a human CMV enhancer, 5′ UTR region, soLux luciferase reporter gene, BGH poly A 3′ UTR, R6K origin of replication and a kanamycin resistance gene (Kan^R). Restriction sites (BamHI, HindIII and NheI) used in construction of promoter variants are indicated. *Numbers* indicate size of elements in DNA base pairs. pDNA was complexed with liposomes of GL67A and delivered to the mouse lung by instillation. Luciferase activity in the mouse lung, normalised to total protein concentration, is shown at day 1 (**b**) and day 14 (**c**) post-dosing. Luciferase reporter gene expression was under the control of the promoters indicated. Luciferase activity in naïve animals acts as negative controls. TNF-α levels in bronchoalveolar lavage fluid collected at day 1 post-dosing is shown (**d**). Mock-treated animals act as negative controls. n=5–6 mice per group. *Vertical dashed lines* delineate commonly used enhancer/promoters from the novel CpG-free promoters

dosing, a BAL was performed and levels of TNF-α in the BALF were assessed. In line with previous studies utilising CpG-free pDNAs [3], we observed that in vivo delivery of all seven of the CpG-free promoters resulted in low levels of TNF-α similar to that of naïve mice (Fig. 2d). In contrast, delivery of the CpG-rich CMV, RSV, SV40 and UbC promoters resulted in a highly significant induction of TNF-α levels ranging between 44- to 187-fold (p<0.001; A/T) (Fig. 2d).

CpG-free promoter activity following in vivo liver gene transfer

To ensure the observations made above could be applied generally, the transcriptional activity of a subset of the selected promoters was examined in an alternative model of in vivo gene transfer. Transgene expression from the four most potent promoters in the lung (hCBOX, hTDOX, hREG1B and mFABP) was measured in the liver following naked plasmid DNA delivery via the high-pressure tail vein method [12, 13]. Transcriptional activity was compared with constructs containing the commonly used CpG-rich CMV and UbC enhancer/promoters and CpG-free hCEFI enhancer/promoters; all previously found to be effective in this system [22]. At day 1 post-dosing, the hCEFI enhancer/promoter directed the highest levels of Lux activity of the positive controls (p<0.001; A/D). Encouragingly, all four of the novel CpG-free promoters directed higher activity than the hCEFI enhancer/promoter (p<0.001 for mFABP, p< 0.01 for others; A/D) (Fig. 3).

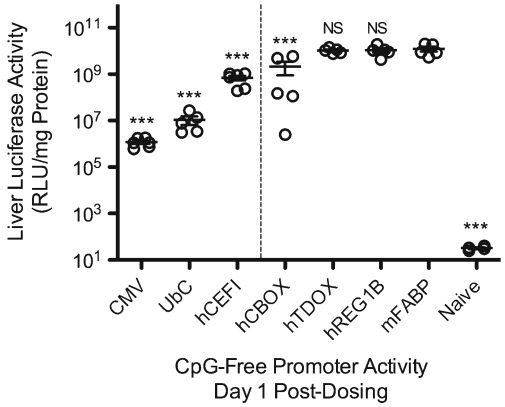

Fig. 3 CpG-free promoter activity following high-pressure tail vein delivery of naked plasmid DNA to the mouse liver. Plasmid DNA was diluted to 0.1 mg/ml in D-PBS and a volume equal to 10 % bodyweight was injected over 5 s to the tail veins of mice. Luciferase activity in the mouse liver, normalised to total protein concentration, is shown at day 1. n=6 mice per group. *Vertical dashed line* delineates commonly used enhancer/promoters from the novel CpG-free promoters

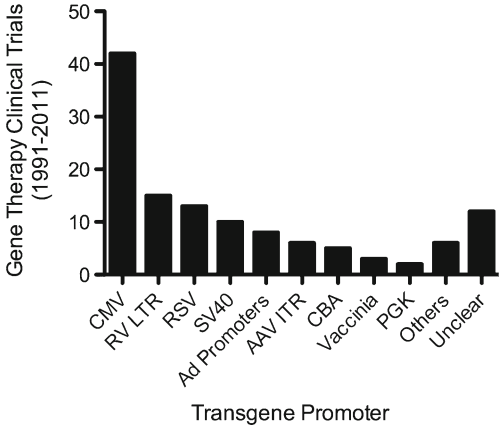

Fig. 5 Promoters used in gene therapy clinical trials (1991-2011). One hundred twenty-two gene therapy clinical trial publications were identified from the Gene Therapy Clinical Trials Database (http://www.wiley.com/legacy/wileychi/genmed/clinical/). Where possible, the promoters used in each gene therapy vector were identified. For a full list of publications by promoter, see Supplementary Fig. 2

Duration of CFTR mRNA expression directed by mFABP promoter following aerosol-mediated lung gene transfer

The results for the in vivo lung gene transfer studies performed above were obtained using an instillation procedure that rapidly delivers a bolus of fluid to the lungs. Subsequently we evaluated the transcriptional activity of the mFABP promoter, the most potent of the novel CpG-free promoters described here in a more clinically relevant model of lung gene transfer, where pDNA was complexed with 25 kDa PEI and delivered by inhalation of an aerosol produced by a jet nebuliser [11]. Transcriptional activity of the mFABP promoter, directing the expression of a CpG-free form of the cystic fibrosis transmembrane conductance regulator (*CFTR*) cDNA, was compared with that achieved by pCF1-CFTR, a construct containing the CpG-rich CMV enhancer/promoter and pGM169 a construct containing the CpG-free hCEFI enhancer/promoter, which between them have been used in four previous non-viral clinical trials involving patients with the common inherited lung disease cystic fibrosis [1, 2, 23, 24].

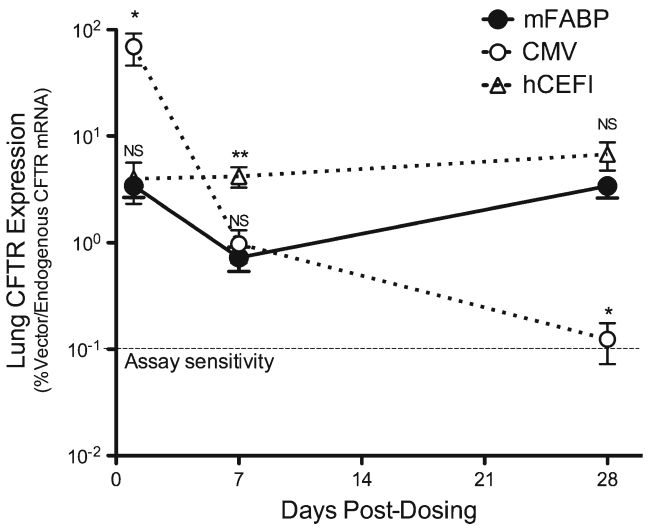

Fig. 4 mFABP promoter activity following pDNA/PEI aerosolisation to the mouse lung. mFABP-CFTR or pCF1-CFTR pDNA (2 mg) was complexed with 25 kDa branched PEI at an N/P ratio of 10:1 in a final volume of 10 mL. pDNA/PEI complexes were aerosolised into an 8-L perspex chamber over ~30 min containing un-restrained groups of mice. Absolute levels of pDNA vector-derived lung mRNA were determined at 1, 7 and 28 days post-dosing by TaqMan qRT-PCR and normalised to absolute levels of murine *Cftr* (100 %=1 copy vector mRNA: 1 copy murine *Cftr* mRNA). n=6 mice per group. *Horizontal dashed line* indicates assay sensitivity

Lungs were harvested at 1, 7 and 28 days post-aerosol dosing, and absolute levels of vector-derived *CFTR* mRNA were determined by qRT-PCR and normalised to endogenous murine *Cftr* mRNA [14]. Robust transcriptional activity from the mFABP promoter was observed at day 1 post-aerosol and remained essentially unchanged throughout the study (Kruskal–Wallis test, $p>0.05$) (Fig. 4). At day 1 post-aerosol, activity from the CMV promoter was ~20-fold higher than from the mFABP promoter ($p=0.002$; MW) but fell during the study such that it was at a similar level to that of mFABP by day 7 ($p=0.329$; MW) and was 15-fold lower than mFABP by day 28 ($p=0.009$; MW) (Fig. 4). Activity from the hCEFI promoter was similar to that of the mFABP promoter at days 1 and 28 post-aerosol ($p>0.05$; MW).

Discussion

Promoter selection is undoubtedly a key factor in the efficacy of gene transfer vectors. With the availability of many

complete mammalian genomes, there exists an almost unlimited choice of promoters available. Despite this, very few promoters have been assessed in gene therapy models, and the number of different promoters used in clinical trials appears to be small. Using a number of public sources, we were able to identify 122 publications from gene therapy clinical trials performed between 1991 and 2011. In these trials, we could identify the use of only 15 different promoters. Variations of the CMV enhancer/promoter were the most popular (34 %) with the retroviral LTR (12 %), RSV LTR (11 %) and the SV40 promoter (8 %) being the next most common (Fig. 5 and Supplementary Fig. 2). Of note is that a significant number of clinical trial publications (10 %) failed to mention which promoter was used and subsequent searches of cited literature could not identify the element selected. While this probably represents only a subset of all studies, it is clear that few promoters have been used in clinical trials for gene therapy and the choice of promoter has most likely not been based on any rigorous assessment of what was available. In the case of non-viral gene therapy for CF, persistent gene expression is perceived as an absolute requirement for clinical progress in order to maximise the interval between treatments. Most clinical trials for CF, however, have used promoters of viral origin such as CMV [1, 2, 25, 26] or RSV [27, 28]. Most of these studies reported evidence of CFTR expression, but levels were rarely maintained for more than a few days. Similar transient activity with viral promoters has been observed in mouse models of lung gene transfer [9, 29]. Subsequently, we have shown that some human promoters (such as UbC), or compound elements comprising a viral enhancer and a human promoter (such as hCEFI), can direct long-term gene expression in the mouse lung [3, 9] and in human clinical trials [24]. Thus, the concept that selecting an appropriate promoter can have an appreciable impact on gene expression profile following in vivo gene transfer has been established.

By using the database screening approach described here, we identified 18 naturally occurring CpG-free promoters (Table 1) and chose six (Table 2) for evaluation in mouse models of in vivo gene transfer. Five of these six were shown to direct transcriptional activity 24 h following delivery to the lung via instillation, with levels of transgene expression directed by the mFABP promoter approaching that achieved with the commonly used CMV enhancer/promoter (Fig. 2b). Gene transfer by this route using conventional CpG-rich pDNA constructs is associated with an acute inflammatory reaction. However, levels of pro-inflammatory TNF-α provoked by delivery of the six novel CpG-free promoters were essentially at the naive background, 43–118-fold lower than those provoked by a construct containing the CpG-rich CMV enhancer/promoter (Fig. 2d). The majority of promoters historically evaluated in this model appear to suffer from pronounced transcriptional silencing [3, 9, 10, 29], which was readily apparent for the commonly used CMV, RSV and SV40 enhancer/promoters (Fig. 2b, c). Interestingly, five of the six new CpG-free promoters evaluated—hTSHB, hCBOX, hTDOX, hREG1B and mFABP—retained significant transcriptional activity for at least 14 days after delivery (Fig. 2c).

To further assess their effectiveness, the activity of the four most transcriptionally potent promoters in the lung studies was evaluated in the mouse liver following high-pressure tail vein delivery. All were shown to direct transcriptional activity in the liver 24 h after delivery, with levels of transgene expression exceeding those achieved with the commonly used CMV, UbC and hCEFI enhancer/promoters (Fig. 3). Finally, mFABP, the most consistent of the novel CpG-free promoters, was evaluated in a clinically relevant aerosol delivery model utilising *CFTR* as a potential therapeutic transgene. Transcriptional activity from an mFABP-CFTR construct eclipsed that directed by pCF1-CFTR, a CMV promoter construct used in three previous clinical trials, and was similar to that achieved with an hCEFI promoter containing construct used in a recent clinical study.

Together, these results demonstrate that not only do naturally occurring CpG-free promoters exist but they can be effective in models of gene therapy in which many commonly used promoters are ineffective. Moreover, these promoters are novel to gene therapy and were selected merely by their CpG content. This suggests that for situations where CpG-content is not a concern, promoter choice could be limitless and a search based on suitable criteria is likely to identify candidate promoters that may be useful for preclinical and clinical application. Our specific interest was to screen for CpG-free promoters, but the approach can be used to screen for the presence of any sequence or combination of sequences. For example, if novel promoters were required for use in a model of inflammatory disease, it may be beneficial to specify promoters with or without binding sites for specific transcription factors; to this end, the binding sites for NF-kB have been included in gene transfer vectors to increase nuclear uptake or increase expression [30, 31]. Forty-four promoters containing at least four NF-kB binding sites were identified in our databases in a matter of minutes, without the need for in-depth knowledge of bioinformatics or programming tools (data not shown). Another approach that has been used with a degree of success is the design of synthetic promoters with specific tissue expression [32–34]. In a recent study, Santilli et al. created synthetic promoter for X-GCD gene therapy based around the presence of specific transcription factor binding sites [34]. In the PromoSer database, we were able to rapidly identify eight natural promoters that contain between three and five of these sites (data not shown).

In conclusion, we have developed a method to identify promoter elements, which replaces traditional bioinformatics tools with a simple desktop database that can be

easily mastered. The database provides an unbiased analysis of promoter sequences and led to the identification of novel CpG-free promoters that were functionally equivalent, or better than commonly used promoters. By adapting the promoter search criteria, this approach could be used in other applications requiring cell or tissue-specific transgene expression, such as development of transgenic animal models or recombinant protein production.

Acknowledgments The work was supported by a grant from the Cystic Fibrosis Trust to the UK Cystic Fibrosis Gene Therapy Consortium.

Conflict of interest The authors declare no conflicts of interest.

References

1. Alton EW, Stern M, Farley R, Jaffe A, Chadwick SL, Phillips J, Davies J, Smith SN, Browning J, Davies MG et al (1999) Cationic lipid-mediated CFTR gene transfer to the lungs and nose of patients with cystic fibrosis: a double-blind placebo-controlled trial. Lancet 353:947–954
2. Ruiz FE, Clancy JP, Perricone MA, Bebok Z, Hong JS, Cheng SH, Meeker DP, Young KR, Schoumacher RA, Weatherly MR et al (2001) A clinical inflammatory syndrome attributable to aerosolized lipid-DNA administration in cystic fibrosis. Hum Gene Ther 12:751–761
3. Hyde SC, Pringle IA, Abdullah S, Lawton AE, Davies LA, Varathalingam A, Nunez-Alonso G, Green AM, Bazzani RP, Sumner-Jones SG et al (2008) CpG-free plasmids confer reduced inflammation and sustained pulmonary gene expression. Nat Biotechnol 26:549–551
4. Yew NS, Zhao H, Przybylska M, Wu IH, Tousignant JD, Scheule RK, Cheng SH (2002) CpG-depleted plasmid DNA vectors with enhanced safety and long-term gene expression in vivo. Mol Ther 5:731–738
5. Halees AS, Leyfer D, Weng Z (2003) PromoSer: a large-scale mammalian promoter and transcription start site identification service. Nucleic Acids Res 31:3554–3559
6. Halees AS, Weng Z (2004) PromoSer: improvements to the algorithm, visualization and accessibility. Nucleic Acids Res 32:W191–W194
7. Cavin Perier R, Junier T, Bucher P (1998) The eukaryotic promoter database EPD. Nucleic Acids Res 26:353–357
8. Schmid CD, Perier R, Praz V, Bucher P (2006) EPD in its twentieth year: towards complete promoter coverage of selected model organisms. Nucleic Acids Res 34:D82–D85
9. Gill DR, Smyth SE, Goddard CA, Pringle IA, Higgins CF, Colledge WH, Hyde SC (2001) Increased persistence of lung gene expression using plasmids containing the ubiquitin C or elongation factor 1 alpha promoter. Gene Ther 8:1539–1546
10. Pringle IA, Raman S, Sharp WW, Cheng SH, Hyde SC, Gill DR (2005) Detection of plasmid DNA vectors following gene transfer to the murine airways. Gene Ther 12:1206–1214
11. Davies LA, McLachlan G, Sumner-Jones SG, Ferguson D, Baker A, Tennant P, Gordon C, Vrettou C, Baker E, Zhu J et al (2008) Enhanced lung gene expression after aerosol delivery of concentrated pDNA/PEI complexes. Mol Ther 16:1283–1290
12. Liu F, Song Y, Liu D (1999) Hydrodynamics-based transfection in animals by systemic administration of plasmid DNA. Gene Ther 6:1258–1266
13. Song YK, Liu F, Zhang G, Liu D (2002) Hydrodynamics-based transfection: simple and efficient method for introducing and expressing transgenes in animals by intravenous injection of DNA. Methods Enzymol 346:92–105
14. Rose AC, Goddard CA, Colledge WH, Cheng SH, Gill DR, Hyde SC (2002) Optimisation of real-time quantitative RT-PCR for the evaluation of non-viral mediated gene transfer to the airways. Gene Ther 9:1312–1320
15. Bazzani RP, Cai Y, Hebel HL, Hyde SC, Gill DR (2011) The significance of plasmid DNA preparations contaminated with bacterial genomic DNA on inflammatory responses following delivery of lipoplexes to the murine lung. Biomaterials 32:9854–9865
16. Davies LA, Hyde SC, Nunez-Alonso G, Bazzani RP, Harding-Smith R, Pringle IA, Lawton AE, Abdullah S, Roberts TC, McCormick D et al (2012) The use of CpG-free plasmids to mediate persistent gene expression following repeated aerosol delivery of pDNA/PEI complexes. Biomaterials 33:5618–5627
17. Faul F, Erdfelder E, Lang AG, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39:175–191
18. Saxonov S, Berg P, Brutlag DL (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci USA 103:1412–1417
19. Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S (1996) Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. Nucleic Acids Res 24:3439–3452
20. Brunak S, Engelbrecht J, Knudsen S (1991) Prediction of human mRNA donor and acceptor sites from the DNA sequence. J Mol Biol 220:49–65
21. McLachlan G, Davidson H, Holder E, Davies LA, Pringle IA, Sumner-Jones SG, Baker A, Tennant P, Gordon C, Vrettou C et al (2011) Pre-clinical evaluation of three non-viral gene transfer agents for cystic fibrosis after aerosol delivery to the ovine lung. Gene Ther 18:996–1005
22. Riu E, Chen ZY, Xu H, He CY, Kay MA (2007) Histone modifications are associated with the persistence or silencing of vector-mediated transgene expression in vivo. Mol Ther 15:1348–1355
23. Zabner J, Cheng SH, Meeker D, Launspach J, Balfour R, Perricone MA, Morris JE, Marshall J, Fasbender A, Smith AE et al (1997) Comparison of DNA–lipid complexes and DNA alone for gene transfer to cystic fibrosis airway epithelia in vivo. J Clin Invest 100:1529–1537
24. Davies JC, Davies G, Gill DR, Hyde SC, Boyd AC, Innes JA, Porteous DJ, Cheng SH, Scheule RK, Higgins TH, Griesenbach U, Alton EW (2011) Safety and expression of a single dose of lipid-mediated CFTR gene therapy to the upper and lower airways of patients with CF. Pediatric Pulmonol 46(S34):281–281
25. Porteous DJ, Dorin JR, McLachlan G, Davidson-Smith H, Davidson H, Stevenson BJ, Carothers AD, Wallace WA, Moralee S, Hoenes C et al (1997) Evidence for safety and efficacy of DOTAP cationic liposome mediated CFTR gene transfer to the nasal epithelium of patients with cystic fibrosis. Gene Ther 4:210–218
26. Konstan MW, Davis PB, Wagener JS, Hilliard KA, Stern RC, Milgram LJ, Kowalczyk TH, Hyatt SL, Fink TL, Gedeon CR et al (2004) Compacted DNA nanoparticles administered to the nasal mucosa of cystic fibrosis subjects are safe and demonstrate partial to complete cystic fibrosis transmembrane regulator reconstitution. Hum Gene Ther 15:1255–1269
27. Gill DR, Southern KW, Mofford KA, Seddon T, Huang L, Sorgi F, Thomson A, MacVinish LJ, Ratcliff R, Bilton D et al (1997) A placebo-controlled study of liposome-mediated gene transfer to the nasal epithelium of patients with cystic fibrosis. Gene Ther 4:199–209

28. Hyde SC, Southern KW, Gileadi U, Fitzjohn EM, Mofford KA, Waddell BE, Gooi HC, Goddard CA, Hannavy K, Smyth SE et al (2000) Repeat administration of DNA/liposomes to the nasal epithelium of patients with cystic fibrosis. Gene Ther 7:1156–1165
29. Yew NS, Wysokenski DM, Wang KX, Ziegler RJ, Marshall J, McNeilly D, Cherry M, Osburn W, Cheng SH (1997) Optimization of plasmid vectors for high-level expression in lung epithelial cells. Hum Gene Ther 8:575–584
30. Kuramoto T, Nishikawa M, Thanaketpaisarn O, Okabe T, Yamashita F, Hashida M (2006) Use of lipoplex-induced nuclear factor-kappaB activation to enhance transgene expression by lipoplex in mouse lung. J Gene Med 8:53–62
31. Breuzard G, Tertil M, Goncalves C, Cheradame H, Geguan P, Pichon C, Midoux P (2008) Nuclear delivery of NFkappaB-assisted DNA/polymer complexes: plasmid DNA quantitation by confocal laser scanning microscopy and evidence of nuclear polyplexes by FRET imaging. Nucleic Acids Res 36:e71
32. Worthington J, Robson T, Scott S, Hirst D (2005) Evaluation of a synthetic CArG promoter for nitric oxide synthase gene therapy of cancer. Gene Ther 12:1417–1423
33. Szymanski P, Anwer K, Sullivan SM (2006) Development and characterization of a synthetic promoter for selective expression in proliferating endothelial cells. J Gene Med 8:514–523
34. Santilli G, Almarza E, Brendel C, Choi U, Beilin C, Blundell MP, Haria S, Parsley KL, Kinnon C, Malech HL et al (2010) Biochemical correction of X-CGD by a novel chimeric promoter regulating high levels of transgene expression in myeloid cells. Mol Ther 19:122–132