

# **An Investigation into Persistent Transgene Expression from Plasmid Vectors in the Lung**

Thesis submitted in partial fulfilment of the requirements for the degree  
of Doctor of Philosophy



**Catarina A. Oliveira**

Gene Medicine Research Group

Nuffield Division of Clinical Laboratory Sciences

and

Linacre College

University of Oxford

Hilary Term, 2015

---

## **An Investigation into Persistent Transgene Expression from Plasmid Vectors in the Lung**

**Catarina A. Oliveira**

Nuffield Division of Clinical Laboratory Sciences and Linacre College,  
University of Oxford

Thesis submitted for the degree of Doctor of Philosophy, Hilary Term, 2015

### **Abstract**

Gene therapy for chronic lung diseases will require vectors capable of persistent transgene expression in the absence of inflammation but few promoters have been identified that can fulfil both of these requirements. The development of plasmids devoid of CG dinucleotides (CpGs) has resulted in markedly reduced levels of pro-inflammatory cytokines. Furthermore, a novel synthetic enhancer/promoter, termed hCEFI, has been shown to direct high levels of sustained transgene expression in the murine lung, following aerosol delivery of non-viral formulations. A CpG-free plasmid containing the hCEFI enhancer/promoter has been evaluated in clinical trials for cystic fibrosis lung gene therapy. The aim of this thesis was to investigate the reasons for successful persistent expression achieved with this plasmid. An understanding of these reasons is also useful for the development of new gene therapy treatments. Cell culture models did not recapitulate the transgene expression profiles obtained *in vivo*; therefore, a mouse aerosol model of lung gene transfer was utilised. New variants of the hC enhancer present in the hCEFI promoter were constructed and tested *in vivo*. Studies demonstrated that a minimal enhancer sequence, comprising the final 93 bp of the full-length 302 bp hC enhancer, was necessary and sufficient for mediating persistent and high levels of transgene activity in the mouse lung. Support for the inclusion of multiple copies of the minimal enhancer to increase transgene activity was also obtained. Observations in a mouse model for gene transfer to the liver showed that the minimal enhancer could not mediate transgene activity at the same level as the full-length hC enhancer, indicating that the minimal enhancer does not uniformly direct transgene activity but may be specific to the target organ. An analysis of predicted transcription factor binding sites was performed to support such tissue-specific differences. Overall, these studies demonstrate that only a very small portion of an enhancer is required to direct high-level activity in the lung. This has implications for the future design of both non-viral and viral vectors for use in gene therapy applications.

## Acknowledgements

I would like to begin by thanking my supervisors Prof. Deborah Gill and Prof. Steve Hyde for giving me the chance to study for this D.Phil. and for their time, support and patience during this project. The members of the Gene Medicine Research Group have made my time here enjoyable and their help, encouragement and friendship has not gone unnoticed. I'd first like to thank Dr. Ian Pringle for his time answering my many questions, putting up with my whistling and humming, and for his support in the lab and at BMS, especially with performing the hydrodynamic tail vein injections. I'd like to thank Dr. Stephanie Sumner-Jones and Dr. Lee Davies for their camaraderie, understanding and assistance when I needed it. In particular, thank you to Stephanie for your time in helping me with TaqMan, and to Lee for your time in talking about all things related to airway gene delivery. A big thank you to Ian and Lee for reading through parts of my thesis and for your feedback- very much appreciated! I'd also like to acknowledge Dr. Reto Bazzani. Even though our time together in the lab was short, you have become a good friend and I'm glad to have you in my life- thank you for the long-distance support and encouragement. To my fellow students, JF Gélinas, Natasha Davie, Jack Tan and Agata Antepowicz: thanks for the laughs, the hugs, the conversations and the nights out! Thanks also to Charlotte Rush and Karen Bamford for their support and help over the years, and to all the other members of Gene Medicine Group whose path I crossed during my time here, particularly Anna Lawton and Mary Connolly.

To my parents, Lena and Frank, and sisters, Andrea and Rosana: thank you for the support and love you've provided not only for the past 4 years but also throughout my life. To my children, Eva and Edward: thank you for your super cuddles, excellent kisses and fun laughs, which always brighten up my day. To my husband and best friend, Mark: thank you for your love and support. Your help, encouragement, and patience have been instrumental in allowing me to achieve this goal.

Big thanks to my friends who were always up for a chat and/or helping with the kids over the years: Dr. Lisa Cooper, Dr. Alison Crook, Conchita Sancho, Niamh Tuite, Tessa Finlayson-Gray, Dr. Judy Kim and Dr. Fran Fieldhouse.

My D.Phil. was funded by the Medical Research Council. I would also like to thank the NDCLS Graduate Studies Committee for their further support.

Finally, to my cousin Joseph Cecilio, the reason for doing this project: you inspire me. We all love you very much.

## Declaration

This thesis has been composed by myself and has not been used in any previous application for a degree. The results presented herein were obtained by myself, except where acknowledged in the text. All sources of published or unpublished information have been specifically acknowledged by means of a reference or personal communication notation, respectively.

---

## List of Abbreviations

AAT	alpha-1 antitrypsin
AAV	adeno-associated virus
Ad	adenovirus
ALI	air-liquid interface
ANOVA	analysis of variance
ASL	airway surface liquid
ASPA	Animal (Scientific Procedures) Act 1986
BGH	bovine growth hormone
BLI	bioluminescence imaging
BMS	Biomedical Services
bp	base pair(s)
CAG	CMV enhancer/chicken $\beta$ actin promoter
CAR	coxsackie-adenovirus receptor
cDNA	complementary DNA
CF	cystic fibrosis
CFTR/Cftr	cystic fibrosis transmembrane conductance regulator
ChIP	chromatin immunoprecipitation
chrDNA	chromosomal DNA
CpG	cytosine-guanosine dinucleotide
CTL	cytotoxic T lymphocyte
DC-Chol:DOPE	3 $\beta$ -[N-(N',N'- dimethylaminoethane)carbonyl]cholesterol: dioleoylphosphatidyl- ethanolamine
DMPE-PEG <sub>5000</sub>	1,2-dimyristoyl-sn-glycero-3-phosphoethanolamine-N-[methoxy (polyethyleneglycol) 5000]
DOPE	1,2-dioleoyl-sn-glycero-3-phosphoethanolamine
EboZ	Ebola virus, Zaire strain
EF1 $\alpha$	elongation factor 1 $\alpha$

---

EF	synthetic, codon-optimised CpG-free version of the EF1 $\alpha$ promoter
EFI	EF promoter and intron
EGFP	enhanced green fluorescent protein
EGTA	ethylene glycol tetraacetic acid
EHD	electrohydrodynamic
F/HN	fusion and hemagglutinin neuraminidase envelope proteins
FBS	fetal bovine serum
FIV	feline immunodeficiency virus
GL67	Genzyme lipid 67; Cholest-5-en-3-ol(3 $\beta$ )-3-[(3-aminopropyl)[4-[(3-aminopropyl)amino]butyl] carbamate]
GL67A	Aerosol formulation of GL67; 1:2:0.05 molar ratio of GL67:DOPE:DMPE-PEG <sub>5000</sub>
GFP	green fluorescent protein
GP	glycoprotein
GTA	gene transfer agent
GTF	general transcription factor
HAT	histone acetyltransferase
HBV	hepatitis B virus
hC	synthetic, codon-optimised CpG-free version of the hCMV enhancer
hCEFI	synthetic, codon-optimised CpG-free version of the hCMV enhancer, EF1 $\alpha$ promoter and intron
hCMV	human cytomegalovirus
HCV	hepatitis C virus
HD-Ad	helper-dependent adenovirus
HIV	human immunodeficiency virus
HSV TK	herpes simplex virus thymidine kinase
HTV	hydrodynamic tail vein
IDLV	integrase-deficient lentiviral vector

---

I/E	immediate/early
IL	interleukin
IFN	interferon
LB	Luria Bertani
LF	Lipofectamine 2000
LV	lentivirus
MBD	methyl-CpG-binding domain
mC	synthetic, codon-optimised CpG-free version of the mCMV enhancer
mCEFI	synthetic, codon-optimised CpG-free version of the mCMV enhancer, EF1 $\alpha$ promoter and intron
mCMV	murine cytomegalovirus
MCS	multiple cloning site
mFABP	murine fatty acid binding protein
MFI	mean fluorescence intensity
MIDGE	minimalistic, immunological-defined gene expression
N:P	ratio of lipid nitrogen to DNA phosphate content
NF $\kappa$ B	nuclear factor $\kappa$ B
NK	natural killer
NLS	nuclear localisation signal
NTPs	nucleoside triphosphates
PBS	phosphate buffered saline
PCL	periciliary liquid
pDNA	plasmid DNA
pDNA/GL67A	plasmid DNA complexed with liposomes of GL67A
PEI	polyethylenimine
pFAR	plasmids free of antibiotic resistance markers
photons/sec/cm <sup>2</sup> /sr	photons/second/cm <sup>2</sup> /steradian
PIC	pre-initiation complex

---

PLL	poly-L-lysine
Pol II	RNA polymerase II
polyA	polyadenylation
rAAV	recombinant adeno-associated virus
rAd	recombinant adenovirus
RLU	relative light units
rLV	recombinant lentivirus
rpm	revolutions per minute
RSV	respiratory syncytial virus
RT	room temperature
SATB1	special AT-rich sequence-binding protein 1
SCID	severe combined immunodeficiency
SD	standard deviation
SEAP	secreted alkaline phosphatase
SeV	Sendai virus; murine parainfluenza virus type I
SIN	self-inactivating
SIV	simian immunodeficiency virus
S/MAR	scaffold matrix attachment region
soCFTR2	synthetic, codon-optimised CpG-free version of the CFTR gene
soLux	synthetic, codon-optimised CpG-free version of the firefly luciferase gene
SP-B	surfactant protein B
TEER	trans-epithelial electrical resistance
TF	transcription factor
TFBS	transcription factor binding site
TI	transcriptional interference
TLR	toll-like receptor
TNF $\alpha$	tumour necrosis factor alpha

TSS	transcription start site
UbC	ubiquitin C
UKCFGTC	UK Cystic Fibrosis Gene Therapy Consortium
VSV-G	vesicular stomatitis glycoprotein
WFI	water for injection

---

## Table of Contents

<b>Chapter 1: Introduction</b> .....	<b>1</b>
<b>1.1. The lung as a target for gene therapy vectors</b> .....	<b>1</b>
1.1.1. Lung anatomy and structure .....	1
1.1.2. Proliferation and repair of the lung epithelium .....	2
1.1.3. Diseases of the lung targeted by gene therapy .....	6
1.1.4. Targeted cell types .....	6
1.1.5. Barriers to gene transfer .....	7
<b>1.2. Gene transfer to the lung for the treatment of cystic fibrosis</b> .....	<b>9</b>
1.2.1. CF lung disease .....	10
1.2.2. Current treatment strategies for CF lung disease .....	10
1.2.3. Vectors for CF clinical gene therapy .....	11
1.2.4. Adenoviruses and helper-dependent adenoviruses .....	12
1.2.5. Adeno-associated viruses .....	13
1.2.6. Retroviruses and Lentiviruses .....	15
1.2.7. Non-viral gene delivery systems .....	18
1.2.8. Naked DNA .....	18
1.2.9. Physical methods to increase transfection efficiency .....	19
1.2.10. Chemical methods to increase transfection efficiency .....	20
1.2.11. Cationic polymers .....	20
1.2.12. Cationic lipids .....	22
1.2.13. Other non-viral vectors for airway gene transfer .....	25
1.2.14. Use of the murine lung as a model for CF gene transfer .....	26
<b>1.3. Plasmid design for long-term transgene expression</b> .....	<b>27</b>
1.3.1. Self-replicating vectors .....	29
1.3.2. Integrating vectors .....	32
1.3.3. Minimal plasmid DNA vectors .....	33
1.3.4. CpG-depleted plasmids .....	35

---

1.3.5. Enhancer/promoter constructs .....	39
<b>1.4. Transcriptional regulation of transgene activity .....</b>	<b>43</b>
1.4.1. Histone modifications .....	44
1.4.2. Transcription factories .....	45
1.4.3. DNA methylation of the vector .....	45
1.4.4. Transcription factors .....	46
<b>1.5. Aims of this study .....</b>	<b>51</b>
<b>Chapter 2: Materials and Methods .....</b>	<b>53</b>
<b>2.1. Materials .....</b>	<b>53</b>
2.1.1. Reagents and solutions .....	53
2.1.2. Centrifugation .....	53
<b>2.2. Cell culture .....</b>	<b>53</b>
2.2.1. Cell culture conditions for the A549 cell line .....	53
2.2.2. Cell culture conditions for the LA-4 cell line .....	54
2.2.3. Cell culture conditions for the air-liquid interface cultures .....	54
2.2.4. Transient transfection of cultured cell lines with plasmid DNA vectors .....	54
2.2.5. Preparation of cell culture lysates for quantification of luciferase activity .....	55
2.2.6. Transient transfection of ALI cultures with plasmid DNA vectors .....	55
2.2.7. Preparation of ALI cultures for quantification of luciferase activity .....	55
<b>2.3. Plasmid cloning techniques .....</b>	<b>56</b>
2.3.1. PCR amplification .....	56
2.3.2. Gel electrophoresis .....	56
2.3.3. Restriction enzyme digests .....	57
2.3.4. Gel extraction of DNA fragments .....	57
2.3.5. Ligation of DNA fragments .....	58
2.3.6. Site-directed mutagenesis .....	58
<b>2.4. Amplification of plasmid DNA .....</b>	<b>58</b>
2.4.1. Preparation of selective Luria Bertani plates .....	58

---

2.4.2. Bacterial transformation of plasmid DNA.....	59
2.4.3. Inoculation into growth media.....	59
2.4.4. Purification of plasmid DNA from transformed bacteria.....	59
2.4.5. Spectrophotometric quantification of DNA concentration in solution.....	59
<b>2.5. Animal studies .....</b>	<b>60</b>
2.5.1. Mouse strains .....	60
2.5.2. Mouse anaesthesia .....	60
2.5.3. Preparation of lipoplexes for intranasal instillation .....	61
2.5.4. Preparation of polyplexes for aerosolisation.....	61
2.5.5. Delivery to the mouse lung via intranasal instillation .....	61
2.5.6. Delivery to the mouse lung via aerosolisation .....	62
2.5.7. Delivery to the mouse liver via hydrodynamic tail vein injection.....	62
2.5.8. Harvesting of lung tissue and preparation of lysates.....	62
2.5.9. Harvesting of liver tissue and preparation of lysates .....	63
2.5.10. Preparation of mice for bioluminescence imaging.....	63
<b>2.6. Quantification of transgene activity .....</b>	<b>63</b>
2.6.1. The quantification of luciferase activity from lysates .....	63
2.6.2. The quantification of total protein from lysates .....	64
2.6.3. The quantification of luciferase activity using bioluminescence imaging.....	64
2.6.4. The visualisation of GFP activity using fluorescence microscopy .....	65
<b>2.7. Determination of plasmid DNA levels.....</b>	<b>65</b>
2.7.1. Extraction of DNA from liver lysates .....	65
2.7.2. Quantification of DNA using quantitative real-time PCR .....	65
2.7.3. Quantification of bacterial genomic DNA in plasmid DNA preparations using GT115- specific quantitative real-time PCR.....	66
<b>2.8. Statistical analysis .....</b>	<b>67</b>
<b>Chapter 3: Investigation of lung cell culture models .....</b>	<b>70</b>
<b>3.1. Introduction.....</b>	<b>70</b>

---

<b>3.2. Results</b> .....	<b>77</b>
3.2.1. Establishing A549 cell culture and transfection conditions .....	77
3.2.2. Time-course studies of pG4-hCEFI-soLux and pG4-mCEFI-soLux in A549 cells ....	79
3.2.3. Determining LA-4 transfection conditions .....	84
3.2.4. Time-course study using in LA-4 cells .....	86
3.2.5. Establishing transfection conditions for ALI cultures .....	88
3.2.6. Comparison of luciferase activity in ALI cultures .....	93
<b>3.3. Discussion</b> .....	<b>95</b>
<b>Chapter 4: Investigation of lung transgene activity using bioluminescence imaging</b>	<b>98</b>
<b>4.1. Introduction</b> .....	<b>98</b>
<b>4.2. Results</b> .....	<b>101</b>
4.2.1. Using BLI to detect luciferase activity after dosing with pG4-mCEFI-soLux.....	101
4.2.2. Comparison of luciferase activity of pG4-mCEFI-soLux and pG4-hCEFI-soLux....	107
4.2.3. Construction of plasmids encompassing enhancer deletions.....	110
4.2.4. Luciferase activity in A549 cells from plasmids containing enhancer variants .....	117
4.2.5. Time-course studies of plasmids containing deletions of the hC enhancer in the mouse lung .....	120
4.2.6. Time-course studies of plasmids containing deletions of the mC enhancer in the mouse lung .....	125
<b>4.3. Discussion</b> .....	<b>130</b>
4.3.1. Use of bioluminescence imaging after delivery of non-viral vectors to the lung .....	130
4.3.2. Dynamic range of BLI compared with standard luminometry .....	130
4.3.3. Variant upstream regulatory elements and their effects on transgene activity .....	132
4.3.4. Conclusions .....	133
<b>Chapter 5: Transcription factor binding site analysis of the hC and mC enhancer variants</b>	<b>135</b>
<b>5.1. Introduction</b> .....	<b>135</b>

---

<b>5.2. Results</b> .....	<b>137</b>
5.2.1. Identification of predicted transcription factor binding sites using MacVector .....	137
5.2.2. Use of the Genomatix software suite.....	141
5.2.2.2. ModelInspector to identify joint promoter modules .....	153
5.2.2.3. Identification of overrepresented TFBSs .....	156
<b>5.3. Discussion</b> .....	<b>161</b>
5.3.1. Transcription factor binding site analysis.....	161
5.3.2. Identification of sites in the hC enhancer which may mediate high levels of transgene activity.....	161
5.3.3. Overrepresented binding sites as a source of competition.....	163
5.3.4. Conclusions .....	165
<b>Chapter 6: Effect of deletions of the minimal hC enhancer on luciferase activity in the mouse lung</b> .....	<b>166</b>
<b>6.1. Introduction</b> .....	<b>166</b>
<b>6.2. Results</b> .....	<b>167</b>
6.2.1. Plasmid vectors used in Chapter 6 studies .....	167
6.2.2. Effect of the mC enhancer placed upstream of the minimal hC enhancer .....	171
6.2.3. Luciferase activity from the 70 bp region of interest .....	173
6.2.4. Quantification of GT115 genomic DNA contamination .....	175
6.2.5. Investigation of the effect of a spacer region.....	177
6.2.6. Luciferase activity after mutation of the NFkB binding site .....	180
6.2.7. Luciferase activity from a 50 bp enhancer .....	182
<b>6.3. Discussion</b> .....	<b>184</b>
<b>Chapter 7: Utility of the minimal hC enhancer</b> .....	<b>187</b>
<b>7.1. Introduction</b> .....	<b>187</b>
<b>7.2. Results</b> .....	<b>189</b>
7.2.1. Plasmid vectors used in Chapter 7 studies .....	189

---

7.2.2. Using the minimal hC enhancer with an alternative CpG-free promoter .....	192
7.2.3. Comparison of luciferase activity after hydrodynamic tail vein injection .....	194
7.2.3.1. Establishing high-level transgene persistence in immuno-deficient mice .....	194
7.2.3.2. Quantification of plasmid DNA in the mouse liver .....	197
7.2.3.3. Utility of the minimal hC enhancer in the liver .....	199
7.2.3.4. Gene expression in mouse lung and liver .....	202
7.2.4. Influence of the adaptive immune system on transient transgene activity from pG4-mCEFI-soLux .....	205
7.2.5. Study utilising the minimal enhancer as a tandem dual enhancer .....	207
<b>7.3. Discussion .....</b>	<b>209</b>
7.3.1. Use of the mFABP promoter .....	209
7.3.2. Transgene activity in the mouse liver .....	210
7.3.3. The role of the adaptive immune response in transient transgene activity .....	211
7.3.4. Gene expression differences between mouse lung and liver .....	212
7.3.5. Use of the minimal enhancer as a dual tandem enhancer .....	213
7.3.6. Conclusions .....	213
<b>Chapter 8: Discussion .....</b>	<b>215</b>
<b>8.1. Identification of a model for investigation of persistent transgene expression ..</b>	<b>216</b>
<b>8.2. The use of <i>in silico</i> methods identified potentially interesting TFBSs .....</b>	<b>218</b>
<b>8.3. The hC enhancer contains very few overrepresented binding sites compared with the mC enhancer .....</b>	<b>220</b>
<b>8.4. The minimal hC enhancer did not mediate high-level activity in the liver .....</b>	<b>222</b>
<b>8.5. Implications .....</b>	<b>222</b>
<b>8.6. Conclusion .....</b>	<b>224</b>
<b>Chapter 9: References .....</b>	<b>225</b>

---

## List of Figures

Figure 1.1. Diagram of lung structure and potential extracellular barriers to gene transfer.....	4
Figure 1.2. Diagram of a generic modular plasmid.....	28
Figure 1.3. Diagram of a generic modular plasmid containing an S/MAR element. ....	31
Figure 1.4. Lung transgene activity in a mouse aerosol model after delivery of pDNA/GL67A lipoplexes. ....	42
Figure 1.5. Model of transcriptional activation. ....	50
Figure 3.1. Schematic showing detail of a MucilAir ALI culture.....	73
Figure 3.2. Plasmid constructs used in Chapter 3 studies. ....	76
Figure 3.3. Comparison of luciferase activity after transfection of A549 cells harvested from high- or low-density parental cultures using PEI or Lipofectamine 2000 (LF). ....	78
Figure 3.4. Comparison of luciferase activity after transfection of (A) high and (B) low densities of A549 cells with pG4-mCEFI-soLux or pG4-hCEFI-soLux over an extended period. ....	82
Figure 3.5. Comparison of luciferase activity after transfection of low-confluence A549 cells with pG4-mCEFI-soLux or pG4-hCEFI-soLux over a short duration.....	83
Figure 3.6. Comparison of luciferase activity after transfection of low- and high-density LA-4 cells with pG4-hCEFI-soLux.....	85
Figure 3.7. Time-course of luciferase activity after transfection of low-confluence LA-4 cells with pG4-mCEFI-soLux or pG4-hCEFI-soLux.....	87
Figure 3.8. Comparison of luciferase activity after transfection of ALI cultures with pG4-mCEFI-soLux or pG4-hCEFI-soLux. ....	91
Figure 3.9. Comparison of GFP activity after transfection of ALI cultures with pEGFP-N1...	92
Figure 3.10. Time-course of luciferase activity in ALI cultures after transfection with pG4-mCEFI-soLux or pG4-hCEFI-soLux in the presence of 10 mM EGTA.....	94
Figure 4.1. Plasmids used in BLI studies. ....	104
Figure 4.2. Comparison of luciferase activity after dosing of pG4-mCEFI-soLux using (A) BLI or (B) standard luminometry.....	106

---

Figure 4.3. Comparison of luciferase activity after aerosolisation of pG4-mCEFI-soLux or pG4-hCEFI-soLux using (A) BLI or (B) standard luminometry. ....	109
Figure 4.4. Diagram of plasmid DNA constructs containing hC enhancer variants. ....	113
Figure 4.5. Diagram of plasmid DNA constructs containing mC enhancer variants. ....	115
Figure 4.6. Diagram of the plasmid DNA construct containing no enhancer. ....	116
Figure 4.7. A time-course experiment comparing luciferase activity in A549 cells after transfection with plasmids containing deletions of the (A) hC enhancer or (B) mC enhancer. ....	119
Figure 4.8. Luciferase activity after aerosol administration of plasmids containing deletions of the hC enhancer using (A) BLI or (B) standard luminometry. ....	123
Figure 4.9. Luciferase activity after aerosol administration of enhancer variants of pG4-mCEFI-soLux variants using standard luminometry. ....	128
Figure 5.1. Sequence alignment of the hC and mC enhancer sequences. ....	138
Figure 5.2. A schematic of the matrix concept used in the Genomatix software suite. ....	142
Figure 5.3. Predicted transcription factor binding sites which occur once in the hC enhancer and are absent in the mC enhancer. ....	149
Figure 5.4. Predicted transcription factor binding sites which occur once in the mC enhancer and are absent in the hC enhancer. ....	151
Figure 6.1. Enhancer/promoter regions of plasmid constructs utilised in the studies in Chapter 6. ....	170
Figure 6.2. Comparison of luciferase activity after aerosol administration of plasmids containing the (A) hCd3 enhancer, (B) full-length mC and hCd3 dual enhancer and (C) hCd4 enhancer. ....	172
Figure 6.3. Comparison of luciferase activity after aerosol administration of plasmids containing the (A) hCd3 enhancer, (B) hCd5 enhancer containing the 70 bp region of interest and (C) hCd4 enhancer. ....	174
Figure 6.4. Comparison of luciferase activity after aerosol administration of plasmids containing the (A) hCd3 enhancer, (B) hCd5 enhancer containing the 70 bp region of	

interest, (C) hCd6 enhancer containing the 70 bp region of interest and a 23 bp spacer and (D) hCd4 enhancer.....	179
Figure 6.5. Comparison of luciferase activity after aerosol administration of plasmids containing the (A) hCd5 enhancer containing the 70 bp region of interest, (B) hCd7 enhancer containing the 70 bp region of interest with a mutated NF $\kappa$ B binding site and (C) hCd4 enhancer.....	181
Figure 6.6. Comparison of luciferase activity after aerosol administration of plasmids containing the (A) hCd5 enhancer containing the 70 bp region of interest, (B) hCd8 enhancer containing the 3' 50 bp of the 70 bp region of interest and (C) hCd4 enhancer.....	183
Figure 7.1. Enhancer/promoter regions of plasmid constructs utilised in the studies in Chapter 7.....	191
Figure 7.2. Comparison of luciferase activity after aerosol administration of plasmids containing the mFABP promoter and (A) hC enhancer, (B) hCd3 enhancer and (C) no enhancer. ....	193
Figure 7.3. Comparison of luciferase activity after hydrodynamic tail vein injection dosing of pG4-hCEFI-soLux in SCID and BALB/c mice. ....	196
Figure 7.4. Quantification of plasmid DNA in the liver of SCID and BALB/c mice.....	198
Figure 7.5. Comparison of luciferase activity in SCID and BALB/c mice after dosing by hydrodynamic tail vein injection of plasmids containing the (A) hC enhancer and (B) hCd3 enhancer.....	201
Figure 7.6. Comparison of luciferase activity after aerosol dosing of pG4-mCEFI-soLux in BALB/c or SCID mice.....	206
Figure 7.7. Comparison of luciferase activity after aerosol administration of plasmids containing the (A) dual enhancer using the hCd3 sequence in tandem, (B) hCd3 enhancer and (C) hCd4 enhancer.....	208

---

## List of Tables

Table 1.1. Proposed niches and the respective types of facultative stem cells in the mouse lung.....	5
Table 2.1. Plasmid DNA names and sources.....	69
Table 3.1. Studies utilising ALI cultures to investigate reasons for low transfection efficiency after non-viral gene transfer to the lung. ....	74
Table 4.1. Overview of plasmid name and enhancer length. ....	111
Table 4.2. Statistical information for Figure 4.8. ....	124
Table 4.3. Statistical information for Figure 4.9. ....	129
Table 5.1. Number of potential transcription factor binding sites identified by MacVector. .	139
Table 5.2. Output names of potential transcription factor binding sites identified by MacVector. ....	140
Table 5.3. Number of potential transcription factor binding sites identified by MatInspector. ....	145
Table 5.4. Output names of potential families of transcription factor binding sites identified by MatInspector.....	146
Table 5.5. Output names of potential transcription factor binding sites identified by MatInspector.....	147
Table 5.6. Detail of TFBSs uniquely present in the 3' 93 bp of the hC enhancer.....	150
Table 5.7. Detail of TFBSs uniquely present in the 5' 126 bp of the mC enhancer.....	152
Table 5.8. Output names of potential interacting transcription factor binding sites identified by ModelInspector. ....	154
Table 5.9. Number of overrepresented TFBSs in the EF promoter, and hC- and mC-derived enhancers.....	157
Table 5.10. The top two overrepresented binding sites found in the EF promoter and all hC- and mC-derived enhancers. ....	158
Table 5.11. Matrix families represented in the EF promoter and hC enhancer variants. ....	159
Table 5.12. Matrix families represented in the EF promoter and mC enhancer variants. ...	160

Table 6.1. Overview of plasmid name, original source and enhancer length.....	168
Table 6.2. Summary of bacterial chrDNA present in pDNA preparations.....	176
Table 7.1. Overview of plasmid name, original source and enhancer length.....	190
Table 7.2. Gene expression data obtained via the BioGPS public repository.....	204

## **Chapter 1: Introduction**

### **1.1. The lung as a target for gene therapy vectors**

The basic premise of gene therapy involves the delivery of a recombinant nucleic acid to a target cell type or tissue in order to correct a disease phenotype. Attempts to translate this concept into a successful therapeutic approach have generally been hampered by the lack of therapeutically relevant levels of the desired transgene products and also due to adverse immune responses and associated acute toxicity after vector delivery. However, through enhanced vector design and a deeper understanding of immunological and inflammatory obstacles, researchers are beginning to achieve some clinical success. The lung has been the subject of intense investigation for gene delivery research for the potential treatment of inherited and acquired pulmonary disorders. The relative accessibility of the lung via well-established procedures such as nebulisation is an advantage for the delivery of gene transfer agents. However, the lung is a complex organ that has evolved numerous extra- and intracellular barriers to invasion of particles and infectious agents. Hence, gene transfer agents need to be able to negotiate the internal milieu of the lung in order to be further developed for clinical benefit.

#### **1.1.1. Lung anatomy and structure**

The human lung is a complex organ consisting of a number of highly specialised cell types (see Figure 1.1). The main regions of the lung can be divided into two parts: the conducting airways of the lower respiratory tract and the respiratory zone where gas exchange with capillary endothelial cells takes place. The large conducting area begins with the trachea, which divides first into two bronchi and then subsequently into smaller airway passages. The small conducting airways (bronchioles) consist of a single layer of cuboidal ciliated cells, basal cells and columnar secretory Clara cells (see Figure 1.1). In the gaseous exchange region, the alveolar cells are lined with large, squamous, gas exchanging type I pneumocytes, which make up to 99 % of the parenchyma. Type I cells are also at the interface with pulmonary capillaries to facilitate gas exchange. Interspersed among the type I

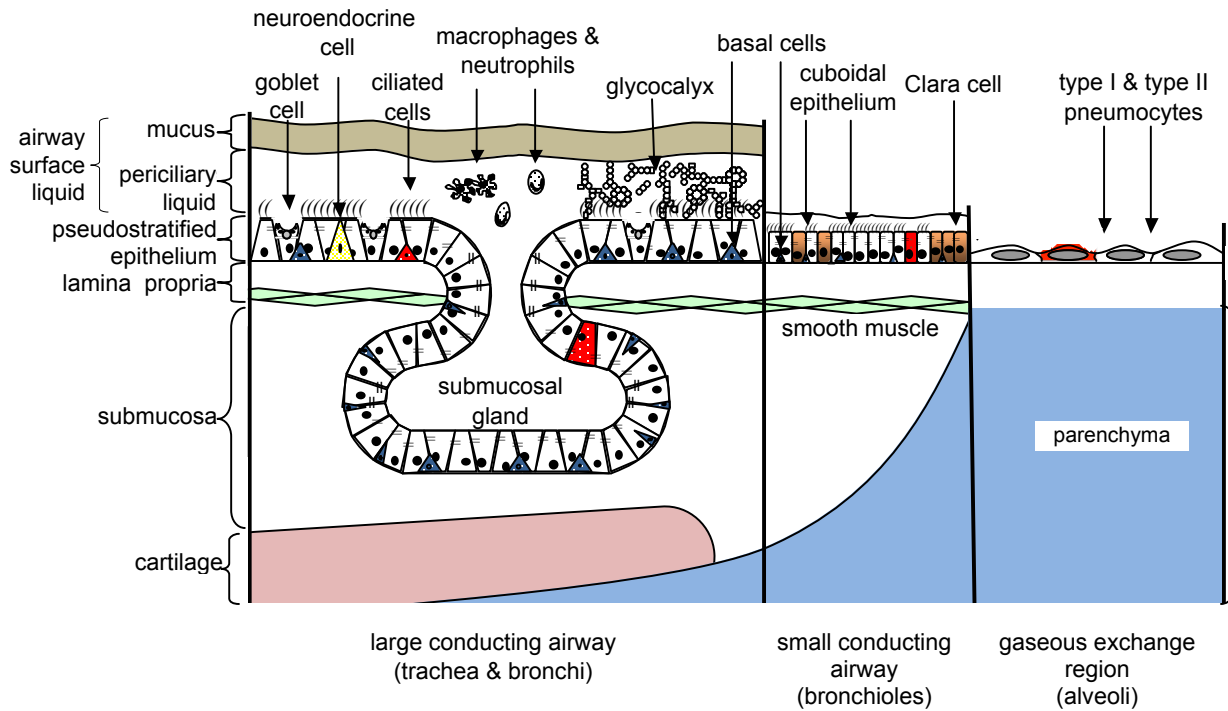
cells are cuboidal, surfactant-secreting type II pneumocytes. In mice, the same cells are present but differ in their pattern of distribution along the respiratory tree (Liu and Engelhardt, 2008).

### **1.1.2. Proliferation and repair of the lung epithelium**

The lung is continuously challenged by exposure to environmental and biological insults, such as pollution and the presence of infectious agents such as viruses or bacteria. In response to exposure to these potentially harmful particles, the damaged epithelium of the lung must be maintained through a process of cell proliferation. The identification of a single adult stem cell population capable of regenerating the different epithelial cells of the lung remains elusive. Advances in techniques which allow for single cell labelling and tracking *in vivo* in the mouse have identified a number of 'facultative' stem cells, which are differentiated cells that are normally quiescent but can act as stem/progenitor cells in response to injury (Rawlins and Hogan, 2006). This putative population can give rise to daughter cells that differentiate into one or more cell types (Rawlins and Hogan, 2006) adjacent to their respective micro-environmental niches (Liu and Engelhardt, 2008). Table 1.1 contains details on the putative stem cell niches of the mouse lung.

In terms of gene delivery for the treatment of chronic diseases, it would be beneficial to target the 'stem cell' niches described above. However, there is presently no viable approach to specifically target this population *in vivo*. Studies have suggested that mature ciliated cells are terminally differentiated and do not self-renew or contribute to other lineages (Rawlins and Hogan, 2008; Rawlins et al., 2007). Under specific conditions, epithelial cell half-life has been estimated to be 6 months in the mouse trachea and 17 months in the mouse bronchioles (Rawlins and Hogan, 2008). Therefore, potential for long-lived transgene expression after a single administration may be possible. However, for chronic conditions that will require life-long therapy, repeat administration of any gene therapy vector will be required to replace the transgene in the cells lost through turnover until a definitive stem cell

population can be identified and targeted. These studies were performed on mice housed in a specific pathogen-free facility, but it is conceivable that in an environment where the airway epithelium is constantly exposed to harmful particles, damage and physiological changes may occur on a frequent basis. As a result, cell turnover rate may increase. Interestingly, studies have shown that expression of basal cell markers were found in the distal lungs of mice after virus infection, where basal cells are not typically found (Byers et al., 2013). This may indicate that the cellular composition of the lung epithelium is influenced by environmental exposure (Byers et al., 2013; Kumar et al., 2011). This further supports the requirement for repeated administration of gene therapy vectors to treat chronic conditions.



**Figure 1.1. Diagram of lung structure and potential extracellular barriers to gene transfer.**

Epithelia of the adult human lung consist of different cell types in the various anatomical compartments. The surface epithelium of the large conducting airway is pseudostratified and consists of ciliated and non-ciliated secretory columnar cells such as neuroendocrine cells, basal cells and goblet cells, which secrete mucus. This region represents the first barriers to incoming particles, and includes the airway surface liquid, composed of a mucus gel layer and periciliary liquid layer, epithelial glycocalyx and cilia. Neutrophils and alveolar macrophages may also act as barriers to foreign agents. Additionally, tight junctions (=) prevent entry of pathogens and particles through the epithelium. The small conducting airway consists of a single layer of cuboidal ciliated cells, basal cells and columnar secretory Clara cells. Different putative regenerative stem cell populations (red) include basal cells in the trachea and bronchi, an unknown cell type in the submucosal glands, variant Clara cells in the small conducting airway and type II cells in the alveolar region (for details, see Section 1.1.2 and Table 1.1). Figure has been adapted from Gill *et al.*, 2004 and Rawlins & Hogan, 2006.

**Table 1.1. Proposed niches and the respective types of facultative stem cells in the mouse lung.**

Location	Proposed niche	Facultative stem cell type	Self-renewal or differentiating potential	Reference(s)
Large conducting airway	Dynamic, not well defined	Basal cells	Basal cells Clara-like cells Ciliated cells	(Hong et al., 2004; Rock et al., 2010)
Small conducting airway	Neuroendocrine bodies	Variant Clara cells	Clara cells Ciliated cells	(Hong et al., 2001)
Gaseous exchange region	Alveolar Type II niche	Subset of alveolar Type II cells	Type I cells Type II cells	(Desai et al., 2014; Evans et al., 1975)

### **1.1.3. Diseases of the lung targeted by gene therapy**

Gene therapy approaches for treatment of acute and chronic lung diseases are currently being investigated. Acute or acquired conditions such as lung cancer and bronchiolitis obliterans may benefit from a single dose of a vector capable of rapid and high levels of gene expression of a transiently expressed gene. However, there is little anticipated need for repeated administration, or persistent action; a single dose of gene therapy with an acute effect is sufficient as long as the treatment is effective after one administration. Chronic conditions, such as asthma, cystic fibrosis (CF) and chronic obstructive pulmonary disease, however, may require life-long benefit from a persistently expressing transgene, which will be difficult to maintain in the absence of treatment of a stem cell population in the lung (Section 1.1.2). Even gene transfer vectors that are highly efficient, delivering to a high percentage of epithelial cells, will eventually need to be re-administered once the terminally differentiated cells are replaced in the epithelium.

### **1.1.4. Targeted cell types**

The field generally accepts that the nature of the disease will dictate both the cell type to be targeted and the vector required to deliver their genetic cargo (Ibraheem et al., 2014). Even when the cell target(s) are well defined, the delivery of the therapeutic gene is hampered by a number of barriers. Depending on the respiratory disease, and the therapeutic goal, possible target cells in the lung can encompass epithelial cells, alveolar cells, macrophages, or endothelial cells (reviewed in Sanders et al., 2009). The selected route of administration will also depend largely on the distribution and function of the native protein (Davies et al., 2001). For CF, native and functional cystic fibrosis transmembrane conductance regulator (CFTR) protein is expressed in various regions and cell types throughout the airway, but it is thought that the ciliated airway epithelial cells lining the small airways are the most likely target cell for gene replacement (Griesenbach and Alton, 2012). In the case of Surfactant protein B (SP-B) deficiency, type II alveolar cells in the gaseous exchange region of the lung

are the targets for gene therapy (Hidalgo et al., 2015). Therefore, vector delivery must be optimised for the targeted cell type.

### **1.1.5. Barriers to gene transfer**

For efficient gene transfer to take place, the vector must firstly be able to accumulate at the target site, gain access to the target cell type, be trafficked through the cell and, in most cases, the genetic material must enter the nucleus (Aneja et al., 2009). For this to occur in the airways, vectors need to overcome a multitude of physical and immune barriers before efficient gene transfer can take place. Indeed, in the case of non-viral vectors, it has been estimated at least  $1 \times 10^5$  plasmid molecules per cell are required in the extracellular compartment to ensure that a few molecules are taken up into the nucleus of non-mitotic cells (Lechardeur et al., 2005).

#### **1.1.5.1. Extracellular barriers**

The first physical barrier that a vector encounters in the large airway is the airway surface liquid (ASL) that covers the conducting airways and is composed of two layers: a viscous mucus gel layer and the periciliary liquid (PCL) (Matsui et al., 1998b) (Figure 1.1). Inhaled materials are captured by mucins present in the mucus gel layer and are transported by the cilia to the oesophagus where elimination occurs via coughing (Sanders et al., 2009). Several antimicrobial peptides, such as defensins, lysozyme, lactoferrin and secretory leukocyte proteinase inhibitor, are also present in the mucus (Hiemstra et al., 2015). Such peptides contribute to innate immunity and play a major role in host defense, and as such, represent an additional barrier to successful entry of vectors into the cell (Virella-Lowell et al., 2000). The PCL bathes the cilia and is important in mucociliary clearance (Matsui et al., 1998a). Beneath the mucus layer, the glycocalyx, composed of carbohydrate, glycoproteins and polysaccharides (Linden et al., 2008), has also been shown to impede entry of vectors into the cell (Pickles et al., 2000). Thus, a large proportion of gene transfer vectors entering the lungs will be prevented from reaching the target cells. Even arrival at the cell surface does not guarantee efficient transfer if there is a paucity of cell receptors on the apical

surface (Gill and Hyde, 2014). It may be necessary for the vector to reach the basolateral surface of the target cell, a process which can be facilitated by the usage of tight-junction openers (Wang et al., 2000). Although effective in cell culture and *in vivo* models, this latter approach is considered controversial when considering the treatment of humans clinically, especially in the case of CF patients given that the lung is chronically infected with bacteria, which may result in systemic invasion of such pathogens. Additionally, the general toxicity profiles of different tight-junction modulators may automatically preclude their routine use in the clinic (Johnson et al., 2003b). Finally, the cell's plasma membrane is also a physical barrier that hinders the free diffusion of non-viral vectors into the cell.

#### **1.1.5.2. Intracellular barriers**

Upon successful cell entry, most vectors must then traverse the intracellular environment in order to reach the nucleus. Non-viral vectors require the use of microtubule networks and cytoplasmic adapter proteins, such as dynein, to transport the plasmid DNA (pDNA) to the nucleus (Badding et al., 2012). Viral gene delivery vectors exploit the natural ability of viruses to efficiently deliver genetic material to the nucleus. For example, lentiviruses can naturally penetrate the intact nuclear membrane and transduce non-dividing cells. Many non-viral vectors are taken up via endocytosis, pinocytosis or phagocytosis, or via means that physically allow the pDNA to enter the cell and may undergo degradation in the mature endosome (lysosome); if successfully released from the endosome they can be degraded in the cytoplasm by cellular nucleases (Pollard et al., 2001). Also, diffusion rate of pDNA in the cell cytoplasm is low (Lukacs et al., 2000). In addition to endosomal degradation, cytoplasmic-based degradation can also occur (Escoffre et al., 2010). Finally, nuclear import is a major barrier for gene therapy vectors (Brunner et al., 2002). Once inside the nucleus, the vector must be trafficked to active sites of transcription (Aneja et al., 2009). While a number of techniques, such as electroporation and magnetofection (see Section 1.2.9), have been exploited to enhance the number of vector molecules which can physically enter the cell, the myriad of intracellular barriers described above still limit the quantity and quality of

vector which can enter the nucleus. Therefore, entry into the cell does not equate to successful gene transfer and therapeutic benefit.

## **1.2. Gene transfer to the lung for the treatment of cystic fibrosis**

Cystic fibrosis (CF), a monogenic disorder affecting approximately 1 in 2500 live births in the UK, is caused by mutations in the *CFTR* gene (Riordan et al., 1989). In 2010, the median age of death was 29 years ([www.cysticfibrosis.org.uk](http://www.cysticfibrosis.org.uk)). Mutations in the *CFTR* gene impair the function of CFTR, an epithelial chloride channel required for proper function of ion transport in epithelial cells throughout the body. However, mortality arises from eventual lung destruction from chronic and non-resolving activation of the innate immune system in response to persistent bacterial infection (Koch and Hoiby, 1993). After the successful cloning of the *CFTR* gene in 1989 (Riordan et al., 1989), the expectation was that the correction of the lung defect with the correct copy of the *CFTR* gene would be relatively straightforward (Lindee and Mueller, 2011) and as such, CF has been at the forefront of research and development of lung gene therapies. The benefit of progressing gene therapy for the treatment of CF lung disease is partially due to the large number of known mutations of the *CFTR* gene. Over 1900 mutations have been identified, encompassing a range of channel deficiencies, from complete absence of the CFTR protein at the cell membrane to functional mutants which are able to reach the cell membrane but have a shortened half-life due to protein instability (<http://www.genet.sickkids.on.ca/cftr/Home.html>). Gene complementation to correct the faulty gene has the potential to treat not only the currently known mutations, but also any that may be discovered in the future.

Studies found that certain CF mutations are associated with milder clinical phenotypes, and that a small amount of residual CFTR function at the membrane is sufficient to reduce lung symptoms (Sheppard et al., 1995). Cell culture mixing experiments also suggest that as little as 6 to 10 % of non-CF cells mixed with CF cells can restore CFTR-mediated chloride secretion to non-CF levels (Johnson et al., 1992). Additionally, in a study which inter-crossed

mice carrying different *Cftr* alleles, it was observed that only 5 % of normal protein expression resulted in a correction of up to 50 % of the normal chloride ion transport defect (Dorin et al., 1996). Researchers in the field were encouraged by these studies, which together suggest that modest levels of CFTR expression may have a significant clinical impact in treating CF lung disease.

### **1.2.1. CF lung disease**

Many hypotheses exist which try to link defective CFTR function and airway disease in CF; these hypotheses look at the role of defective ion transport and their effect on antimicrobial function/secretion, mucociliary clearance, colonisation of *Pseudomonas aeruginosa*, and bacterial clearance in the human lung (reviewed in Verkman et al., 2003). Furthermore, non-inflammatory mechanisms related to cellular dysfunction in the airway wall due to absent or reduced levels of CFTR protein may contribute to airway remodelling, which is characterised by extensive structural airway changes, including thickened airway walls, narrowed airway lumens, air trapping and abnormal widening of the airways (bronchiectasis) (Murphy and Atala, 2013). The precise understanding of innate immunity in CF has been hampered due to uncertainties regarding the roles of inflammation and infection and the onset of lung disease.

### **1.2.2. Current treatment strategies for CF lung disease**

Thus far, the only pharmacological product which is licensed to treat the underlying molecular basis of CF is ivacaftor (trade name Kalydeco, Vertex Pharmaceuticals, Boston, MA, USA) (Davis et al., 2012). Ivacaftor is a small molecule which acts as a potentiator by increasing the probability of open CFTR chloride channels at the epithelial cell surface (Yu et al., 2012). Significant clinical benefit was demonstrated in patients carrying the G551D mutation, which affects 4 to 5 % of the CF population (Ramsey et al., 2011). Further clinical studies were initiated to determine whether ivacaftor could be used in combination with lumacaftor, a small molecule that was developed to increase the trafficking of CFTR protein

to the cell surface (Van Goor et al., 2010). The aim was to treat patients who carry the most common CF mutation, a 3-base pair deletion of the amino acid phenylalanine at position 508 (F508del-CFTR), which affects up to 70 % of the CF population (Boyle et al., 2014). To date, the results suggest substantially more modest clinical benefit than observed in G551D patients with ivacaftor alone for patients who are homozygous for the F508del-CFTR mutation but not those who are heterozygous (Pettit and Fellner, 2014).

Other current treatments for CF lung disease are not curative; instead, they focus on the treatment of symptoms. Such therapies include physiotherapy, which uses airway clearance techniques to clear the viscous mucus from the lungs to minimise the risk of infection. A range of inhaled drugs may also be used on a daily basis; these include bronchodilators, expectorants, steroids, anti-mucolytics and antibiotics. The frequency of such treatments increases with active exacerbations and place a high therapeutic burden on the patient.

### **1.2.3. Vectors for CF clinical gene therapy**

Overall, 25 Phase I/II clinical trials involving over 470 patients have been completed; these studies have led to the refinement and clinical development of new gene delivery vectors. Thus far, the clinical success of viral and non-viral approaches has been hampered by a number of physiological reasons (Section 1.1.5). In addition to the barriers in place in a normal, healthy lung, efficacious gene transfer to CF lungs faces additional difficulties due to the excessive production of sputum; both viral and non-viral vectors have been shown to be impeded by these viscoelastic, gel-like airway secretions (Kitson et al., 1999). Viral gene delivery vectors are based on modified viruses with airway cell tropism and exploit the natural ability of viruses to efficiently enter host cells, deliver genetic material and replicate in an efficient manner. However, all viral vectors generate an inflammatory and immune response to some degree, and those tested clinically thus far have not been able to be re-administered without a significant impact on transgene expression (Sumner-Jones and Gill, 2014). Non-viral vectors have been developed to avoid the use of viral proteins, which in

turn may increase the probability of successful re-administration. The most important attribute required by any vector is the ability to mediate targeted gene delivery and transgene expression without widespread distribution to non-target tissues or the induction of harmful adverse reactions. A number of vectors have been assessed to identify one that may satisfy both the safety and efficacy components of successful gene therapy.

#### **1.2.4. Adenoviruses and helper-dependent adenoviruses**

Recombinant adenovirus (rAd) is a non-enveloped, replication-incompetent, double-stranded DNA virus that is able to effectively transduce cells independent of active cell division. Upon entry into the nucleus, the viral genome remains episomal. Due to this and the constant turnover of airway epithelia (Section 1.1.2), repeat administration would be required for chronic lung diseases such as CF. Adenoviruses have natural tropism for lung epithelial cells and, as such, were thought to be the ideal vector to be evaluated clinically for CF lung disease. Indeed, rAd vectors were the first gene transfer agents to be assessed in human trials. Although rAd was found to be safe, they were only moderately effective in transducing human airway cells (Perricone et al., 2001), due to the lack of two crucial receptor proteins - coxsackie-adenovirus receptors (CAR) (Pickles et al., 1998) and  $\alpha_v\beta_5$  integrins (Goldman and Wilson, 1995) on the apical surface of epithelial cells. Additionally, cell entry into human lung epithelial cells (Pickles et al., 2000) and to the mouse lung *in vivo* (Stonebraker et al., 2004) were hampered by components of the glycocalyx.

In total, ten clinical trials assessing the efficacy of CF gene transfer with rAd vectors have been performed. In a pivotal proof-of-concept trial, the nasal epithelium of three patients was targeted by a first generation rAd carrying the *CFTR* gene (Zabner et al., 1993). Although correction of the chloride channel function was transient, this study demonstrated that using a gene therapy strategy to correct the underlying basis of CF lung disease is possible. Further studies targeting both the nose and lung went on to show that transgene expression was not sustained and that repeat administration was not feasible (Harvey et al., 1999). This

lack of efficacy after re-administration was determined to be due to inflammation, neutralising antibodies and cytotoxic T-lymphocyte (CTL) responses after the initial administration of the vector (Yang et al., 1995).

Helper-dependent adenovirus (HD-Ad) vectors were developed to limit the immune response to rAd (Parks et al., 1996). Also known as 'gutless' vectors, HD-Ad are depleted of all viral genes apart from the inverted terminal repeats and the viral packaging sequences (Kochanek et al., 1996). Studies in mice demonstrated reduced toxicity and immunogenicity (Palmer and Ng, 2005). However, HD-Ad capsid proteins are targeted by neutralising antibodies (Sinn et al., 2011) and as such, the question of repeat administration still remains to be addressed.

### **1.2.5. Adeno-associated viruses**

Adeno-associated viruses (AAV) are non-pathogenic, replication-deficient, single-stranded DNA viruses belonging to the parvovirus family. Wild-type AAV is capable of genome integration, but recombinant AAV (rAAV) vectors were widely thought to remain episomal, either as circular molecules or linear concatemers (Schnepp et al., 2005). However, recent studies have shown that rAAV can integrate into the mouse liver at levels higher than previously thought (Chandler et al., 2014). Furthermore, an integration-site analysis was performed on muscle biopsies that were obtained from human subjects who participated in the clinical trials for assessment of the rAAV1 vector termed Glybera (uniQure, Amsterdam, NL) (Gaudet et al., 2012), the first gene therapy product to be marketed in the Western world, being used to treat patients with lipoprotein lipase deficiency. The integration-site analysis revealed that low levels of integration occurred randomly, with a preference for integration into the mitochondrial genome (Kaepffel et al., 2013). Further investigation will be required to understand the wider implications of this level of integration, especially from a safety point of view.

One of the limiting factors to date precluding the widespread use of rAAV in CF research has been the limited packaging capacity of the vector. Although conflicting results have been reported, it is generally thought that the capacity of rAAV vectors is approximately 5 kb (Dong et al., 1996), which decreases the utility of rAAV in CF research due to the fact that human *CFTR* cDNA approaches this size. One approach to circumvent this limitation is the development of *CFTR* mini-genes, which in turn allows for the incorporation of longer, more efficient, promoters to direct higher transgene expression levels (Zhang et al., 1998). Further strategies include the exploitation of the propensity for rAAV to form concatemers *in vivo* by employing a *trans*-splicing approach, or by utilising the host cell's homologous recombination machinery (Duan et al., 2001). The underlying principle of both of these strategies is that two viruses are used to deliver the required elements and cDNA. If both viruses enter the same cell, the genomes may recombine to generate the full-length therapeutic gene. A subsequent study demonstrated efficient transduction of airway epithelial cells in the mouse (Halbert et al., 2002), thus providing proof-of-principle of utilising such an approach. Despite these advances in the field, the suitability of their use in the human lung remains unknown.

Clinically, rAAV2 has been delivered to the nose (Wagner et al., 1999; Wagner et al., 1998) and lungs (Aitken et al., 2001; Flotte et al., 2003; Moss et al., 2007). Treatment was generally well tolerated but levels of transduction and transgene expression were low. Alternative serotypes have been tested in order to increase transduction efficiency; for instance, rAAV1 and rAAV7 appear to favour the transduction of skeletal muscle, while rAAV6 and rAAV5 infect airway epithelial cells more efficiently (Ura et al., 2014). Nevertheless, despite advancements in the development of novel rAAV serotypes and the observation that showed that certain serotypes can be repeatedly administered to mice (Limberis and Wilson, 2006), recent clinical trials assessing the efficacy and safety of rAAV in the lung have not been performed. A recent study which utilised rAAV vectors devoid of all cytosine-phosphate-guanosine (CpG) dinucleotides resulted in attenuated adaptive immune responses and prolonged transgene expression after muscular injection into mice (Faust et

al., 2013). Given that CpG-free pDNA vectors mediate high-level and persistent transgene activity in the mouse lung in the absence of detectable inflammation (see Section 1.3.4) (Hyde et al., 2008), the administration of CpG-free rAAV vectors into the lung may also circumvent the known inflammatory issues.

### **1.2.6. Retroviruses and Lentiviruses**

Retroviruses are a family of enveloped, single-stranded RNA viruses that integrate into the host genome. The ability of retroviruses to integrate was deemed to be advantageous, especially in the context of chronic conditions which may require life-long treatment, and hence, persistent transgene expression. To increase the safety aspect of retroviruses, these vectors have been engineered to be replication-incompetent, which avoids the complication of possible viral proliferation and subsequent infection of adjacent cells. After establishing that retroviral-mediated correction of the CFTR defect was possible in cell culture using CF airway epithelial cells (Drumm et al., 1990), further experiments performed in a rat xenograft model of fully differentiated human epithelial cells showed that retroviruses were inefficient in transducing fully differentiated, non-dividing airway epithelial cells (Engelhardt et al., 1992). The reason for the relatively low number of transduced cells may be due to the requirement of active cell division at the time of transduction with retroviruses (Miller et al., 1990).

The use of integrating viral vectors is associated with the risk of genotoxicity and insertional mutagenesis. Two separate gene therapy studies for the treatment of X-linked severe combined immunodeficiency (SCID-X1) showed that vector-mediated up-regulation of host cellular oncogenes led to uncontrolled clonal expansion of mature T-cells, which then led to the development of leukaemia (Hacein-Bey-Abina et al., 2003a; Howe et al., 2008). The onset of these severe adverse events was thought to be due to integration of the retroviral provirus close to the *LMO2* proto-oncogene, which exerted an enhancer activity on the *LMO2* promoter (Hacein-Bey-Abina et al., 2003b). To reduce the potential for insertional mutagenesis, self-inactivating (SIN) retroviral vectors have been developed (Thornhill et al.,

2008). Human clinical trials assessing the safety and efficacy of these modified vectors are currently underway (Hacein-Bey-Abina et al., 2014).

Lentiviruses are members of the retrovirus family and recombinant lentiviral (rLV) vectors, based on the human immunodeficiency virus (HIV) and non-human lentiviruses, have been developed. Unlike the gamma-retroviruses described above, rLV is able to transduce non-dividing cells (Naldini et al., 1996), as well as non-dividing epithelial cells from a human bronchial xenograft (Goldman et al., 1997). Lentiviral vectors are able to integrate into the host genome, which may be an advantage, but may also be considered a safety risk due to potential genotoxicity. To reduce the risk of insertional mutagenesis, rLV vectors have been developed in which the integrase protein has been mutated to reduce the efficiency of integration (Vargas et al., 2004). These vectors have been termed integrase-deficient lentiviral vectors (IDLVs). The D64V integrase mutant is the most commonly used mutation for generating IDLVs. This mutant yields integration events close to background levels when compared with wild-type lentivirus (Apolonia et al., 2007). Such IDLVs have been investigated for gene transfer into the lung (Pringle et al., 2014). However, transgene activity from the IDLV vectors in the lung was five- to ten-fold lower when compared with integrase-competent vectors (Pringle et al., 2014).

Recombinant retroviruses can be further modified to increase the affinity for specific cell types. The process, known as pseudotyping, is performed by replacing the native envelope glycoprotein with one of a different enveloped virus, potentially changing the tropism and transduction properties of the vector. For example, the most common glycoprotein used to pseudotype rLV is the vesicular stomatitis glycoprotein envelope (VSV-G), which increases the vector tropism to a wide range of tissue and cell types (Naldini et al., 1996). Unfortunately, VSV-G-pseudotyped rLV vectors are unable to transduce fully-differentiated, non-dividing airway epithelial cells (Johnson et al., 1992) without the addition of tight-junction openers (Limberis et al., 2002). This may be explained by the lack of heparin sulfate

proteoglycans on the apical cell surface, which have been shown to be involved with the binding of the VSV-G protein (Copreni et al., 2008).

To improve transduction of proximal airway epithelial cells, envelope glycoproteins (GP) derived from viruses that target receptors found on the apical surface of airway cells have been exploited. Glycoproteins isolated from the filovirus family, such as the Marburg virus and the Zaire strain of the Ebola (EboZ) virus, have been used to pseudotype HIV-derived vectors (Chan et al., 2000). Subsequent studies demonstrated that an HIV-based vector pseudotyped with the EboZ viral GP was able to apically transduce both *ex vivo* human tracheal cells (Kobinger et al., 2001) and murine airway epithelial cells *in vivo* (Medina et al., 2003). Furthermore, lentivirus vectors derived from non-human primates, such as feline immunodeficiency virus (FIV)-derived vectors, were able to be successfully pseudotyped with both the Marburg and EboZ GP to transduce the apical surface of *ex vivo* human airway epithelial cells (Sinn et al., 2003). Unfortunately, the utility of using these filovirus-based pseudotypes has been limited by very low viral titers during production (Sinn et al., 2005), precluding their further investigation in clinical trials.

Another approach has exploited the natural tendency of the paramyxovirus family to infect airway epithelial cells. The fusion and hemagglutinin neuraminidase envelope proteins (F/HN) of the Sendai virus (SeV, also known as murine parainfluenza virus type I) have been used to pseudotype a simian immunodeficiency virus (SIV)-based lentiviral vector (Kobayashi et al., 2003). Further studies using a modified version of this vector demonstrated: a) transduction of polarised, differentiated *ex vivo* airway epithelial cells; b) effective transduction of murine nasal epithelial cells *in vivo*, leading to persistent expression up to 15 months after a single dose; and c) successful repeat administration to the mouse nose after three doses (Mitomo et al., 2010).

Overall, viral vectors have demonstrated that gene transfer to airway cells is both feasible and efficient. Indeed, rLV vectors in particular have been the focus of recent research due to a number of advantages, such as the ability to transduce non-dividing cells, as well as being able to mediate long-term transgene expression, and by being seemingly less immuno-stimulatory when compared with other viral vectors. Despite these advantages, rLV still have the theoretical risk of insertional mutagenesis and genotoxicity. Additionally, immune responses to the vector as well as the transgene product are possible. Therefore, concurrent work has continued to progress in the field of non-viral vector development for a range of gene therapy applications, including treatment of lung diseases.

### **1.2.7. Non-viral gene delivery systems**

Despite their relative inefficiency, non-viral vectors offer a number of potential advantages over viral approaches, including: a) an improved safety profile due to a reduced risk of toxicity, immunogenicity or oncogenicity; b) no observable limit to packaging; c) more amenable to pharmaceutical formulation due to its relative stability; and d) demonstrable clinical evidence of successful repeat administration. Research has led to the development of numerous non-viral gene delivery systems, including the administration of naked pDNA or using physical or chemical methods to increase gene transfer. These methods demonstrated various levels of success in mediating efficient transfection into target host cells of the airway epithelium.

### **1.2.8. Naked DNA**

Non-viral pDNA vectors consist of covalent closed circles of double-stranded pDNA generated in bacteria. In airway cells, delivery of naked pDNA was shown to be effective in transfecting mouse airway epithelial cells via intratracheal delivery, without the need for damaging the epithelium (Meyer et al., 1995), although in the same study, expression was higher if the pDNA was complexed with cationic liposomes (Section 1.2.12). This may have been due to inefficient uptake of naked pDNA due to its large size and hydrophilic negatively

charged phosphate groups (reviewed in Al-Dosari and Gao, 2009). The use of naked pDNA for nebulisation into the lung is generally ineffective due to shearing by aerosolisation (Crook et al., 1996), although electrohydrodynamic (EHD) aerosol generation delivered intact naked pDNA to the lungs of mice (Davies et al., 2005). Despite this encouraging result, the EHD device has not yet been commercialised for use in the clinic.

### **1.2.9. Physical methods to increase transfection efficiency**

Physical methods that create a transient pore in the cell membrane, thus facilitating the transit of plasmid molecules into the cell, have also been investigated. Electroporation uses electric fields to promote gene transfer of naked DNA into target cells; typically, the animal receives a dose of plasmid followed by a number of electrical pulses from two electrodes that are applied to the tissue. The use of electroporation has been investigated in mice (Dean et al., 2003; Kaufman et al., 2010; Pringle et al., 2007) and rats (Gazdhar et al., 2006; Machado-Aranda et al., 2005), demonstrating that electroporation can be safe and well tolerated depending on the electrical strength and frequency of the pulse. Gene transfer was observed throughout the lung at levels higher than naked pDNA alone, but transgene expression was maximal in alveolar cells, which may limit the utility for CF gene therapy. In order to demonstrate clinical relevance, large animal models that are closer in size, anatomy and physiology to human, have also been investigated, including sheep (Pringle et al., 2007) and pig lungs (Dean et al., 2011). Other physical methods including both low- (Xenariou et al., 2010) and high-frequency (Xenariou et al., 2007) ultrasound (sonoporation) have been tested *in vivo* demonstrating increased transgene expression after delivery of naked pDNA. The use of magnetic forces (magnetofection) generated variable results when using naked pDNA in the mouse lung (Dames et al., 2007; Xenariou et al., 2006). Translating such physical methods into clinically relevant procedures for CF lung gene transfer has not been pursued, although they may be useful for some other specific disease applications.

### **1.2.10. Chemical methods to increase transfection efficiency**

Chemicals have been used in conjunction with pDNA in an attempt to mimic the processes that allow a virus to efficiently infect mammalian cells; the most commonly used chemicals include cationic lipids and polymers resulting in plasmid compaction to aid in transport into the cells and to protect pDNA from degradation by nucleases or shear forces generated during aerosolisation. The combination of pDNA with cationic lipid to form lipoplexes, and with polymers to form polyplexes, is relatively straightforward. The formation of stable complexes is driven by electrostatic interactions between the positively charged chemical and the negatively charged DNA (Tros de Ilarduya et al., 2010). Transfection efficiency needs to be optimised based on the method of delivery and cell type. Typically, conditions such as size of the complex, the ratio of cationic vector to pDNA molecules and net ionic charge are investigated to achieve maximal levels of transfection. Although the complete mechanism of lipo- or polyplex entry into the cell and subsequent transcription is not fully understood, the generally-accepted model includes: a) binding of the cationic complex to the anionic cell membrane via electrostatic interactions, b) uptake into the cell via endocytosis or endocytosis-like mechanisms, c) retention of the complex in perinuclear endosomes, d) release of the complex from the endosome, e) dissociation of the complex from the cationic carrier, and f) entry of the DNA into the nucleus (reviewed in Elouahabi and Ruyschaert, 2005). One further consideration for gene delivery to the airways is that aerosolised non-viral formulations must be resistant to degradation due to shear forces that are generated during nebulisation.

### **1.2.11. Cationic polymers**

Chemicals such as polyethylenimine (PEI) and poly-L-lysine (PLL) are examples of frequently used cationic polymers. One theory explaining the release of PEI/pDNA complexes from the endosomal compartment involves the 'proton sponge' effect, where proton accumulation in the endosome buffers the low pH, and leads to endosomal swelling, rupture, and the release of the polyplex into the cytoplasm (Boussif et al., 1995). Nuclear

entry of the pDNA is not fully understood, but may involve passive DNA entry into the nucleus during cell division, or active transport through the nuclear pore either by DNA sequence-dependent transport, or by nuclear localisation signals mediated by transcription factors in the cytoplasm (see Section 1.4.4) (Elouahabi and Ruyschaert, 2005).

For airway gene transfer, PEI has received considerable attention, as it has been shown to be effective at mediating high levels of transgene expression both in cell culture and *in vivo* (Boussif et al., 1995). However, the effectiveness of different forms of this cationic polymer appears to be dependent on the mode of delivery. For example, Wiseman *et al.* demonstrated that the linear form of 22 kDa PEI leads to higher reporter gene expression after instillation when compared with branched 25 kDa PEI (Wiseman et al., 2003). Conversely, in nebulisation studies, branched 25 kDa PEI mediated higher levels of transgene activity when directly compared with both 22 and 25 kDa linear PEI (Rudolph et al., 2005). These results indicate that the molecular structure, such as the degree of branching and molecular weight (Rudolph et al., 2005), as well as route of administration are important to assess in such studies. Additionally, it has been shown that molecular structure and method of administration contribute to the toxicity of the polyplex in mice. Both tracheal (Rudolph et al., 2000) and nasal (Davies et al., 2012; Davies et al., 2008) instillation of 25 kDa PEI polyplexes to mice resulted in severe immune cell infiltration or increased levels of pro-inflammatory cytokines, whereas aerosol administration did not produce any histological changes in the mouse lung (Gautam et al., 2000), even after ten aerosolised doses (Davies et al., 2012) or when concentrated polyplex formulations containing over 8 mg/mL of pDNA were administered (Davies et al., 2008). Interestingly, repeat administration of polyplexes did not necessarily result in high levels of sustained transgene expression. However, by increasing the dosing interval it was possible to avoid the effects of the refractory period following dosing (Davies et al., 2012). Despite these advances in aerosolisation, the use of PEI has not been evaluated in the clinic for airway gene transfer. This is due to its non-degradable nature, which can cause toxic side effects (Oliveira et al., 2014); however,

degradable derivatives of PEI are currently being pursued in order to circumvent such issues (Islam et al., 2014).

### **1.2.12. Cationic lipids**

Cationic liposomes are another class of chemicals utilised in airway gene transfer protocols. Liposomes are composed of three basic domains: 1) a positively charged polar head group which binds to the negatively charged DNA, 2) a hydrophobic anchor to allow the complex to cross the cell membrane, and 3) a linker which joins the polar and non-polar regions. The mixing of liposomes and pDNA in an aqueous solution results in the formation of lipoplexes. Neutral lipids, such as DOPE (1,2-dioleoyl-sn-glycero-3-phosphoethanolamine), are added to increase transfection efficiency. Although the exact reasons for improved gene transfer are unknown, it is thought that the addition of DOPE may increase the fusogenic ability of the liposome in the processes of cell uptake, endosomal escape through destabilisation of the endosomal membrane, and dissociation of the lipid from the pDNA (Fasbender et al., 1997a).

The efficacy and safety of lipoplexes has been investigated in a number of clinical trials for treatment of CF. Of particular interest was the dosing of pDNA and DC-Chol/DOPE (3b-[N-(N',N'- dimethylaminoethane)carbamoyl]cholesterol:dioleoylphosphatidyl- ethanolamine) liposomes to the nasal epithelium of CF patients, whereby repeat administration was demonstrated without apparent loss of efficacy after subsequent doses (Hyde et al., 2000). However, transient transgene expression was observed. Further investigations into lipoplex formulations led to the development of cationic lipids that were more effective at gene transfer and particularly effective after aerosol delivery. A systematic analysis of novel cationic lipids were tested after instillation to the mouse lung, which identified the DC-cholesterol analogue Genzyme Lipid 67 (GL67; Cholest-5-en-3-ol (3P)-,3-[(3-aminopropyl)[4-[(3-aminopropyl)amino]butyl] carbamate]) as a promising candidate lipid. In mouse studies, GL67/DOPE mediated gene transfer 100-fold higher compared with early formulations based on DC-cholesterol/DOPE; importantly, in contrast to rAdV and rAAV

vectors, GL67/DOPE was also able to be repeatedly administered without a reduction in transgene expression (Lee et al., 1996). Dose-dependent inflammation was observed after instillation into the mouse lung which was due to the lipids, rather than the pDNA, although the combination of GL67/DOPE and pDNA led to the highest inflammatory response (Scheule et al., 1997). Interestingly, a decrease in lung toxicity was observed when GL67/DOPE lipoplexes were aerosolised to mouse lungs when compared with instillation (Eastman et al., 1997b). Further improvements required the inclusion of polyethylene glycol-containing lipid (DMPE-PEG<sub>5000</sub>; 1,2-Dimyristoyl-sn-Glycero-3-Phosphoethanolamine-N-[methoxy (Polythene glycol) 5000]), resulting in an optimised formulation suitable for concentrated aerosol delivery (GL67A; GL67/DOPE/DMPE-PEG<sub>5000</sub>) (Eastman et al., 1997a). Mice that received aerosolised GL67A showed no signs of an inflammatory response, whereas mice that received a 10-fold lower dose via instillation exhibited moderate levels of inflammation although equivalent levels of transgene expression were detected (Eastman et al., 1997a). This inflammation may be due to the fact that instillation results in pooling of the bolus in the distal lung rather than full-organ distribution as is the case with aerosolisation (Cheng and Scheule, 1998).

GL67/DOPE was delivered to the nasal epithelium in a clinical trial when complexed with pDNA containing *CFTR* cDNA, and compared with pDNA alone; surprisingly, the lipoplexes conferred only a small advantage over the pDNA alone on the basis of electrophysiological changes (Zabner et al., 1997). This study, in conjunction with a study in healthy patients (Chadwick et al., 1997), provided impetus for two clinical studies in CF patients assessing the safety of GL67A in the lungs. In the first study, safety of GL67A-containing lipoplexes was assessed compared with GL67A alone (Alton et al., 1999). In contrast to the result observed by Chadwick and colleagues where aerosol delivery of GL67A was well tolerated in healthy patients (Chadwick et al., 1997), the delivery of GL67A to CF patients was marked by mild influenza-like symptoms that resolved within 36 hours of administration along with mild airway symptoms (Alton et al., 1999). Similar results were observed in the group that

received pDNA complexed with GL67A (Alton et al., 1999). Encouragingly, correction of the basic electrophysiological defect was detected in patients that received the lipoplex (Alton et al., 1999). Similar results were observed in a second study where the side effects observed were attributed to lipid-DNA administration. The reasons for this clinical response are unknown but could be attributed to the processing of the lipoplexes through macrophages, or due the presence of unmethylated CpGs in the pDNA (Alton et al., 1999). This led to the development of pDNA completely devoid of CpGs, termed pG4-hCEFI-soCFTR2 (Hyde et al., 2008), which was evaluated in a Phase I/IIa study. A single aerosol of a low dose of the pG4-hCEFI-soCFTR2/GL67A lipoplex to the lungs of CF patients led to a negligible inflammatory response (Alton et al., 2013b) and resulted in electrophysiological correction in the nose for up to 13 weeks following a single dose to the nose (Davies et al., 2011), which is the longest reported correction of the underlying chloride channel defect in CF patients. In support of a human multi-dose clinical trial, pre-clinical toxicology aerosol studies with this formulation were performed in sheep (nine doses) (Alton et al., 2013a) and mice (12 doses) (Alton et al., 2014), demonstrating that systemic immune responses were not detected after repeat administration. Additionally, gene expression persisted for over 4 months after the final lipoplex dose in mice (Alton et al., 2014). This formulation has also been evaluated in a Phase IIb multidose trial; the patients in this trial received 12 monthly doses of the lipoplex to evaluate patient clinical benefit (Alton et al., 2013b). The results of this clinical trial confirm that repeat administration at monthly intervals is safe; furthermore, stabilisation of lung function was observed when compared with lung function in patients who received saline as placebo (Alton et al., 2015). While the overall improvement of the lipoplex formulation was modest in comparison with the placebo (3.7 % improvement in lung function), the results indicate a baseline from which to further improve clinical outcome. Further possibilities may include a higher dose of pG4-hCEFI-soCFTR2/GL67A, with a higher dosing frequency.

### 1.2.13. Other non-viral vectors for airway gene transfer

The final physical barrier to successful gene transfer in post-mitotic cells is the nuclear envelope. The nuclear pore complex allows the transport of molecules across the nuclear envelope, where small particles are able to pass by passive diffusion. The cationic lipids and polymers described thus far have a size in the order of 100 nm (Guo and Huang, 2011). Therefore, smaller gene transfer agents would be advantageous to increase the amount of pDNA that may be available for transcription in the nucleus.

Compacted nanoparticles have been developed which consist of a single molecule of pDNA compacted with PEG-substituted 30-mer lysine polymers, and which were shown to transfect post-mitotic cells by directly passing through the 25 nm nuclear membrane pore (Liu et al., 2003). Furthermore, instillation studies *in vivo* demonstrated high levels of transgene expression when compared with delivery of naked pDNA (Ziady et al., 2003a) with a favourable safety profile (Ziady et al., 2003b). This nanoparticle was evaluated in a human clinical trial after administration to the nasal mucosa (Konstan et al., 2004) where it was well tolerated. Electrophysiological correction was observed in eight of 12 treated subjects (Konstan et al., 2004). In a comparison with GL67A and PEI after aerosol delivery to the sheep lung, nanoparticles resulted in lower transgene expression than GL67A and PEI (McLachlan et al., 2011). The doses were not matched; however, the dose delivered for nanoparticles (40 mg) was similar to that of GL67A (52.8 mg) and five times higher than what was delivered with PEI (8 mg). No further clinical studies have been performed with nanoparticles.

Efforts to increase transfection efficiencies of non-viral vectors in the lung have led to a proliferation of new non-viral vectors, including the use of biologically derived products such as chitosan. Gene transfer mediated by a combination of chitosan with low molecular weight PEI was able to suppress lung tumourigenesis *in vivo* (Jiang et al., 2009). Exploitation of the

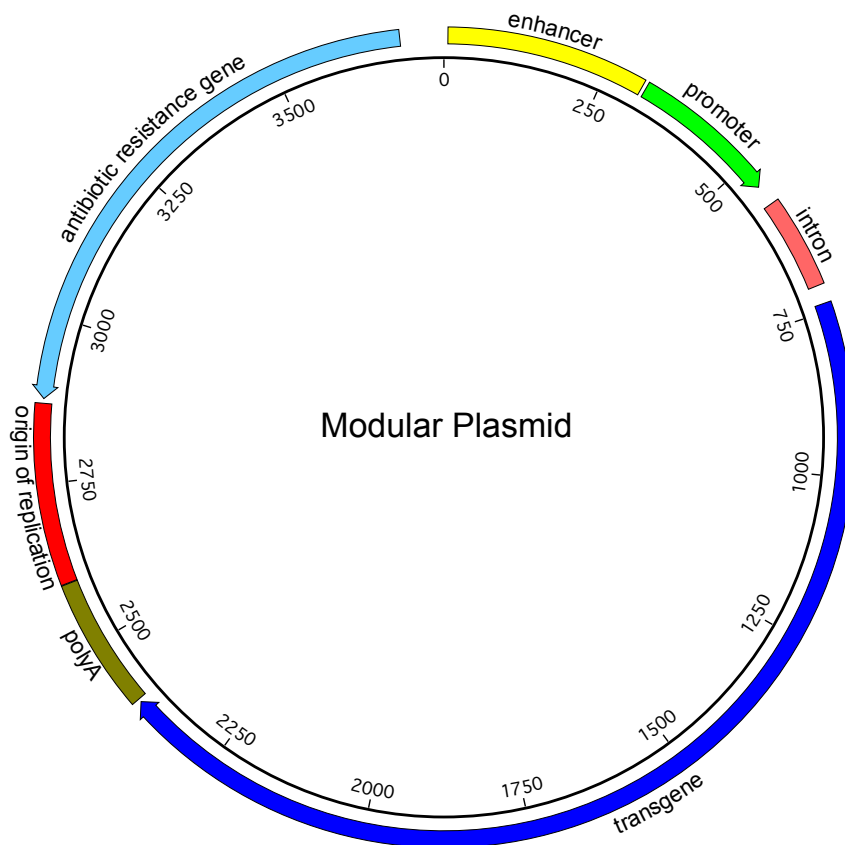
monolayer membrane organization observed in archaeobacteria have demonstrated proof-of-principle for the use of novel, synthetic archaeosomes after nasal administration *in vivo* (Le Gall et al., 2014). These studies are ongoing and acknowledge the current view that pursuing alternatives to viral gene transfer agents is important.

#### **1.2.14. Use of the murine lung as a model for CF gene transfer**

The mouse has been used as a model for airway gene transfer for delivery of both non-viral and viral vectors to the airways to investigate the persistence of gene expression for treatment of lung diseases. Several transgenic mouse strains with a range of *Cftr* mutations have been developed (reviewed in Wilke et al., 2011), although these mice do not develop overt lung disease, largely due to the presence of alternative chloride transport pathways in mouse epithelial cells (Clarke et al., 1994). These transgenic strains do, however, suffer from severe gastrointestinal disease due to the recapitulation of CF intestinal disease and must be fed special diets (Grubb and Boucher, 1999). The expression of human *CFTR* under the control of specific promoters has been shown to correct the intestinal defect observed in the transgenic strains in order to increase survival rate (Zhou et al., 1994). Other animal models have been developed, including the pig (Rogers et al., 2008) and ferret (Sun et al., 2010), both of which more closely model CF lung disease (Yan et al., 2015a). Anatomical differences between humans and these animal models must also be considered. For example, submucosal glands in the mouse are present only in the upper third of the trachea and larynx, whereas in humans, they are found throughout the trachea and bronchi (Liu and Engelhardt, 2008). Nevertheless, as a model for assessing persistence and level of transgene activity after aerosolisation, BALB/c mice have recently been used to support the progression of a non-viral vector into a Phase IIb trial in CF patients (Alton et al., 2014) and thus represents a useful model to further investigate the reasons for transgene expression.

### 1.3. Plasmid design for long-term transgene expression

Achieving sustained transgene expression levels has been a focus for the development of gene therapy for chronic diseases. Whereby transient transgene expression may be beneficial for acute conditions such as cancer, chronic conditions such as CF will require long-term expression due to cell-turnover in the lung (Section 1.1.2). Plasmid DNA is thought to remain in an episomal state within the nucleus. However, the early loss of transgene expression is unlikely to be due to cell turnover and dilution of the plasmid through rounds of cell division. Pivotal studies demonstrated that judicious promoter selection for plasmid gene expression is critical to mediating persistent expression in the mouse lung and also that loss of expression was not simply due to immune-mediated destruction of transfected cells (Gill et al., 2001; Yew et al., 2001). Further studies showed that loss of transgene expression could not solely be attributed to the degradation or loss of pDNA in transfected cells after instillation of naked DNA or lipoplexes *in vivo* (Pringle et al., 2005). Interestingly, studies in the mouse lung suggest that silencing of transgene activity occurs in the absence of *de novo* methylation (Pringle et al., 2005). Further investigations into plasmid promoters have demonstrated the importance of promoter selection for airway gene transfer (Hyde et al., 2008; Pringle et al., 2012a). Therefore, the overall design and selection of constituent elements within the pDNA (Figure 1.2) is key to generating long-term, high-level transgene expression.



**Figure 1.2. Diagram of a generic modular plasmid.**

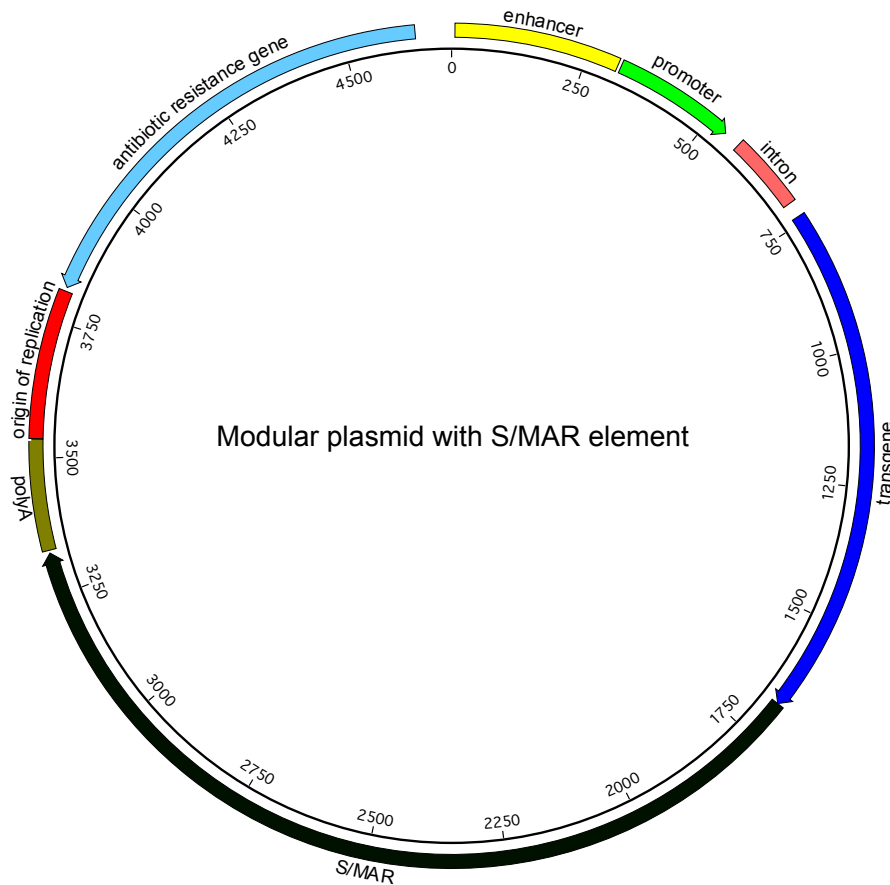
A conventional non-replicating plasmid can be divided into the expression cassette and the bacterial backbone. The expression cassette contains all sequences required for expression of the transgene (dark blue), including an enhancer (yellow) and promoter (light green) to increase the rate of transcription, intronic sequences (pink) to enhance splicing efficiency and a polyadenylation site (dark green) for mRNA stability. The bacterial backbone is solely required for production of the plasmid and contains a bacterial origin of replication (red) and antibiotic resistance gene (light blue).

### 1.3.1. Self-replicating vectors

In some situations, transient expression from non-viral vectors may be the result of cell turn-over, whereby the extra-chromosomal pDNA is lost as it is not passed on to daughter cells. One strategy to circumvent this is to incorporate a scaffold matrix attachment region (S/MAR) into the plasmid (Figure 1.3). This generates a class of self-replicating, non-integrating vectors which are maintained as a functional extra-chromosomal (episomal) entity (Argyros et al., 2008), replicating during cell division via the endogenous cellular replication machinery (Bode et al., 2001). These regions are ubiquitously present in all eukaryotic genomes; thus, the natural properties of a S/MAR have been exploited to increase the likelihood of persistent transgene expression of a synthetic DNA system. A S/MAR-based vector has been shown to co-segregate with chromosomes during mitotic division by anchoring the chromatin to the S/MAR in Chinese hamster ovary (CHO) cells (Jenke et al., 2002). Another experiment in CHO cells showed that pDNA containing an S/MAR element was capable of replicating at low copy number while remaining episomal (Piechaczek et al., 1999); further investigation demonstrated that replication occurred once per cell cycle concomitantly with cellular DNA during interphase (Schaarschmidt et al., 2004). Additionally, S/MARs may play a role in regulating transcription by facilitating the access of enhancers and transcription factors (Argyros et al., 2011).

Several studies have investigated the use of S/MAR-containing plasmids *in vivo*. Genetically-modified pig fetuses were produced with a S/MAR vector which was delivered to embryos using the sperm-mediated gene transfer method, which resulted in a reporter gene being expressed in all tissues in nine of 12 genetically-modified fetuses (Manzini et al., 2006), indicating that the S/MAR plasmids were established as actively replicating episomes. After hydrodynamic tail vein (HTV) injection of a plasmid driven by the tissue-specific AAT promoter containing an S/MAR element, long-term transgene expression was detected. However, transgene activity did not persist after partial hepatectomy performed 3 weeks after

dosing, suggesting that the vectors did not replicate and remained in a passive episomal state (Argyros et al., 2008). Indeed, the reason for persistence may partially be attributed to the choice of promoter, since a similar plasmid containing the human CMV promoter led to transient transgene expression (Argyros et al., 2008). However, the AAT-containing plasmid in the absence of S/MAR also led to very low levels of transgene activity; the authors noted that the addition of the S/MAR element into an AAT-containing plasmid resulted in a small degree of protection from methylation, and subsequent promoter attenuation, although the S/MAR does not confer this degree of methylation-blocking to the human cytomegalovirus (hCMV) promoter (Argyros et al., 2008). This result is in conflict with other studies in the liver (Chen et al., 2008) and lung (Pringle et al., 2005) which demonstrated that silencing is independent of *de novo* methylation. Altogether these data further highlight that the choice of promoter is critical for mediating long-term transgene expression.



**Figure 1.3. Diagram of a generic modular plasmid containing an S/MAR element.**

Diagram of a plasmid containing an S/MAR element (shown in black). The S/MAR element must be placed after the transcription unit in order for episomal replication to take place (Stehle et al., 2003).

### 1.3.2. Integrating vectors

Integration of pDNA into the host genome is an alternative strategy to prolong expression of the therapeutic gene after delivery to actively dividing or stem cell populations. Naturally occurring genomic elements called transposons have been exploited in the development of new plasmid-based integrating vectors. These include the *Sleeping Beauty*, *piggyBac* and phiC31 systems (reviewed in Gersbach and Barbas, 2013). Both the *Sleeping Beauty* and *piggyBac* systems utilise mobile elements that move via a DNA intermediate (transposon) to mediate transgene integration. The transposon system consists of two components: a) a transposon containing a gene-expression cassette flanked by inverted terminal repeats, and b) a source of transposase enzyme (Essner et al., 2005). The transposase can be expressed on the same (*in cis*) or different (*in trans*) plasmid. The transposase catalyzes the excision of the transposon from the plasmid and promotes its integration into the host genome using a precise cut-and-paste mechanism, targeting random TA dinucleotides for *Sleeping Beauty* (Ivics et al., 1997) or TTAA tetranucleotides when using *piggyBac* (Wilson et al., 2007).

Studies assessing transgene expression from both transposon systems in the mouse lung have been performed. After intravenous administration of a plasmid complexed with PEI, the *Sleeping Beauty* system mediated transgene activity in alveolar epithelial cells up to 3 months post-dosing (Belur et al., 2003; Liu et al., 2004). The *piggyBac* system yielded similar results although the transfected cell type was not identified (Saridey et al., 2009). However, as intravenous delivery is well known to deliver pDNA to the lung endothelium (Goula et al., 1998) it is likely that the majority of transfected cells were in the alveolar region rather than the airway. Therefore, the utility for such a system to treat CF is limited. Also, the observed random integration into the host cell genome with these vectors does not confer an advantage over integrating retroviral or lentiviral vectors (Section 1.2.6). Therefore a controlled, stable, unidirectional site-specific system may be preferable in order to reduce safety concerns regarding random integration of vector DNA into genomic sites that could

induce genotoxic or oncogenic events. One such system is the phiC31 integrase, a site-specific recombinase enzyme where integration occurs via the recombination of the *attB* recognition site on a plasmid with a short DNA sequence in the human genome (pseudo-*attP* sites) (Thyagarajan et al., 2001). Further analysis of the recognition sites (30 to 40 bp long) in the human genome identified approximately 100 integration sites in three human cell lines (Chalberg et al., 2006). Studies in the mouse lung after intravenous delivery of phiC31-containing plasmids complexed with PEI resulted in low-level luciferase activity at 2 weeks, but only at very low levels when compared with day 1 values (Aneja et al., 2007). For integrating vectors to be useful for treating CF, expression is required in the airway epithelium. Delivery of naked pDNA incorporating *Sleeping Beauty* transposase via nasal instillation, a delivery route previously shown to result in airway epithelial transgene expression (Hyde et al., 1993), led to persistent transgene expression for up to 6 months after dosing, but only when combined with the UbC promoter (Lawton et al., 2002). This study indicated that the choice of promoter was more critical than vector integration for persistent transgene expression, at least up to 6 months.

### **1.3.3. Minimal plasmid DNA vectors**

The bacterial backbone in a conventional plasmid is solely required for plasmid manufacture and includes a bacterial origin of replication and a growth selection marker, such as an antibiotic resistance gene (Figure 1.2). Although the use of antibiotic resistance genes, such as kanamycin, is approved for clinical use, selection markers are generally discouraged due to a risk of inadvertent dissemination of the resistance gene in human somatic cells and the environment (reviewed in Vandermeulen et al., 2011). To this end, a class of plasmids and bacterial producer cells have been developed which do not require the presence of an antibiotic resistance gene for plasmid maintenance and propagation. The antibiotic-free system, named pFAR (plasmids free of antibiotic resistance markers), led to approximately 1-log higher transgene expression levels after plasmid mass-matched electrotransfer into mouse skin or tumours when compared with a similar vector containing the kanamycin

resistance gene (Marie et al., 2010). A further study investigated transgene expression in mice after HTV injection of equimolar amounts of pFAR plasmids and conventional plasmids (containing a kanamycin resistance gene) harbouring either the CAG (CMV enhancer/chicken  $\beta$  actin) promoter, or the liver-specific AAT promoter (Quiviger et al., 2014). The pFAR plasmids resulted in approximately 500-fold higher reporter protein levels 4 months after dosing, but only when the liver-specific promoter was used (Quiviger et al., 2014). Again, the importance of promoter selection is reinforced by these studies.

In addition to the removal of the antibiotic resistance gene, removal of the bacterial origin of replication has also been investigated. The advantages, beyond the biosafety concerns described above, include improved transgene expression due to: 1) an increased gene delivery efficiency due to decreased plasmid size, 2) a reduction in epigenetic silencing, and 3) reduced risk of triggering an inflammatory reaction to the transgene product (reviewed in Oliveira and Mairhofer, 2013). Minimalistic, immunological-defined gene expression (MIDGE) vectors are minimal size gene transfer units consisting of the expression cassette, flanked by two short hairpin oligonucleotide sequences which results in a linear, covalently closed, dumbbell-shaped molecule (Schakowski et al., 2001). After HTV injection of naked MIDGE vectors into mice, transgene expression was significantly higher in the liver, lung, kidney and heart when compared with equimolar amounts of the corresponding conventional pDNA vector (Schakowski et al., 2007). These results correlated with the number of vector copies per cell detected, indicating that the smaller size of MIDGE vectors could result in improved transfection rates; however, other reasons, such as the relatively low content of CpG sequences, have not been discounted (Schakowski et al., 2007).

Small supercoiled minicircle DNA vectors have been developed which are also devoid of bacterial backbone sequences. Proof-of-principle was established by Darquet *et al.* whereby two to 10-fold higher luciferase reporter gene activity was observed in cell culture studies after transfection with minicircles when compared with conventional pDNA (Darquet et al.,

1997). Several *in vivo* studies in the lung (Vaysse et al., 2006), heart (Huang et al., 2009) and liver (Chen et al., 2003) have further confirmed the results observed in cell culture. These minicircle studies have suggested that increased transgene expression may be due to: a) smaller vector size, thus allowing for more efficient transfection and subsequent nuclear entry, and/or b) the removal of the prokaryotic backbone which traditionally contains CpG sequences recognised by the innate immune system which could lead to initiation of the inflammatory cascade and subsequent transgene silencing (Darquet et al., 1997; Vaysse et al., 2006). Intriguingly, transcriptional silencing in the liver after HTV injection was CpG-independent (Chen et al., 2008). The accumulation of these data led to the development of the hypothesis that transgene silencing is due to the formation of repressive heterochromatin on the bacterial backbone which then spreads throughout the plasmid, which in turn attenuates promoter function (Chen et al., 2008). Minicircle DNA has also been evaluated following aerosol delivery to the mouse lung; however, the use of minicircle DNA did not confer any advantage when compared with conventional CpG-free plasmids complexed with PEI (Bazzani *et al.*, manuscript in preparation). This may indicate that the use of minicircle DNA may be advantageous when depletion of CpGs is not possible, such as in the case of critical sequences required for transcription factor binding to mediate active transcription.

#### **1.3.4. CpG-depleted plasmids**

The presence of unmethylated CpGs has been implicated in transgene silencing after viral (see Section 1.2.5) and non-viral (see Section 1.3.1) vector delivery. Comparison of mammalian and bacterial DNA has shown that CpG frequency in mammalian DNA is ~1:64 while bacterial DNA has the expected frequency of ~1:16 (Krieg et al., 1995). Furthermore, the methylation patterns are species-specific; for example, bacterial CpGs are essentially unmethylated (Krieg et al., 1995) while in mammals, approximately 70 to 80 % of CpG sequences are methylated (see Section 1.4.3) (Antequera and Bird, 1993). Together, these differences combine to make bacterial and mammalian DNA structurally distinct. The

mammalian innate immune system takes advantage of this distinction and recognises unmethylated CpGs as foreign entities (Krieg et al., 1995; Scheule, 2000). CpGs activate macrophages and dendritic cells to express increased levels of pro-inflammatory cytokines and co-stimulatory molecules (Krieg and Kline, 2000). This effect is termed the CpG response and is mediated by Toll-like receptor (TLR) 9, the innate immune receptor for unmethylated CpGs (Hemmi et al., 2000). The culmination of the TLR9 signalling pathway results in the expression of genes encoding for pro-inflammatory cytokines such as  $\text{TNF}\alpha$ , IL-6, IL-12 and type I interferons (IFNs) (Hemmi et al., 2000). Additional studies have also identified the production of other pro-inflammatory cytokines, such as  $\text{IFN}\gamma$ , in response to stimulatory CpGs (Klinman et al., 1996). When neutralising antibodies against  $\text{TNF}\alpha$  and  $\text{IFN}\gamma$  were administered, enhanced levels of gene expression in the mouse lung were observed (Li et al., 1999), thereby further implicating the role of pro-inflammatory cytokines in the loss of transgene expression.

Following delivery of lipid/pDNA complexes to the mouse airways via nasal instillation (Scheule et al., 1997; Yew et al., 1999b), intratracheal administration (Freimark et al., 1998) or intravenous administration (Li et al., 1999), it was cationic complexes that resulted in the greatest induction of pro-inflammatory cytokines, rather than the administration of the cationic lipid or pDNA alone. Furthermore, Freimark *et al.* demonstrated that lipoplexes containing methylated plasmids, or plasmid fragments lacking in CpGs, did not induce high levels of cytokine production (Freimark et al., 1998). Clinical studies in human subjects also demonstrated a synergistic effect of lipid and pDNA on the production of pro-inflammatory cytokines after aerosol delivery to the lung. In the case of GL67A, the best characterised lipid used for airway gene transfer, investigators focussed on the removal of CpGs from pDNA in order to reduce inflammation, and in turn, improve transgene expression. Proof-of-concept for the use of CpG-reduced plasmids resulting in decreased levels of inflammation was demonstrated in a pivotal study performed by Yew and colleagues (Yew et al., 2000). A

significant reduction in pro-inflammatory cytokines was observed after nasal instillation or intravenous administration of CpG-reduced plasmids (containing 256 CpGs) complexed with cationic lipids when compared with its CpG-rich counterpart (containing 526 CpGs) (Yew et al., 2000). Interestingly, reporter gene expression increased two-fold when the CpG-reduced plasmid was administered, but it was not possible to confirm whether this result was due to decreased cytokine levels or to the use of codons optimal for expression in mammalian cells in the synthetic reporter transgene (Yew et al., 2000). The majority of the CpGs that were eliminated in this study were located in the transgene and the kanamycin resistance gene. Further attempts to produce a plasmid completely devoid of CpGs focused on developing synthetic promoters, based on the hCMV promoter (Yew et al., 2002) and the elongation factor 1 $\alpha$  (EF1 $\alpha$ ) promoter (Hyde et al., 2008). Also, a CpG-free form of the R6K origin of replication was developed which supported pDNA replication in *E. coli* containing the R6K *pir* gene product (Wu et al., 1995).

The step-wise reduction of CG dinucleotides in pDNA resulted in a plasmid completely devoid of CpGs, which mediates long-term, high-level transgene expression with minimal inflammation after aerosolisation to the mouse lung with cationic lipids (Hyde et al., 2008) or polymers (Davies et al., 2008). Interestingly, the removal of CpGs from the hCMV enhancer/promoter or enhancer-free EF1 $\alpha$  promoters was not sufficient to mediate long-term transgene activity in the lung after aerosol delivery (Hyde et al., 2008). However, by combining a synthetic, CpG-free, version of the human cytomegalovirus enhancer (hC) (Yew et al., 2002) and a synthetic, CpG-free, version of the human elongation factor 1 $\alpha$  promoter (EF), persistent high-level transgene activity was observed in the mouse lung (see Section 1.3.5) (Hyde et al., 2008). The safety profile, as measured by levels of acute pro-inflammatory cytokines, was also improved. Interestingly, complete CpG removal was essential for these features, as retention of even a single CpG in a plasmid complexed with a

cationic lipid was sufficient to generate an inflammatory response after nasal instillation to the mouse lung (Hyde et al., 2008).

A number of naturally occurring CpG-free promoters have also recently been identified. The most promising in terms of persistence of transgene expression was the promoter of the murine fatty acid binding protein 4 (mFABP) gene. After aerosol delivery to the mouse lung, transgene expression levels of a plasmid containing the hC enhancer and the mFABP promoter were similar to the expression from the hCEFI enhancer/promoter 28 days after administration (Pringle et al., 2012a). Interestingly, neither the hCEFI nor hCmFABP enhancer/promoter combinations mediated persistent transgene expression in the BALB/c mouse liver after HTV injection (Pringle et al., 2012a), indicating that the route of administration and target tissue are important considerations when selecting a suitable promoter sequence.

A previous study in which the effect of CpGs in the transgene and plasmid backbone was systematically investigated showed that the CpG content in the transgene was more important than the plasmid backbone in adversely affecting transgene persistence in the mouse lung (Bazzani *et al.*, manuscript in progress). Additionally, transgene persistence was not affected after aerosolisation to the lungs of mice deficient in the TLR9 (TLR<sup>-/-</sup>) receptor for unmethylated CpGs, suggesting that silencing of CpG-rich transgene expression in the lung is independent of TLR9 (Bazzani *et al.*, manuscript in progress). A possible alternative pathway for this CpG-dependent, yet TLR9-independent, silencing effect is via the production of interferons in response to double-stranded DNA, although the mechanism has not yet been characterised (Ishii and Akira, 2006).

Despite the removal of all CpGs from plasmid vectors, residual low-level inflammation after dosing of lipoplexes (Bazzani et al., 2011; Hyde et al., 2008) or polyplexes (Davies et al., 2012) to the lung has still been observed. GL67A alone was capable of increasing several

cytokines, and additional cytokines were elevated, even in the presence of CpG-free lipoplexes (Bazzani et al., 2011). However, a residual inflammatory response was observed in TLR9<sup>-/-</sup> mice instilled with CpG-free lipoplexes when compared with mice instilled with WFI only, supportive of an acute, mild inflammatory cytokine response evoked in a TLR9- and CpG-independent manner (Bazzani et al., 2011). It appears that TLR9 contributes to the majority, but not the entirety, of the inflammatory responses associated with lipoplex delivery (Zhao et al., 2004b). Instead, the acute response can be partially attributed to unmethylated CpGs present in the residual bacterial chromosomal DNA (chrDNA) present after the plasmid production process, whereby pDNA enzymatically-treated to remove residual chrDNA resulted in lower levels of pro-inflammatory markers in the mouse lung (Bazzani et al., 2011). A similar observation was made after delivery of naked pDNA by hydrodynamic limb vein injection to mouse muscle, where muscle damage was attributed to relatively high levels (5 %) of contaminating bacterial chrDNA (Wooddell et al., 2010). These studies demonstrate that a significant decrease in inflammation can be achieved with a reduction in CpGs, not only from the pDNA, but also by removal of contaminating levels of bacterial chrDNA.

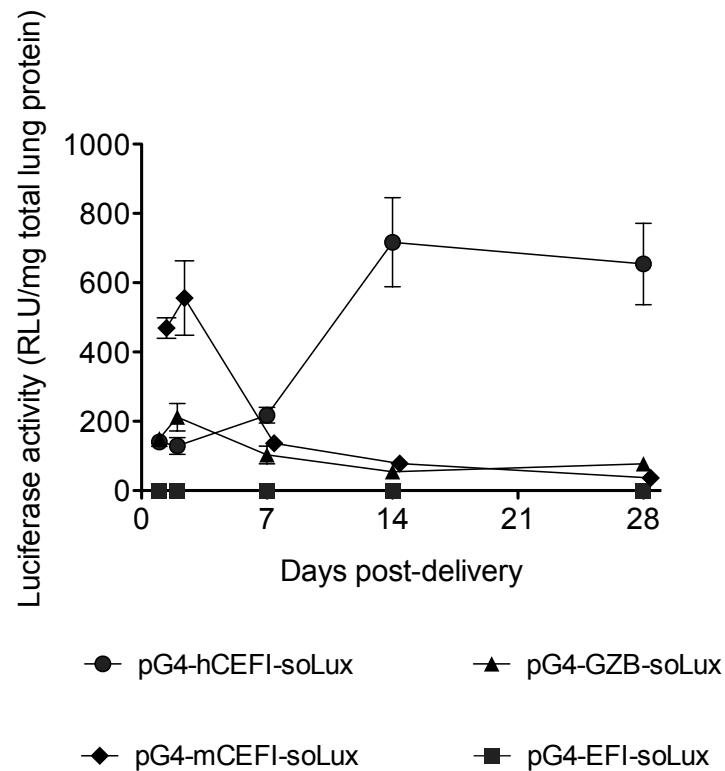
### **1.3.5. Enhancer/promoter constructs**

Strong viral promoters have traditionally been used in gene transfer studies leading to acute high-level gene expression. For example, the hCMV promoter, one of the most widely used promoters, has been used in 34 % of gene therapy trials conducted between 1991 and 2011 (Pringle et al., 2012a). Promoter attenuation, and subsequent silencing of transgene expression, has been attributed to the use of such viral promoters due to the up-regulation of the pro-inflammatory cytokines IFN $\gamma$  and TNF $\alpha$  (Qin et al., 1997). Interestingly, the silencing effect occurred at the mRNA level, did not cause vector DNA degradation or inhibition of the protein synthesis, nor destruction of transfected cells, and did not affect the constitutive cellular promoter from the  $\beta$  actin gene (Qin et al., 1997). Additional pivotal studies demonstrated the importance of judicious promoter selection in mediating long-term transgene activity. In addition to hCMV, other viral promoters (RSV and SV40) were also

shown to mediate transient transgene expression, lasting for only 1 week after dosing pDNA (Gill et al., 2001). This study also demonstrated that the loss of gene expression was not due to destruction of the transgene-expressing cells, nor due to preferential degradation of the plasmid after entering the cells (Gill et al., 2001). However, replacing the viral promoters with human endogenous promoters, such as EF1 $\alpha$  or human cellular ubiquitin C (UbC), resulted in persistent low-level transgene expression for at least 4 or 26 weeks, respectively, after nasal instillation of naked DNA (Gill et al., 2001). The UbC promoter also mediated long-term reporter gene expression after electroporation of naked DNA (Pringle et al., 2007) and after aerosolisation of lipoplexes (Hyde et al., 2008) to the mouse lung. The UbC promoter, however, could not be depleted of CpGs without a loss of transcriptional activity (Hyde et al., 2008) possibly due to the destruction of transcription factor binding sites in the promoter or intron. Yew and colleagues also demonstrated the benefit of using endogenous promoters for gene transfer to the mouse lung; the human *UBB* (UbB) promoter directed sustained transgene expression in the mouse lung after delivery of lipoplexes to immuno-deficient mice after intravenous delivery or nasal instillation (Yew et al., 2001). These studies led to the conclusion that although endogenous promoters may be weaker, resulting in lower levels of early transgene activity when compared with the use of strong viral promoters, they might be preferable when long-term transgene activity is required.

One strategy to generate both high-level and persistent transgene expression is to combine a viral enhancer sequence with a mammalian endogenous promoter to form a hybrid promoter. For example, by combining the wild-type hCMV enhancer with the UbB promoter to form the CUBI hybrid promoter, transgene expression and duration was increased after intranasal instillation (Yew et al., 2001). In the mouse aerosol model, however, the use of the wild-type hCMV enhancer in conjunction with the endogenous UbB, UbC or EF1 $\alpha$  promoters did not increase transgene persistence above the level observed with the UbC promoter alone (Hyde et al., 2008). However, as seen in Figure 1.4, using the synthetic CpG-free

murine CMV enhancer (mC) in conjunction with a CpG-free EF1 $\alpha$  promoter (EF) resulted in high reporter gene activity levels at early time-points, but transgene levels approached background by day 14 (Hyde et al., 2008). Unexpectedly perhaps, replacing the mC enhancer with the synthetic CpG-free hC enhancer resulted in transgene expression levels which persisted at high levels for up to 28 days after aerosol dosing with GL67A (Figure 1.4) (Hyde et al., 2008). A subsequent study demonstrated that the hCEFI enhancer/promoter could direct high-level transgene activity for up to 5 months following a single dose of 24  $\mu$ g of pDNA (Alton et al., 2014). Such persistence in the mouse lung in the absence of detectable inflammation had not been reported previously, and as such, the synthetic, hybrid hCEFI enhancer/promoter was selected for use in clinical vectors for gene therapy trials in CF patients (see 1.2.12).



**Figure 1.4. Lung transgene activity in a mouse aerosol model after delivery of pDNA/GL67A lipoplexes.**

Female BALB/c mice ( $n = 30$  per group) were exposed to 10 mL treatment aerosols (Section 2.5.6) containing 26.5 mg of the respective plasmid complexed with GL67A. Plasmid pG4-hCEFI-soLux contains the synthetic hybrid CpG-free versions of the hCMV enhancer (hC) and the elongation factor 1 $\alpha$  promoter (EF). Firefly luciferase activity was measured in lung lysates at the time-points indicated using luminometry (Section 2.6.1,  $n = 6$  per time-point). Plasmid pG4-mCEFI-soLux is identical to pG4-hCEFI-soLux except that it contains the synthetic hybrid CpG-free version of the mCMV enhancer (mC) instead of the hC enhancer (Hyde et al., 2008). Plasmid pG4-GZB-soLux contains the CpG-free version of the hCMV immediate-early gene enhancer and promoter (Yew et al., 2002), while plasmid pG4-EFI-soLux only contains EF as the promoter. Data is reproduced from Hyde *et al.*, 2008.

#### 1.4. Transcriptional regulation of transgene activity

Modifications to the design of plasmid vectors can improve and customise transgene expression; as described above, the choices of promoter and enhancer sequences are crucial, but the mechanisms are poorly understood. The reasons for loss of transgene activity, in the early days/weeks following gene transfer, may include: a) the loss of the gene transfer vector from the cell through degradation; b) the loss of the vector from the cell population via cell division or cell death; c) the loss of cells expressing the transgene product due to an anti-transgene immune response; or d) promoter attenuation due to inflammation-mediated silencing, or *de novo* methylation of the vector. Studies in the murine lung have shown that the duration of transgene expression in immuno-deficient and immuno-competent mice was similar, indicating that transient expression could not be attributed to an immune response against the transgene product (Gill et al., 2001). Furthermore, loss of transgene expression did not correlate with loss of pDNA (Gill et al., 2001; Pringle et al., 2005), or loss of transfected cells (Gill et al., 2001; Hyde et al., 2008). Crucially, it was shown that transgene expression that had fallen towards background levels could be reactivated by the administration of an adenovirus containing the E4 open reading frame (ORF), confirming that the transient plasmid-mediated expression was not due to the loss or destruction of pDNA (Yew et al., 1999a). The observed rescue of transgene expression was not due to stabilisation of plasmid vector, thereby suggesting that the plasmid was transcriptionally silenced rather than lost or destroyed (Gill et al., 2001).

Perhaps the most intriguing result in the mouse lung was observed by Hyde *et al.*, which indicates that the choice of enhancer can affect the outcome, leading to either persistent, high-level transgene activity (i.e. when using the hC enhancer) or transient transgene expression (i.e. when using the mC enhancer) (Hyde et al., 2008). Interestingly, the lack of induction of pro-inflammatory cytokines was in itself insufficient to mediate long-term persistence in the mouse lung (Hyde et al., 2008). Although the silencing mechanisms are

not completely understood, the study of endogenous transcriptional regulation and epigenetic events have laid down a useful framework to help understand exogenous gene regulation.

#### **1.4.1. Histone modifications**

In the eukaryotic nucleus, DNA is packaged into a nucleoprotein complex known as chromatin. The basic unit of chromatin is the nucleosome, which is composed of an octamer of the four canonical histones around which 147 bp of DNA is wrapped (Lalonde et al., 2014). At its most highly-compacted state as a metaphase chromosome immediately prior to mitosis, double-stranded helical DNA is compacted on the order of 10,000-fold (Alberts et al., 2002). The packaging of DNA into nucleosomes is known to be generally inhibitory to the binding of transcription factors and RNA polymerase II (Pol II) (Luger et al., 1997) and therefore, must accommodate a reversible process to transition from a more condensed inactive state (heterochromatin) to a more open, active structure (euchromatin) (reviewed in Trojer and Reinberg, 2007). The two major mechanisms by which chromatin is modified are via post-translational modifications and ATP-dependent chromatin remodelling.

A large number of post-translational modifications have been described, with acetylation and methylation being most widely studied. Histone modifications may have an effect on histone-DNA interactions, but also histone-histone interactions (Tessarz and Kouzarides, 2014). The deposition and removal of histone post-translational modifications is mediated by co-activators or co-repressors which can recruit specific enzymes, such as histone acetyltransferases (HAT) or histone deacetylases (HDAC), respectively (Lalonde et al., 2014).

Evidence of histone modifications after non-viral gene transfer has been controversial; decreased transgene expression in the mouse liver after HTV injection of naked pDNA was accompanied by increased histone binding, but not necessarily by histone modifications (Ochiai et al., 2007). Conversely, silencing of transgene expression mediated by

conventional pDNA was associated with increased heterochromatin-associated histone modifications and concomitant decrease associated with euchromatin, whereas persistent transgene expression from minicircle DNA (Section 1.3.3) was associated with euchromatin formation (Riu et al., 2007). These results suggest that although exogenous DNA associates with histones to aid the folding of the pDNA into a chromatinised entity, the relevance to subsequent transgene expression or repression is unknown.

#### **1.4.2. Transcription factories**

Discrete areas within the nucleus, termed 'transcription factories', are regions of concentrated polymerases and/or transcription factors (Iborra et al., 1996). Each transcription factory contains one type of polymerase (i.e. I, II or III) and some factories are richer in certain transcription factors than others (Xu and Cook, 2008). Transcription of transgenes may, therefore, be compromised by the area of the nucleus to which the pDNA is recruited. Studies using pDNA as probes demonstrated that the type of promoter contained therein can determine the transcription factory to which the plasmid is dynamically recruited (Larkin et al., 2013) indicating that the mere trafficking of pDNA into the nucleus may not be sufficient to promote transcription, but that the specific promoter may play an important role in different tissues, and may impact expression as a result.

#### **1.4.3. DNA methylation of the vector**

In eukaryotes, inhibition of gene expression is mediated by addition of a methyl group (CH<sub>3</sub>) onto the 5'-position of the cytosine ring of CpGs by the *de novo* methyltransferases DNMT3a and DNMT3b (Newell-Price et al., 2000). DNA methylation inhibits gene expression either directly, by interfering with the binding of transcription factors, or indirectly, by the binding of methyl-CpG-binding domain (MBD) family of proteins to methylated CpGs, which then modify the surrounding chromatin structure by the recruitment of chromatin remodelers, histone deacetylases and methylases, leading to transcriptional repression (see Section 1.4.1) (reviewed in Klose and Bird, 2006). When unmethylated and partially methylated plasmids

were transfected into cells grown in culture to determine how transcription is affected by discrete regions of plasmid methylation, the results indicated that transgene expression from a fully methylated plasmid was decreased by 400-fold compared with an unmethylated plasmid (Hsieh, 1997). The silencing effects were more prevalent when the length of methylated DNA was longer. A number of investigations in the mouse lung (Pringle et al., 2005) and liver (Chen et al., 2008; Ochiai et al., 2006) showed no evidence of *de novo* methylation after administration of CpG-containing non-viral vectors. However, other studies in the mouse liver have indeed shown that CpG-rich regions of some promoter sequences (such as hCMV) are methylated, while the AAT promoter remains unmethylated but only in the presence of an S/MAR element (Argyros et al., 2008). Furthermore, plasmid constructs containing CpGs are able to direct sustained transgene activity in the mouse lung, although the methylation status of these plasmids was not assessed (Gill et al., 2001, Bazzani et al., manuscript in preparation). Therefore, the role of CpGs in the silencing of transgene activity is yet to be fully understood.

#### **1.4.4. Transcription factors**

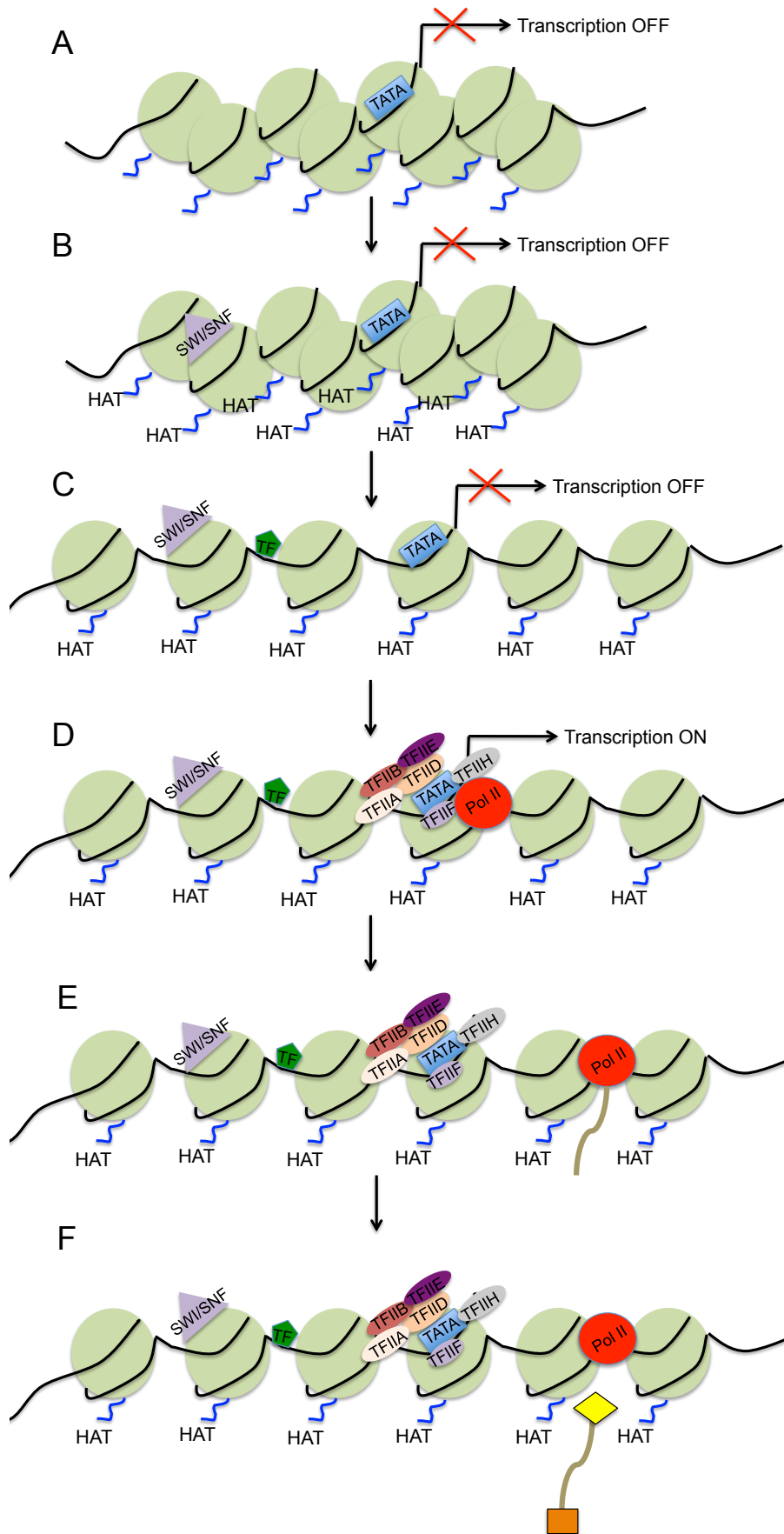
Gene expression is modulated by sequence-specific DNA-binding transcription factors (TFs), which can act either directly by interacting with components of the transcription machinery to repress or initiate transcription, or indirectly on the transcription machinery by recruiting factors that modify chromatin structure via post-translational modifications (Emerson, 2002; Orphanides and Reinberg, 2002). Transcriptional activation requires DNA accessibility for binding of the pre-initiation complex and Pol II (Section 1.4.1). Figure 1.5 shows the process of euchromatin formation, subsequent binding of TF proteins or general transcription factors, and eventual mRNA production. The translation from RNA to protein in the cytoplasm allows for the measurement of a transgene-derived protein, such as luciferase, which can then be used as a surrogate marker for active transcription.

Transcription factors may have a multifunctional role to play in transcription and can be modulated by other regulators found nearby; thus, a single activated TF can induce transcription of one gene while repressing another (Orphanides and Reinberg, 2002). A number of TFs can act as mediators of chromatin compaction and as such, can repress transcription (Trojer and Reinberg, 2007). Furthermore, the distance of the transcription factor binding sites (TFBSs) on the plasmid relative to the TSS may also impact transcriptional regulation. In a cell culture-based study which used transient transfections of pDNA to predict endogenous promoter function, it was demonstrated that when the binding site for certain transcription factors, such as YY1, was placed further away from the TSS, repression of transcription occurred, whereas when the binding site was placed closer to the TSS, transcriptional activation was observed (Whitfield et al., 2012). This result demonstrates that the same TF can act as both an activator and repressor of transcription, dependent on its location on the same promoter (Whitfield et al., 2012).

An additional role for TFs in gene transcription is one as a nuclear-targeting protein. Several studies have attempted to exploit this activity and increase exogenous transgene import into the nucleus, overcoming one of the major bottlenecks in non-viral gene transfer to non-dividing cells. Nuclear import of macromolecules larger than 40 kDa (such is the case with many non-viral vectors) requires energy and is mediated by nuclear localisation signals (NLSs), which are specific recognition sequences that bind to proteins synthesised in the cytoplasm to facilitate their movement into the nucleus (Kalderon et al., 1984). For example, after being translated in the cytoplasm, TFs that are solely active in the nucleus are trafficked into the cell's nucleus. Therefore, a number of TFs may act as NLSs and may bind to TFBSs present on the plasmid to facilitate its entry into the nucleus (Munkonge et al., 2003). However, the mere presence of an NLS may not be sufficient to effect the nuclear localisation of a particular protein, as the context in which the protein is made available is important (Jans and Hubner, 1996). For example, the transcription factor nuclear factor  $\kappa$ B

(NF $\kappa$ B) is present as an active form in the nucleus, but is inactive in the cytoplasm due to being complexed with its inhibitor, I $\kappa$ B (Baeuerle, 1998). Following activation by various stimuli, such as cytokines, NF $\kappa$ B is translocated into the nucleus (Ghosh and Baltimore, 1990). Non-viral vectors have been developed that include the binding site for NF $\kappa$ B, which can bind to the endogenous NF $\kappa$ B present in the cytoplasm and transport the pDNA molecule to the nucleus (Mesika et al., 2001). This approach has been successful in the increasing of transgene activity in cell culture models (Cramer et al., 2012; Mesika et al., 2001).

As described above, the choice of both the enhancer and promoter are important in generating the required transgene expression (see Section 1.3.5). A promoter sequence specifies under what circumstances transcription will be activated, dictating not only the quantity of Pol II recruited, but also whether re-initiation of transcription can occur (Yean and Gralla, 1997). A promoter includes a number of TFBSs, as well as a TSS and a TATA box, which is an A/T-rich region located about 30 bp upstream of the TSS, both of which are required for the recruitment of GTFs for transcription to be initiated (Blackwood and Kadonaga, 1998). The enhancer element contains numerous binding sites for TFs, many of which may be repeated numerous times (Stinski and Roehr, 1985). The presence of these TFs in exogenous gene expression may be responsible for the same mechanisms described above in endogenous eukaryotic systems. Furthermore, TFs are cell-type specific (Heintzman et al., 2009), which may become a critical consideration when designing plasmid constructs for certain tissues, or when comparing transgene expression across different cell types.



**Figure 1.5. Model of transcriptional activation.**

**A:** Heterochromatin is associated with a more 'closed' structure and repression of transcription. Histones (light green circles) are typically deacetylated (histone tails are shown as blue wavy lines). Promoter DNA contains a TATA box (blue rectangle), an A/T rich region located about 30 bp upstream of the transcription start site. **B:** ATP-dependent chromatin remodelers (such as SWI/SNF, purple triangle) or covalent histone modifying enzymes (such as histone acetyltransferases, shown as HAT) bind to the compact chromatin. **C:** The binding of such chromatin-remodelling complexes results in local increases of chromatin accessibility (i.e. euchromatin) and allows for transcription factor binding to take place (dark green pentagon). **D:** Transcription factor proteins may aid in the binding of general transcription factors to the TATA box, resulting in the formation of the pre-initiation complex and the subsequent recruitment of Pol II. Initiation of transcription can begin. **E:** Pol II initiates synthesis of full length mRNA via the addition of the two initiating nucleoside triphosphates (NTPs) and formation of the first phosphodiester bond of mRNA. The Pol II enzyme moves 5' to 3' along the gene sequence and extends the transcript (brown wavy line), in a process known as elongation. **F:** Upon reaching the end of a gene, Pol II stops transcription (termination), the mRNA transcript is cleaved and capped (orange square) and polyadenylation takes place (yellow diamond). The mature mRNA is then exported to the cytoplasm, where it is translated into protein by ribosomes. Figure is adapted from Emerson, 2002 and Orphanides & Reinberg, 2002.

## 1.5. Aims of this study

Persistent transgene activity in the lung in the absence of significant inflammation has been a long-term goal of the UK CF Gene Therapy Consortium ([www.cfgenetherapy.org.uk](http://www.cfgenetherapy.org.uk)). This was achieved in the mouse lung with the development of a non-viral plasmid, completely devoid of CpGs, containing a novel synthetic enhancer/promoter element termed hCEFI. This enhancer/promoter combination was selected for clinical trials for CF lung gene therapy after many *in vivo* studies, evaluating many plasmid constructs, were performed (Alton et al., 2014; Davies et al., 2008; Gill et al., 2001; Hyde et al., 2008; Lawton et al., 2002; Pringle et al., 2012b). Furthermore, the hCEFI enhancer/promoter was a key component in the non-viral vector which was recently evaluated in a large, multidose clinical trial which was sufficiently powered to detect significant differences in lung function (Alton et al., 2015) (Section 1.2.12). Therefore, the hCEFI enhancer/promoter represents one of the most clinically advanced sequences tested in CF patients to date. The reasons for the successful persistent expression from the hCEFI enhancer/promoter have never been elucidated. If the reasons for its successful persistent expression could be determined, it is possible that the specific reasons could be further exploited to generate more robust and higher-expressing vectors. Such an understanding is critical for the development of new gene therapy treatments for chronic lung conditions such as CF or AAT deficiency.

The studies described in this thesis aim to understand the reasons for the success of the hCEFI enhancer/promoter. The studies described in Chapter 3 aim to develop a cell culture model to investigate persistent transgene activity to avoid the issues associated with the use of the mouse aerosol model, which could in turn facilitate the investigation of the hCEFI enhancer/promoter. The work in Chapter 4 investigates the use of bioluminescence imaging of the mouse lung to measure luciferase activity after aerosol dosing of CpG-free plasmids in order to minimise the number of animals required for such studies. Additionally, the role of particular enhancer sequences to mediate transgene activity was studied by evaluating the

effects of enhancer-sequence deletions on subsequent transcriptional activity after aerosol delivery to the mouse lung. Subsequent transcription factor binding site analysis was performed in studies described in Chapter 5, with an aim to identify particular sites of interest with an additional focus on possible interactions that may enhance or repress transcriptional activity after *in vivo* delivery. Further refinement of the hC enhancer was studied in Chapter 6, with a focus on further short sequences and the effect on transgene activity in the mouse lung. Chapter 7 considers the wider utility of a minimal enhancer sequence in the liver, which is currently a target for other gene therapy-based applications. Additionally, the use of another CpG-free promoter was assessed, as well as the use of a minimal tandem enhancer in the context of the EF promoter. Overall, these studies were used to identify a particular sequence of the hC enhancer that is necessary and sufficient to mediate high-level and persistent transgene activity in the mouse lung after aerosol delivery.

## Chapter 2: Materials and Methods

### 2.1. Materials

#### 2.1.1. Reagents and solutions

All chemicals were supplied by Sigma-Aldrich (Poole, UK) unless otherwise stated. All solutions were prepared with Milli-Q reverse osmosed ultra pure water (Merck Millipore, Watford, UK), nuclease-free water (Promega, Southampton, UK) or endotoxin-free water for injection (WFI) (B. Braun Medical Ltd., Sheffield, UK). Where appropriate, solutions were sterilised by autoclaving for 15 to 30 minutes at 121 °C, or by filtration through a 0.22 µm filter (Nalgene, Rochester, NY, USA). Solutions were stored at room temperature (RT) unless otherwise stated. Where concentrations of solutions are indicated by percentages, these refer to weight:volume (w/v) ratios for solids at RT and volume:volume (v/v) ratios for liquids at RT.

#### 2.1.2. Centrifugation

Centrifugation was carried out using the Eppendorf 5415D microcentrifuge with F45-24-11 or F45-36-8 rotors (Eppendorf, Westbury, NY, USA), 13,200 revolutions per minute (rpm) (16,200 relative centrifugal force (rcf)) at RT; the Beckman J2-HS centrifuge with JA-10 rotor (Beckman, High Wycombe, UK), 12,000 rpm (12,900 rcf) at 4 °C; or the Megafuge 1.0 with 2150 rotor (Thermo Fisher Scientific, Loughborough, UK), 1,000 rpm (174 rcf) at RT.

### 2.2. Cell culture

#### 2.2.1. Cell culture conditions for the A549 cell line

The human lung adenocarcinoma cell line A549 (Giard et al., 1973) (LGC Standards, Middlesex, UK) was maintained in Ham's F-12K (Kaighn's) Medium with L-glutamine (Life Technologies, Paisley, UK), supplemented with 10 % fetal bovine serum (FBS) (supplemented Ham's medium). A549 cells were incubated in 5 % CO<sub>2</sub> at 37 °C. Cells were passaged to the required concentration twice a week and received fresh medium once a

week. Cells were removed from vented capped tissue culture flasks (Corning, Amsterdam, NL) by washing the monolayer with phosphate buffered saline (PBS), followed by adding 10 mL trypsin/EDTA solution diluted 1 in 5 in PBS (Life Technologies). The flask was placed in an incubator at 5 % CO<sub>2</sub> at 37 °C for approximately 5 minutes. The trypsin/EDTA solution was neutralised by the addition of 20 mL supplemented Ham's medium before centrifugation to pellet the cells. Cell pellets were re-suspended in 10 mL supplemented Ham's medium and 500 µL of the cell suspension was removed for a cell count with trypan blue on a haemocytometer. The required number of cells were passaged into new 175 cm<sup>2</sup> flasks containing 20 mL supplemented Ham's medium.

### **2.2.2. Cell culture conditions for the LA-4 cell line**

The mouse lung adenoma cell line LA-4 (Stoner et al., 1975) (kindly provided by Prof. Quentin Sattentau, University of Oxford, UK) was maintained in Ham's F-12K (Kaighn's) Medium with L-glutamine (Life Technologies), supplemented with 15 % FBS (supplemented Ham's medium). Cells were incubated and passaged as for A549 cells (Section 2.2.1).

### **2.2.3. Cell culture conditions for the air-liquid interface cultures**

Fully differentiated human air-liquid interface (ALI) cultures (MucilAir) were purchased from Epithelix (Geneva, Switzerland) and cultured at 37 °C and 5 % CO<sub>2</sub> as per the manufacturer's recommendations. The basolateral culture medium (MucilAir culture medium, Epithelix) was replaced every 2 to 3 days. The health of the cultures was assessed by the visualisation of beating cilia.

### **2.2.4. Transient transfection of cultured cell lines with plasmid DNA vectors**

For the transfection of A549 (Section 2.2.1) or LA-4 (Section 2.2.2) cell lines, cells were seeded in six-well plates (Corning) at the required cell density in supplemented Ham's medium and cultured overnight at 37 °C in 5% CO<sub>2</sub>. Plasmid DNA transfection was carried out using 1.5 µg/mL of pDNA complexed with branched 25 kDa PEI at N/P ratio of 10

(Davies et al., 2007) and diluted in a final volume of 2 mL OptiMEM medium (Life Technologies) per well. The growth medium was removed from the wells and replaced with the transfection mix. Cells were incubated for 24 hours, and the OptiMEM medium containing the pDNA complexes was removed and replaced with fresh supplemented Ham's medium every other day.

#### **2.2.5. Preparation of cell culture lysates for quantification of luciferase activity**

Transfection or cell growth medium (Section 2.2.4) was aspirated and the cell monolayer was rinsed twice in 2 mL PBS. Cells were lysed by the addition of 300  $\mu$ L 1x Reporter Lysis Buffer (RLB) and left for 15 minutes. The lysate supernatant was collected and transferred to a 1.7 mL microcentrifuge tube (Corning) and frozen at  $-80^{\circ}\text{C}$  for 15 minutes. After thawing, the supernatant was collected by centrifugation and transferred to a clean microcentrifuge tube. The cleared lysate was assayed for luciferase activity (Section 2.6.1) and total protein (Section 2.6.2) or frozen at  $-80^{\circ}\text{C}$  for later analysis.

#### **2.2.6. Transient transfection of ALI cultures with plasmid DNA vectors**

Human ALI cultures (Section 2.2.3) were transfected as previously described (Griesenbach et al., 2011) using 10  $\mu$ g of pDNA complexed with 25  $\mu$ l of Lipofectamine 2000 (Life Technologies), diluted in a final volume of 100  $\mu$ L OptiMEM medium. The ALI cultures were placed in fresh MucilAir culture medium and 100  $\mu$ L of the transfection mix was added to the apical side. Cells were incubated for 6 h, and the transfection mix was removed and the basolateral medium was replaced. The cultures were transferred to fresh medium every day. When required, 10 mM ethylene glycol tetraacetic acid (EGTA) was added in the transfection mix (Johnson et al., 2003a).

#### **2.2.7. Preparation of ALI cultures for quantification of luciferase activity**

The transfected ALI cultures (Section 2.2.6) were transferred into black 24-well plates (Porvair, Norfolk, UK) containing fresh MucilAir culture medium. D-Luciferin substrate

(PerkinElmer, Coventry, UK) was prepared at 30 mg/mL in water and filtered through a 0.22  $\mu\text{m}$  filter. The substrate was diluted to a working concentration of 1 mg/mL in MucilAir culture medium and a volume of 100  $\mu\text{L}$  was added to the apical side of the ALI culture (Mitomo et al., 2010). After 15 minutes, the cultures were imaged using an IVIS100 imaging system (PerkinElmer) (Section 2.6.3).

## **2.3. Plasmid cloning techniques**

### **2.3.1. PCR amplification**

Amplification of pDNA sequences was performed by preparing individual reactions containing 100 ng/ $\mu\text{L}$  of the pDNA template, 1x Phusion Buffer (Thermo Fisher Scientific), varying concentrations of 50 mM magnesium chloride (Thermo Fisher Scientific), 10mM dNTPs (NEB, Hitchin, UK), 10  $\mu\text{M}$  forward primer, 10  $\mu\text{M}$  reverse primer, 2 units of Phusion Hot Start II DNA polymerase (Thermo Fisher Scientific) and nuclease-free water to a total volume of 50  $\mu\text{L}$ . The components were added to thin-walled 0.2 mL tubes (Corning), briefly centrifuged in Eppendorf 5415D microcentrifuge and placed in a GeneAmp PCR system 2700 thermal cycler (Life Technologies). Cycling conditions were: 5 minutes at 98 °C, followed by 30 cycles of 98 °C at 1 minute, varying annealing temperatures based on optimal annealing temperature calculated by MacVector software v12.5.1 (MacVector, Cary, NC, USA) at 1 minute, and 72 °C for 2 minutes, with a final extension step of 72 °C for 7 minutes.

### **2.3.2. Gel electrophoresis**

DNA was size-fractionated by agarose gel electrophoresis. Gels consisted of 0.7 % or 2 % agarose (Life Technologies), 500 ng/mL ethidium bromide in 1 x Tris-acetate-EDTA (TAE) buffer (40 mM Tris(hydroxymethyl)aminomethane (Tris)-acetate and 1 mM ethylenediaminetetraacetic acid (EDTA), pH 8.3). Agarose was dissolved by heating in the microwave for 2 to 3 minutes, then allowed to cool to hand-hot, with the ethidium bromide added before the pouring. Prior to electrophoresis, DNA samples were mixed with 1x loading

dye in a final volume of 10 to 20  $\mu\text{L}$ , then electrophoresed at 80 to 120 volts for 1 to 2 hours. To allow DNA fragment size estimation, 500 ng of a 100 bp DNA ladder (range 100 bp to 1517 bp) (NEB), 1 kb DNA ladder (range 500 bp to 10 kb) (NEB), or supercoiled DNA ladder (range 2 to 10 kb) (NEB) were electrophoresed in at least one lane on each gel. Following electrophoresis, DNA bands were visualised by UV (312 nm wavelength), using a GeneFlash Gel Documentation System (Syngene, Cambridge, UK), and images were captured and saved on a CompactFlash card (SanDisk Corporation, California, USA).

### **2.3.3. Restriction enzyme digests**

All restriction enzyme digests were performed using enzymes purchased from New England Biolabs Ltd. (NEB), according to the manufacturer's instructions. Typically, a 20  $\mu\text{L}$  reaction volume (containing 2  $\mu\text{g}$  pDNA) was incubated for 1 hour at the temperature recommended by the supplier, using no more than 10 % final enzyme concentration.

### **2.3.4. Gel extraction of DNA fragments**

Typically, 30  $\mu\text{L}$  digested DNA (Section 2.3.3) was mixed with 1x loading dye (1.2 mM sucrose (VWR, Lutterworth, UK) and 3.6 nM bromophenol blue) and electrophoresed on a 0.7 % or 2 % agarose gel (Section 2.3.2), containing 100  $\mu\text{L}$  ethidium bromide, alongside a 1 kb or 100 bp ladder, as appropriate. Bands were visualised with long-wavelength UV light (366 nm), using a UVL-56 hand-held gel illuminator (UVP Inc., Upland, CA, USA). The required bands were excised using a scalpel and placed into pre-weighed 1.7 ml microcentrifuge tubes. The QIAQuick Gel Extraction Kit and protocol (Qiagen, Crawley, UK) was used for fragment purification according to the manufacturer's instructions. The DNA fragments were eluted with nuclease-free water and stored at  $-20\text{ }^{\circ}\text{C}$ . A sample of each purified fragment (3  $\mu\text{L}$ ) was subjected to agarose gel electrophoresis (Section 2.3.2) to confirm that the protocol had yielded the expected product.

### 2.3.5. Ligation of DNA fragments

DNA fragments were purified (Section 2.3.4) and ligated using T4 DNA ligase and associated buffer (NEB) according to the manufacturer's instructions. Typically, 1  $\mu$ L of T4 DNA ligase was mixed with 1  $\mu$ L vector and 7  $\mu$ L insert in a final volume of 10  $\mu$ L and incubated for at least 1 hour, and up to 18 hours, at 18 °C.

### 2.3.6. Site-directed mutagenesis

Plasmid DNA was subjected to site-directed mutagenesis using the Q5 Site-Directed Mutagenesis Kit (NEB) according to the manufacturer's instructions. Amplification of required pDNA sequences was performed by preparing individual reactions containing 10 ng/ $\mu$ L of the pDNA template, 1x Master Mix, 10  $\mu$ M forward primer, 10  $\mu$ M reverse primer, and nuclease-free water to a total volume of 25  $\mu$ L. The primers were designed using the program NEBaseChanger (NEB). The components were added to thin-walled 0.2 mL tubes (Corning) and briefly centrifuged in an Eppendorf 5415D microcentrifuge and placed in a GeneAmp PCR system 2700 thermal cycler. Cycling conditions were: 30 seconds at 98 °C, followed by 25 cycles of 98 °C at 10 seconds, 61 °C at 90 seconds, and 72 °C for 30 seconds, with a final extension step of 72 °C for 2 minutes. The PCR product was then subjected to kinase, ligase and *DpnI* treatment as per the manufacturer's instructions.

## 2.4. Amplification of plasmid DNA

### 2.4.1. Preparation of selective Luria Bertani plates

A volume of 400 mL sterile Luria Bertani (LB) agar (1 % bacto-tryptone (Life Technologies), 0.5 % bacto-yeast extract (Merck Millipore), 1% sodium chloride (NaCl, Thermo Fisher Scientific), 1.5 % bacto-agar (Life Technologies)) was microwaved until melted, then allowed to cool in a water bath at 50 °C. Kanamycin (50 mg/mL, filtered through a 0.22  $\mu$ m filter and stored at 4 °C) was added to a final concentration of 50  $\mu$ g/mL. A volume of 20 to 25 mL of

melted agar was poured into 9 cm petri dishes (Sterilin, Newport, UK) and once set, the plates were dried for about 1 hour at 42 °C.

#### **2.4.2. Bacterial transformation of plasmid DNA**

All pDNA constructs were grown in *E. coli* TOP10 (Life Technologies) or GT115 chemically competent cells (InvivoGen, Toulouse, France). Cells were transformed according to the manufacturer's instructions. Dilutions of transformed cells were spread on LB agar plates (Section 2.4.1) containing 50 µg/mL kanamycin to isolate single colonies for amplification.

#### **2.4.3. Inoculation into growth media**

A single transformed *E. coli* colony, grown overnight on selective LB agar plates (Section 2.4.2), was inoculated into either 5 mL or 400 mL 2x yeast extract/tryptone bacterial medium (1.6 % bacto-tryptone, 1 % bacto-yeast extract and 1 % NaCl) containing 50 µg/mL kanamycin. Cultures were incubated in an orbital shaking incubator Innova 4300 (New Brunswick Scientific, Edison, NJ, USA) at 180 rpm (n/a rcf), overnight at 34 °C.

#### **2.4.4. Purification of plasmid DNA from transformed bacteria**

Plasmid DNA from 5 mL cultures (Section 2.4.3) was purified using the Wizard Plus SV MiniPrep DNA Purification System (Promega), according to manufacturer's instructions. The DNA was eluted in 100 µL nuclease-free water and stored at -20 °C. Large-scale DNA preparations for *in vivo* studies were purified from 400 mL cultures using the EndoFree Plasmid Mega Kit (Qiagen), according to manufacturer's instructions. Final pellets of DNA were resuspended in 500 µL WFI and stored at -20 °C.

#### **2.4.5. Spectrophotometric quantification of DNA concentration in solution**

DNA concentration was measured spectrophotometrically using a NanoDrop Spectrophotometer 2000 (Thermo Fisher Scientific) (range 0.002 to 15 mg/mL), using the relationship between the absorption of UV light by DNA at 260 nm and the DNA

concentration, with an optical density at 260 nm ( $OD_{260}$ ) of 1.0 for a 50  $\mu\text{g}/\text{mL}$  solution of dsDNA. Each 2  $\mu\text{L}$  sample was read in duplicate or triplicate, and the average was used to determine the final concentration. The purity of the preparation was measured by calculation of the  $OD_{260}/OD_{280}$  ratio, rejecting preparations with values of  $< 1.8$ , which may be indicative of potential protein contamination.

## **2.5. Animal studies**

All procedures involving laboratory mice were carried out at the Biomedical Services Unit (BMS) (University of Oxford, John Radcliffe Hospital, Oxford, UK). All procedures were carried out under United Kingdom Home Office approved project and personal licenses for performing experiments on animals under the terms of the Animals (Scientific Procedures) Act 1986. Mice were housed in accordance with the United Kingdom Home Office ethical and welfare guidelines and fed on standard chow and water *ad libitum*. When SCID mice were used, water was supplemented with the antibiotic Septrin (5 mL in 200 mL water) (University of Oxford Vet Services, Oxford, UK).

### **2.5.1. Mouse strains**

Female BALB/c or SCID mice were age 6-12 weeks at point of procedure and purchased from Harlan Laboratories UK (Loughborough, UK), and were allowed to acclimatise for at least 7 days prior to procedures being performed.

### **2.5.2. Mouse anaesthesia**

For delivery to the lung via intranasal instillation (Section 2.5.5), to the liver for hydrodynamic tail vein injection (Section 2.5.7), and during live imaging (Section 2.5.10), mice were anaesthetised using inhalation of Isoflurane (University of Oxford Vet Services, Oxford, UK), until a balanced state of anaesthesia was achieved, as determined by level of response to foot pad pinch.

### 2.5.3. Preparation of lipoplexes for intranasal instillation

All solutions for delivery were prepared in WFI. Liposomes were generated by hydrating a vial containing GL67A (Eastman et al., 1997a) (1:2:0.05 molar ratio of GL67:DOPE:DMPE-PEG5000) (OctoPlus, Leiden, NL) to 1.2 mM with WFI and mixing until completely dissolved as previously described (Pringle et al., 2005). The liposome solution was then kept on ice throughout. Prior to mixing, 800  $\mu$ L of the 1.2 mM liposome solution was transferred to a 7 mL sterile polystyrene bijou tube (Sterilin) and incubated for 5 minutes at 30 °C. A volume of 800  $\mu$ L of pDNA at 1.6 mg/mL was added and the combined solution was incubated for a further 15 minutes at 30 °C to allow complex formation. The final 100  $\mu$ L mixture delivered to the mouse (Section 2.5.5) contained 80  $\mu$ g pDNA, at a pDNA:GL67A mM ratio of 2.4:0.6 (equivalent to a 1:0.25 ratio of lipid nitrogen (N):DNA phosphate (P)) as previously described (Lee et al., 1996).

### 2.5.4. Preparation of polyplexes for aerosolisation

Polyplexes were prepared as per Davies *et al.* (Davies et al., 2008). A volume of 10 mL polyplex was formed at a PEI nitrogen:DNA phosphate (N:P) molar ratio of 1:10. Plasmid DNA was diluted to a concentration of 0.4 mg/mL in WFI and 25 kDa branched PEI was prepared at a concentration of 4.3 mg/mL in sterile water at pH 7.4 and filtered through a 0.22  $\mu$ m filter. Polyplexes were prepared for each aerosol by mixing 5 mL pDNA solution with 5 mL PEI solution using continuous mixing. Complexes were allowed to form for at least 20 minutes, and up to 1 hour, at RT. Approximately 10 mL of the final mixture containing 2 mg pDNA was aerosolised (Section 2.5.6).

### 2.5.5. Delivery to the mouse lung via intranasal instillation

A volume of 100  $\mu$ L lipoplex (Section 2.5.3) was delivered to the lung via intranasal instillation as described previously (Lee et al., 1996). In brief, the anaesthetised mouse was held vertically with mouth closed. A single, continuous droplet was maintained during topical application of the 100  $\mu$ L lipoplex to the nose (using a Gilson P200 pipette (Gilson Inc.,

Middleton, WI, USA)), with the liquid being drawn into the nasal cavity and lungs as the mouse inhaled for breathing.

#### **2.5.6. Delivery to the mouse lung via aerosolisation**

Aerosol delivery of polyplexes (Section 2.5.4) was performed using a continuous, unrestrained whole body exposure system as previously described (Davies et al., 2008). Mice (up to 30 individuals at one time) were placed in an 8-L Perspex exposure chamber and exposed to aerosol, generated using an Aerotech II nebulizer (Biodex, Shirley, NY, USA), and operating at 40 pound-force per square inch (psi), with air from a Timeter PCS 414 air compressor (Allied Healthcare Products, St. Louis, MO, USA) as the driving gas.

#### **2.5.7. Delivery to the mouse liver via hydrodynamic tail vein injection**

Plasmid DNA was diluted to 10 µg/ml in PBS. An anaesthetised mouse was temporarily restrained in a mouse restrainer with access to the tail vein. A volume equivalent to 10 % bodyweight (approximately 1.8 to 2.0 mL) was injected into the tail vein of the anaesthetised mouse over a period of approximately 5 seconds (Liu et al., 1999; Zhang et al., 1999). Luciferase activity was measured by bioluminescence imaging (BLI) (Section 2.5.10).

#### **2.5.8. Harvesting of lung tissue and preparation of lysates**

Mice were euthanized by cervical dislocation at indicated time-points after vector delivery (Sections 2.5.5 and 2.5.6) and the lungs were removed (Gill et al., 2001). The tissue was rinsed in PBS until the majority of blood on the surface of the lung was removed, then immediately immersed in 300 µL 1 x RLB and placed on ice. Lungs were homogenised using Lysing Matrix D tubes (MP Biomedicals, Cambridge, UK) and processed using the FastPrep Instrument (MP Biomedicals) for 45 seconds at a speed setting of 4.0 m/s. The homogenates were frozen at -80 °C for at least 15 minutes. Upon thawing to RT, the homogenates were mixed for 10 seconds and transferred to QIAshredder columns (Qiagen) and centrifuged at 10,000 rpm (9,300 rcf) for 2 minutes. The lysates were frozen at -80 °C for at least 30

minutes, then thawed to RT and centrifuged for 10 minutes at 10,000 rpm (9,300 rcf). The supernatant was transferred to a clean microcentrifuge tube and assayed for luciferase activity (Section 2.6.1) and total protein (Section 2.6.2), or frozen at -80 °C for later analysis.

### **2.5.9. Harvesting of liver tissue and preparation of lysates**

Mice were euthanized by cervical dislocation at indicated time-points after reagent delivery (Section 2.5.7) and the livers were removed. The liver was minced with a scalpel, split into two aliquots (~600 mg each) and placed into TeenPrep Lysing Matrix tubes (MP Biomedicals) containing 9 mL RNeasy Lysis Buffer (RLT) Plus (Qiagen) with 10  $\mu$ L 14.3 M  $\beta$ -mercaptoethanol per 1 mL RLT Plus. Livers were processed using the FastPrep Instrument for 2 x 45 seconds at a speed setting of 6.5 m/s with inversion of the tube after each run. Liver homogenates were frozen at -80 °C for at least 1 hour. After thawing to RT, a volume of 400  $\mu$ L of homogenate was transferred to QIAshredder columns and centrifuged at  $\geq$  10,000 rpm (9,300 rcf) for 2 minutes. The cleared lysate was transferred to a clean 1.7 mL tube and assayed for vector-derived DNA (Section 2.7).

### **2.5.10. Preparation of mice for bioluminescence imaging**

D-Luciferin substrate was prepared at 15 mg/mL in PBS and filtered through a 0.22  $\mu$ m filter. Anaesthetised mice (Section 2.5.2) were administered 100  $\mu$ L via the intranasal route and left for 10 minutes prior to imaging (for quantification of lung luciferase activity) or with 2 x 150  $\mu$ L via the intraperitoneal route and left for 5 minutes prior to imaging (for quantification of liver luciferase activity) (Buckley et al., 2008) (Section 2.6.3). For quantification of background in the bioluminescence imager, naïve-treated mice received no substrate.

## **2.6. Quantification of transgene activity**

### **2.6.1. The quantification of luciferase activity from lysates**

The quantity of firefly luciferase activity in cell culture lysates (Section 2.2.5) or mouse lung lysates (Section 2.5.8) was determined using the Luciferase Assay Reagent System

(Promega) and the Glomax 20/20 Luminometer (Promega). A sample containing 20  $\mu$ L of the cleared lysate was pipetted into a clean 1.7 mL microcentrifuge tube and placed in a Glomax 20/20 Luminometer. A volume of 100  $\mu$ L of Luciferase Assay Substrate was added to each 20  $\mu$ L cell lysate sample. After 2 seconds, light output in relative light units (RLU) was measured over 10 seconds. All samples were read in duplicate and luciferase activity was normalised to total protein in the same sample, as determined in Section 2.6.2. For comparative purposes, 100 RLU per mg total protein corresponds to 2467 ng recombinant luciferase (Promega) per mg total protein (Pringle et al., 2012b).

### **2.6.2. The quantification of total protein from lysates**

Total protein in cell culture lysates (Section 2.2.5) or mouse lung lysates (Section 2.5.8) was determined using the Detergent-Compatible (DC) protein assay kit (Bio-Rad, Hemel Hempstead, UK). Cell culture lysates were assayed undiluted, while mouse lung lysates required a 40-fold dilution in Milli-Q water. A volume of 5  $\mu$ L of lysate was transferred into a 96-well plate (Thermo Fisher Scientific) in triplicate. Lyophilised BSA resuspended in Milli-Q water (5 mg/mL) (Bio-Rad) was used to generate a 7-point standard curve (range 1.75 mg/mL to 0.25 mg/mL in 0.25 mg/mL increments). Each standard was analysed in triplicate on the same plate as the samples. Reagents of the DC protein assay kit were added according to the manufacturer's instructions and the plate was incubated for 15 minutes at RT to allow the colorimetric assay to develop. Absorbance (750 nm) was measured on a SpectraMAX 250 spectrophotometer plate reader and SOFTmax Pro software v5.0.1 (Molecular Devices, Winnersh Triangle, UK) was used to quantify protein in samples by comparison to the BSA standard curve.

### **2.6.3. The quantification of luciferase activity using bioluminescence imaging**

Bioluminescence (photons/sec or photons/sec/cm<sup>2</sup>/sr) was measured in ALI cultures (Section 2.2.7) or living mice (Section 2.5.10) using an IVIS100 system (PerkinElmer) and the

software programme Living Image v4.2 (PerkinElmer). To quantify luciferase activity, photon emission was measured by marking a standardised area for quantification around the ALI culture (Mitomo et al., 2010), or the appropriate organ (Buckley et al., 2008; Griesenbach et al., 2008). Naïve samples (no treatment) were determined at identical time-points to experimental groups. Typically, mean naïve values are plotted for comparison.

The settings for image capture were dependent on the application:

For ALI cultures (Section 2.2.7): binning of 8, 30 to 60 second exposure

For mouse lung quantification (Section 2.5.10): binning of 8, 4 minute exposure

For mouse liver quantification (Section 2.5.10): binning of 8, 10 second exposure

#### **2.6.4. The visualisation of GFP activity using fluorescence microscopy**

Transfected ALI cultures (Section 2.2.6) were viewed under a Leitz DMIRBE (Leica, Milton Keynes, UK) fluorescent microscope equipped with a GFP filter set (Davies et al., 2007). Image acquisition was performed using a Nikon Coolpix 4500 camera (Nikon, Kingston upon Thames, UK).

### **2.7. Determination of plasmid DNA levels**

#### **2.7.1. Extraction of DNA from liver lysates**

Total DNA was prepared from the liver lysates (Section 2.5.9) using the AllPrep DNA isolation protocol (Qiagen) as per the manufacturer's instructions. The DNA was eluted in 100  $\mu$ L Buffer EB and the concentration measured using the Nanodrop spectrophotometer (Section 2.4.5) and diluted to 3 ng/ $\mu$ L in nuclease-free water. Levels of vector-specific DNA were quantified using quantitative real-time PCR (Section 2.7.2).

#### **2.7.2. Quantification of DNA using quantitative real-time PCR**

Absolute levels of vector-derived DNA extracted from liver lysates (Section 2.7.1) were determined by real-time quantitative PCR with the ABI PRISM 7700 Sequence Detection

System and Sequence Detector software v2.3 (Life Technologies) using vector or endogenous DNA specific primers and fluorogenic probes. The sequences of the forward (F), reverse (R) and probe (P) oligonucleotides were designed using Primer Express Software v.3.0 (Life Technologies) and are as follows: for vector-specific luciferase: F: 5'-GAA TAT GGC TCA TGG ATC CCC-3'; R: 5'-TCT TTG ACA ATT AAA CAT TGG CAT AGT-3'; P: 5'-FAM-CCT ATA GTG AGT TGT ATT ATA C-NFQ-3' (Pringle et al., 2005).

Triplicate 10  $\mu$ L reactions were performed for each DNA sample. Each 10  $\mu$ L reaction consisted of 2  $\mu$ L of DNA (6 ng total DNA), 1x TaqMan Universal Mastermix, 300 nM forward primer, 300 nM reverse primer, and 100 nM probe. PCR plates were heat-sealed with a clear seal (ABGene, Epsom, UK) using the ALPS 50V micro-plate sealer (Thermo Fisher Scientific), briefly centrifuged in a PerfectSpin P centrifuge (VWR) and placed in an ABI PRISM 7700 Sequence Detection System (Life Technologies). Cycling conditions were: 2 minutes at 50 °C, then 10 minutes at 95 °C, followed by 40 cycles of 1 seconds at 95 °C and 1 minute at 60 °C.

The number of copies of plasmid-derived DNA was calculated using *in vitro* transcribed portions of the target DNA to construct absolute copy number-based standard curves as previously described (Hyde et al., 2008; Rose et al., 2002). The amount of plasmid-derived DNA was normalised to total DNA in the same sample, as determined in Section 2.4.5.

### **2.7.3. Quantification of bacterial genomic DNA in plasmid DNA preparations using GT115-specific quantitative real-time PCR**

Absolute levels of chromosomal DNA (chrDNA) contamination in plasmid preparations were determined by real-time quantitative PCR using GT115-specific primers and fluorogenic probes as previously described (Bazzani et al., 2011). Quadruplicate 10  $\mu$ L PCR reactions were performed for each DNA sample. Each 10  $\mu$ L reaction consisted of 4  $\mu$ L of DNA (40 ng total DNA), 1x TaqMan Universal Mastermix (Life Technologies), 300 nM forward primer, 300

nM reverse primer, and 100 nM probe. Sequences of the forward (F), reverse (R) and probe (P) oligonucleotides used for each DNA target are as following: F: 5'-TGT GAG CGT CGC AGA ACA TT-3'; R: 5'-TTT TAG CAA CGT ACT ATC CAC TCA TGA-3'; P: 5'-VIC-CAT TGA CGC AGG TGA ATC GGA CGC-TAMRA-3'. PCR plates were heat-sealed with a clear seal (ABGene, Epsom, UK) using the ALPS 50V micro-plate sealer (Thermo Fisher Scientific), briefly centrifuged in a PerfectSpin P centrifuge (VWR) and placed in an ABI PRISM 7700 Sequence Detection System (Life Technologies). Cycling conditions were: 2 minutes at 50 °C, then 10 minutes at 95 °C, followed by 40 cycles of 1 seconds at 95 °C and 1 minute at 60 °C.

The number of copies of GT115-derived DNA was calculated from standard curves prepared from *E. coli* GT115 bacterial cultures (Section 2.4.2) and expressed as percentage chrDNA of total amount of DNA.

## 2.8. Statistical analysis

Data were analysed using GraphPad Prism software version 5 for Mac (La Jolla, CA, USA). When three or more groups were tested, ANOVA or Kruskal-Wallis tests were used for parametric and non-parametric tests, respectively. Parametric analysis (presented as group mean  $\pm$  SD) was used wherever possible. Non-parametric data (presented as group median  $\pm$  SD) were analysed using the Kruskal-Wallis test followed by Dunn's multiple comparison post-hoc tests. Multiple groups were analysed using one-way or two-way analysis of variance (ANOVA) followed by Bonferroni's post-hoc tests correction. One-way ANOVA was used to compare differences between groups, while two-way ANOVA was used to compare two factors (i.e. time and vector). Two-way ANOVA assumes that the replicates are sampled from Gaussian distributions; for this reason, log-adjustment was utilised to equalise the standard deviations to assure normality of data and equivalence of variance. Pairs of interest were analysed using two-tailed unpaired t-tests, whereby the assumption of equal variance

was tested in Prism using an F test. Post-hoc tests analysed all independent groups and tested against each other at each time-point within each experiment. Data were considered to be statistically significant different when  $p < 0.05$  (\*  $p < 0.05$ , \*\*  $p < 0.01$  and \*\*\*  $p < 0.001$ ). Except where noted, when statistical significance between groups was identified, the statistical power for such a difference exceeded 80 % using G\*Power software version 3.1 (Faul et al., 2007) (University of Dusseldorf, Germany).

Table 2.1. Plasmid DNA names and sources.

Plasmid Name	Source
pEGFPLuc	Clontech
pEGFP-N1	Clontech
pG4-hCEFI-soLux	Hyde <i>et al.</i> , 2008
pG4-mCEFI-soLux	Hyde <i>et al.</i> , 2008
pG4-EFI-soLux	Hyde <i>et al.</i> , 2008
pG5-mFABP-soLux	Pringle <i>et al.</i> , 2012a
pG5-hCmFABP-soLux	Pringle <i>et al.</i> , 2012a
pG5-hCEFI-soLux	Pringle <i>et al.</i> , 2012a

The plasmids were named according to the backbone (i.e. pG4 = fourth generation), enhancer/promoter combination, and the transgene (soLux).

EGFP: enhanced green fluorescent protein. hCEFI: synthetic, codon-optimised CpG-free versions of the hCMV enhancer, EF1 $\alpha$  promoter and intron. mCEFI: synthetic, codon-optimised CpG-free versions of the mCMV enhancer, EF1 $\alpha$  promoter and intron. EFI: synthetic, CpG-free versions of EF1 $\alpha$  promoter and intron. mFABP: promoter from murine fatty acid binding protein 4 gene. hCmFABP: synthetic, CpG-free version of hCMV enhancer and mFABP promoter.

## Chapter 3: Investigation of lung cell culture models

### 3.1. Introduction

The use of mouse models for the quantification of transgene expression mediated by non-viral vectors after delivery to the lung has been well established (Section 1.2.14). However, researchers have an ethical obligation to replace the use of animals by other models when possible. Where the use of animals cannot be replaced, the numbers used should be reduced and procedures refined to minimise the stress or discomfort that the animal experiences. This principle is known as the '3Rs' (Reduction, Refinement, Replacement) and was originally described by Russell and Burch in 1959 (Russell and Burch, 1959). The 3Rs were formally adopted into UK law in the Animal (Scientific Procedures) Act (ASPA) 1986. In addition, the procurement and husbandry of animals is expensive, and the generation of large quantities of pDNA for gene transfer experiments is also costly. The isolation of transfected mouse lung cells that contain plasmid, and the identification of cells that contain intact plasmid that is being actively transcribed, is difficult due to low levels of transfection after non-viral gene transfer. Using GFP as a reporter gene, Davies *et al.* were only able to observe approximately a few GFP-positive cells per cm<sup>2</sup> of mouse lung tissue following aerosol delivery of PEI/pDNA complexes (Davies et al., 2008). It is likely that pDNA is present in many more cells, but transgene activity is below the limit of GFP detection due to technical issues in the processing of lung sections, such as lung tissue autofluorescence and the lack of any signal amplification step in reporter gene detection (Davies et al., 2007). Nevertheless, the number of detectable transfected mouse lung cells is very low and for these reasons, cell culture models were investigated.

Three lung cell culture models were selected for studying transgene activity. Firstly, the A549 human lung cell line was investigated. The A549 cell line was generated from a human alveolar carcinoma (Giard et al., 1973) and was shown to share several characteristics with a primary culture of type II alveolar cells, including phospholipid composition and similar

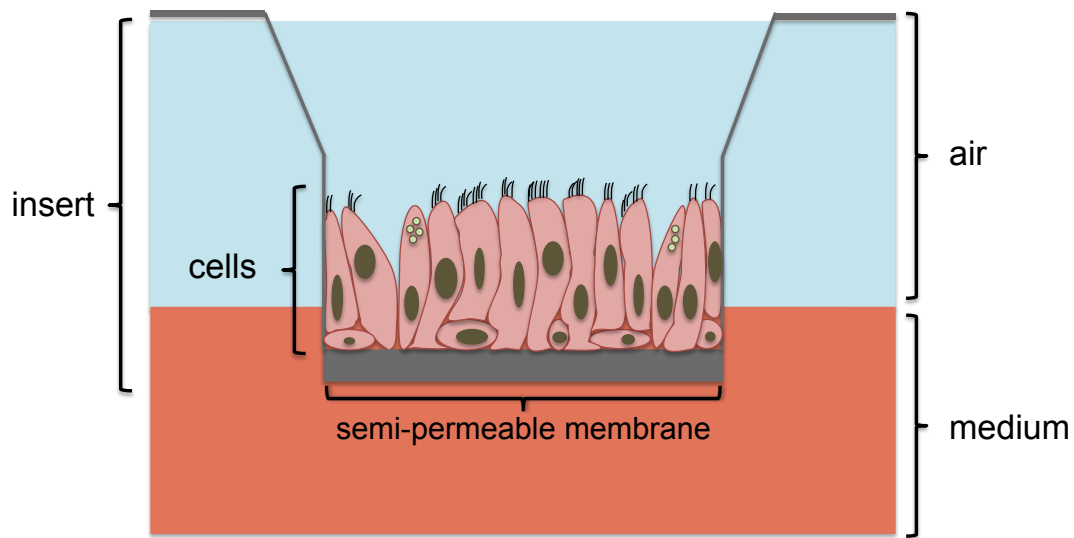
morphological features when assessed by electron microscopy (Nardone and Andrews, 1979). This cell line has been used in many studies for investigating transfection in lung cells prior to non-viral gene transfer *in vivo*, including a recently-published study which investigated the toxicity and transfection efficiency of CpG-free pDNA-chitosan polyplexes prior to instillation to the mouse lung (Wongrakpanich et al., 2014).

The second model investigated in this chapter was the LA-4 mouse lung cell line derived from a benign adenoma of the lung following an intraperitoneal injection of the carcinogen urethane (Stoner et al., 1975), which is typically used as an anaesthetic in non-recovery animal studies. Further characterisation of these cells showed that LA-4 cells have properties of type II alveolar cells (Stoner et al., 1978). This cell line has recently been used in conjunction with the A549 cell line for experiments to investigate cytokine production and cytotoxicity after transfection of mRNA/GL67 complexes (Andries et al., 2013) and has also been used in a number of publications investigating the transfection agent PEI (Beyerle et al., 2011).

The third model that was investigated was the culture of primary human epithelial cells at the air-liquid interface (ALI cultures; see Figure 3.1). The ALI culture model was first described as a method to achieve polarized cells in order to mimic the *in vivo* physiology of pseudostratified respiratory tract epithelium (Whitcutt et al., 1988). Table 3.1 contains a summary of relevant studies using ALI cultures that have aimed to understand the reasons for low transfection rates after non-viral gene transfer to the lung. The ALI cultures used in the experiments in this chapter were commercially sourced (Epithelix, Geneva, Switzerland), with the advantage that cultures are certified negative for the presence of bacteria, mycoplasma, yeast and fungi as well as specified viral pathogens (HIV, HBV and HCV). Additionally, each batch is tested for the presence of beating cilia and mucus, as well as the formation of tight junctions through trans-epithelial electrical resistance (TEER) testing (Constant et al., 2008). The differentiated cells can be kept for up to a year in culture without

loss of phenotype (Constant et al., 2008). Based on this information, the use of primary epithelial cells grown at air-liquid interface was deemed to be relevant as a model of the lung for investigating transgene expression.

The aims of the work presented in this chapter included: (i) to optimise cell growth and/or transfection conditions for A549 cells, LA-4 cells and ALI cultures, (ii) to measure transgene expression in these cells after transfection with plasmids pG4-mCEFI-soLux or pG4-hCEFI-soLux (Figure 3.2), and (iii) to perform time-course experiments to identify whether these cell lines/cultures could recapitulate the *in vivo* result where transgene activity from pG4-hCEFI-soLux persists at high levels but activity mediated by pG4-mCEFI-soLux is transient (Hyde et al., 2008).

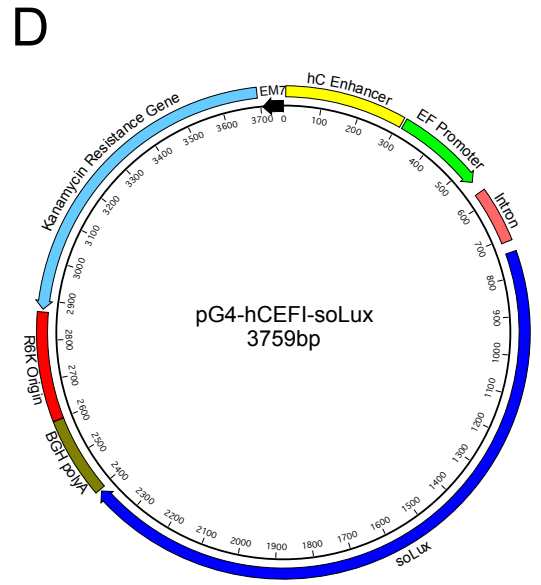
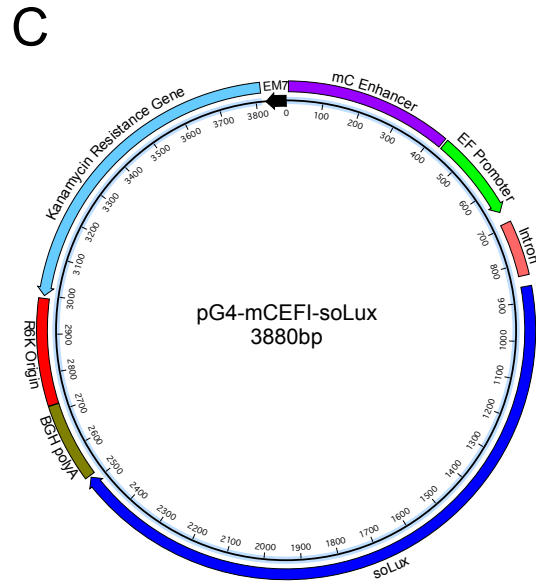
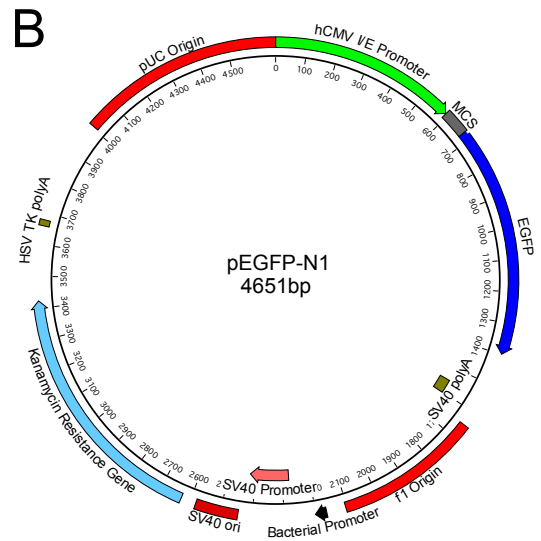
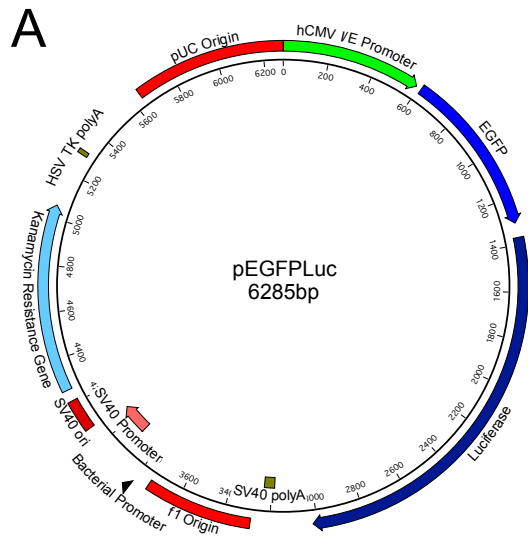


**Figure 3.1. Schematic showing detail of a MucilAir ALI culture.**

Cells in MucilAir ALI cultures are derived from primary human respiratory epithelium isolated from human nasal or bronchial biopsies (Huang et al., 2011). The cells are differentiated airway epithelial cells grown in a 3D culture system and possess many of the morphological and functional characteristics of human airway cells, such as the formation of tight junctions, generation of mucus, and the presence of polarized ciliated and non-ciliated epithelial cells, goblet cells and basal cells which make up the bronchial epithelium *in vivo* (Yamaya et al., 1992). The cell culture insert (6.5 mm diameter), containing the differentiated cells, is placed in a well of a 24-well plate containing cell culture medium so that the basolateral side of the cells on the semi-permeable membrane are submerged (Section 2.2.3). The apical side is exposed to humidified air at 5 % CO<sub>2</sub>.

**Table 3.1. Studies utilising ALI cultures to investigate reasons for low transfection efficiency after non-viral gene transfer to the lung.**

Reference	Experimental Aim	Key Findings
(Matsui et al., 1997)	Study the mechanisms that render differentiated cells difficult to transfect	<ul style="list-style-type: none"> <li>The apical surface of differentiated cells lacks pinocytic or phagocytic mechanisms which are more readily present in undifferentiated cells</li> </ul>
(Fasbender et al., 1997b)	Evaluate gene transfer to mature and immature cells	<ul style="list-style-type: none"> <li>The low rate of cell division in airway epithelia may limit DNA/lipid-mediated gene transfer</li> <li>Limited DNA uptake into the cell across the apical membrane</li> </ul>
(Jiang et al., 1998)	Determine basis for low transfection efficiency via polarization and differentiation status	<ul style="list-style-type: none"> <li>Transfection efficiency is determined by the polarization and mitotic state of cells</li> <li>Nuclear translocation is inefficient</li> </ul>
(Chu et al., 1999)	Investigate binding to apical surface and subsequent intracellular import	<ul style="list-style-type: none"> <li>Gene transfer is enhanced by procedures that allow complexes to interact with basolateral membrane (i.e. tight junction modulators)</li> <li>Binding and internalization of complexes (and subsequent transgene expression) is increased</li> </ul>



**Figure 3.2. Plasmid constructs used in Chapter 3 studies.**

**A:** The pEGFPLuc plasmid (Clontech, Saint-Germain-en-Laye France; Table 2.1) contains the enhanced green fluorescent protein (EGFP) and firefly luciferase transgenes driven by the hCMV immediate-early promoter (hCMV I/E). Other key features include the SV40 polyadenylation signal (polyA) and origin of replication, f1 origin of replication for single-stranded DNA production, bacterial promoter to express kanamycin resistance in *E. coli*, and the kanamycin resistance gene cassette, consisting of the SV40 early promoter and polyA signals from the Herpes simplex virus thymidine kinase (HSV TK) gene. The pUC origin of replication is provided for production in *E. coli*.

**B:** The pEGFP-N1 plasmid (Clontech; Table 2.1) is similar to the pEGFPLuc plasmid except that it does not contain the luciferase transgene. It contains a multiple cloning site (MCS).

**C:** The pG4-mCEFI-soLux plasmid (Hyde et al., 2008; Table 2.1) is completely devoid of CpGs and contains the synthetic version of the mCMV (mC) enhancer, the human elongation factor 1 $\alpha$  (EF) promoter, intron, firefly luciferase (soLux), bovine growth hormone (BGH) polyA, R6K origin of replication and the kanamycin resistance gene.

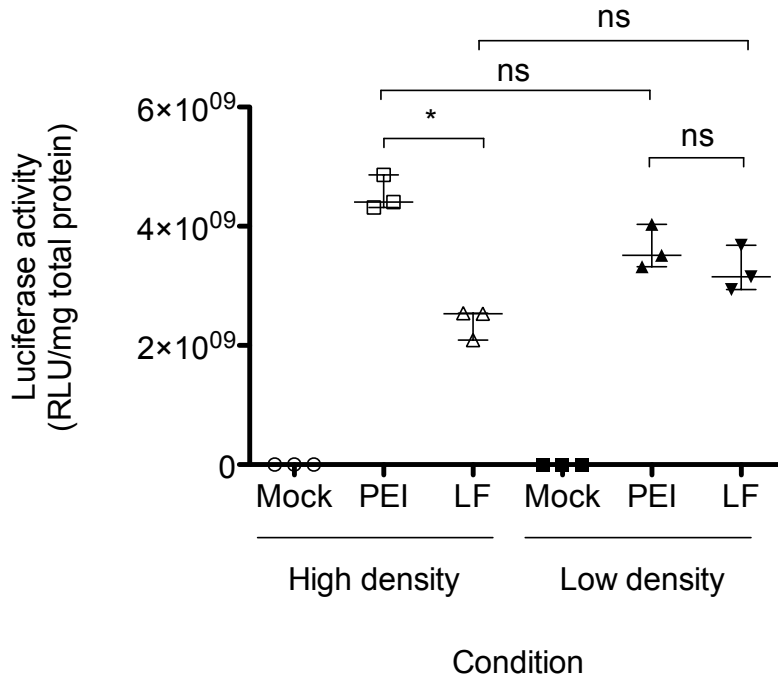
**D:** The pG4-hCEFI-soLux plasmid (Table 2.1) is identical to pG4-mCEFI-soLux except that it contains the synthetic version of the hCMV (hC) enhancer rather than the mC enhancer. Both CpG-free plasmids were sourced from the Gene Medicine Group, University of Oxford (Hyde et al., 2008).

Size in bp is indicated beneath the plasmid name.

## 3.2. Results

### 3.2.1. Establishing A549 cell culture and transfection conditions

A study was undertaken to maximise transfection efficiency in A549 cells to obtain high levels of transfected cells. Two parameters were investigated: the initial seeding density in the parental flask ( $2 \times 10^6$  or  $0.5 \times 10^6$ /flask) and the gene transfer agent for transfection (PEI or Lipofectamine 2000 (LF)). Cells were transfected using the plasmid pEGFP<sub>Luc</sub> (Figure 3.2A) (Section 2.2.4). Figure 3.3 shows that luciferase activity was highest from A549 cells harvested from the parental flask seeded at  $2 \times 10^6$  cells/flask and when PEI as the transfection agent was used ( $p < 0.05$  when compared to LF). Although the difference between the two densities was not significant when PEI was used, there was a trend towards higher luciferase activity when cells were seeded from the high-density parental flask. Therefore, all subsequent studies were performed with parental flasks seeded at  $2 \times 10^6$  cells/flask.



**Figure 3.3. Comparison of luciferase activity after transfection of A549 cells harvested from high- or low-density parental cultures using PEI or Lipofectamine 2000 (LF).**

Flasks containing A549 cells were cultured for three passages at high ( $2 \times 10^6$  cells/flask) or low ( $0.5 \times 10^6$  cells/flask) density (Section 2.2.1). Cells were passaged when confluence approached 80 % (high density) or 50 % (low density), 3 to 4 days after seeding (Section 2.2.1). Cultures were seeded into six-well plates with  $5 \times 10^5$  cells/well and incubated overnight prior to transfection with pEGFP<sub>Luc</sub> (Figure 3.2A) complexed with PEI or LF (Section 2.2.4). Mock-treated cells were treated with transfection medium only in the absence of complexes. Three wells per condition were harvested at 24 hours after transfection (Section 2.2.5). Levels of transgene activity were assessed by luminometry (Section 2.6.1). Data are shown as median  $\pm$  range. Statistical differences between conditions were analysed by Kruskal-Wallis followed by Dunn's multiple comparison post-hoc tests: \*  $p < 0.05$ .

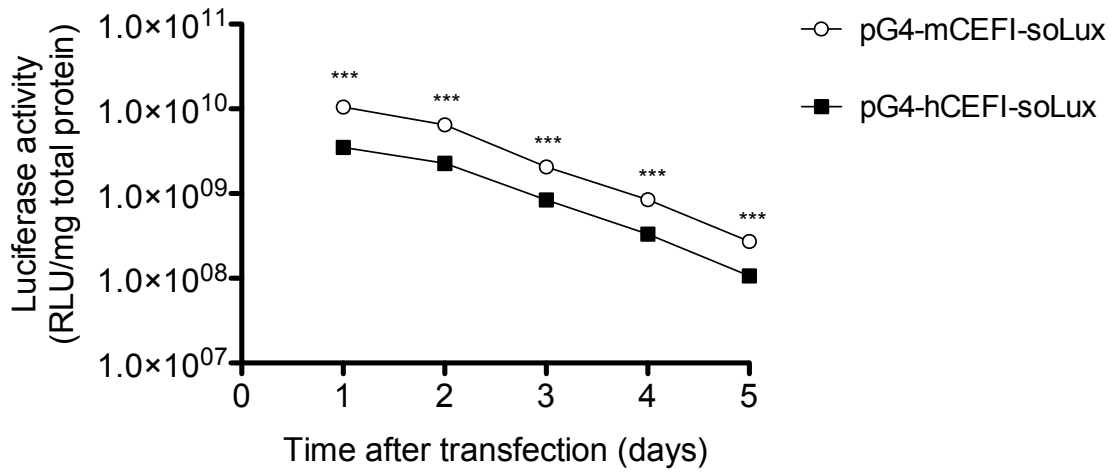
### 3.2.2. Time-course studies of pG4-hCEFI-soLux and pG4-mCEFI-soLux in A549 cells

Using the cell growth ( $2 \times 10^6$  cells/parental flask) and gene transfer agent (PEI) parameters established in Section 3.2.1, a time-course experiment was performed to compare luciferase activity mediated by pG4-mCEFI-soLux (Figure 3.2C) or pG4-hCEFI-soLux (Figure 3.2D). Cells were harvested from the parental flask and seeded at low ( $5 \times 10^4$  cells/well, 25 % confluence) or high ( $5 \times 10^5$  cells/well, 100 % confluence) densities prior to transfection (Section 2.2.4). Results in Figure 3.4 show that luciferase activity from both plasmids was high at early time-points and then declined towards background. There was no evidence that luciferase activity from either plasmid persisted at the higher level under the conditions used. However, there was a trend towards a modest increase in luciferase activity mediated by pG4-hCEFI-soLux between 1 and 2 days in cells seeded at low densities prior to transfection (Figure 3.4B). As the magnitude of this effect was modest, a further study was performed to investigate if the apparent peaks of transgene activity observed in Figure 3.4B were reproducible. Thus, a shorter time-course experiment was performed in triplicate with the addition of earlier sample harvest time-points at 6 hours and 12 hours. Low-confluence A549 cells were transfected with pG4-mCEFI-soLux or pG4-hCEFI-soLux. The results (Figure 3.5) showed that luciferase activity was higher in cells transfected with pG4-mCEFI-soLux, indicating there were significant differences in activity between the two plasmids as early as 12 hours post-transfection. The results also confirmed those observed from the initial low-density time-course experiment (Figure 3.4B) with a reproducible peak in activity at 1 day for cells transfected with pG4-mCEFI-soLux and 2 days for pG4-hCEFI-soLux.

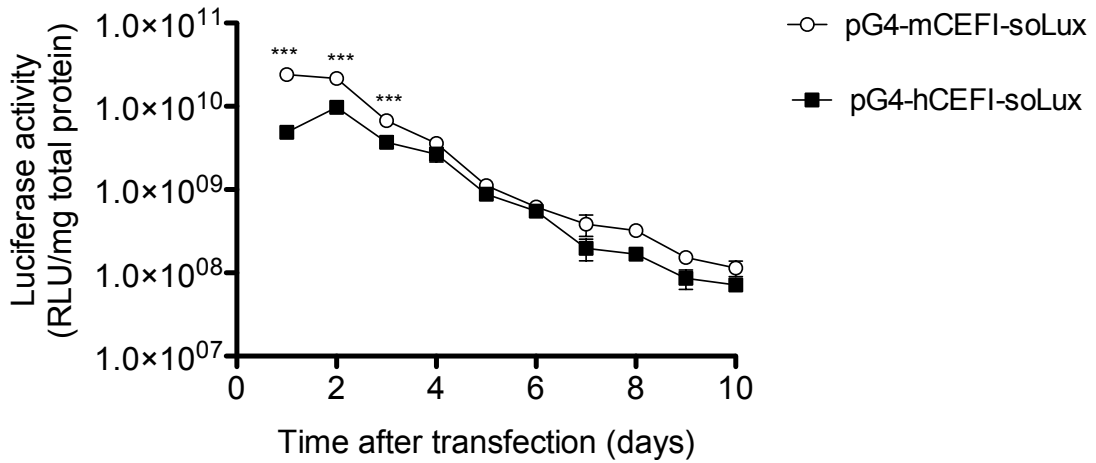
Despite small differences in luciferase activity from these two plasmids at early timepoints, neither plasmid directed transgene activity beyond a few days. Although the peak of transgene activity from both plasmids was reproducible in triplicate experiments (i.e. peak at 1 day for pG4-mCEFI-soLux and 2 days for pG4-hCEFI-soLux), the overall profile of activity did not recapitulate results observed with the aerosol *in vivo* model, which shows persistence

of luciferase activity mediated by pG4-hCEFI-soLux compared with pG4-mCEFI-soLux which rapidly decreases (Hyde et al., 2008). Therefore, the A549 cell culture system does not appear to be a suitable model for mimicking the persistence of pG4-hCEFI-soLux observed *in vivo* and an alternative cell culture model was investigated.

**A**

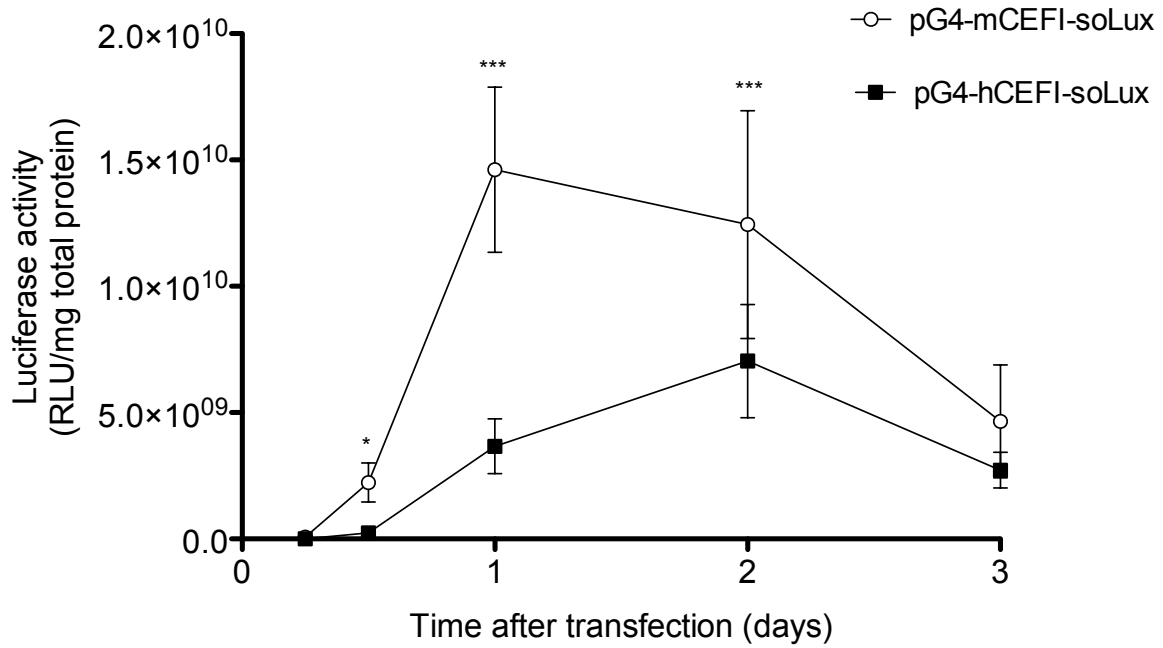


**B**



**Figure 3.4. Comparison of luciferase activity after transfection of (A) high and (B) low densities of A549 cells with pG4-mCEFI-soLux or pG4-hCEFI-soLux over an extended period.**

Flasks containing A549 cells were inoculated with  $2 \times 10^6$  cells/flask and cultured until ~80 % confluent (Section 2.2.1). Subsequently, cells were seeded at either (A) high or (B) low densities ( $5 \times 10^5$  and  $5 \times 10^4$  cells/well, respectively) in six-well plates (Section 2.2.4). Cells were incubated overnight prior to transfection with either pG4-mCEFI-soLux (Figure 3.2C) or pG4-hCEFI-soLux (Figure 3.2D) complexed with PEI (Section 2.2.4). Three wells per condition were harvested daily up to 5 or 10 days for the high or low conditions, respectively (Section 2.2.5). Data are shown as mean  $\pm$  SD. Levels of transgene activity were assessed by luminometry (Section 2.6.1). Statistical differences between conditions were analysed by two-way ANOVA followed by Bonferroni's multiple comparison post-hoc tests: \*\*\*  $p < 0.001$ .



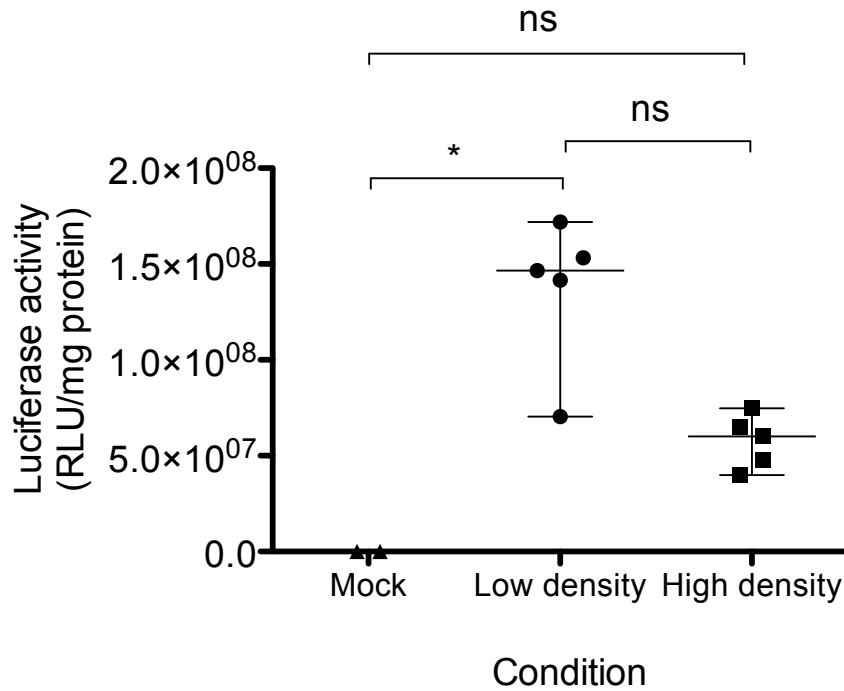
**Figure 3.5. Comparison of luciferase activity after transfection of low-confluence A549 cells with pG4-mCEFI-soLux or pG4-hCEFI-soLux over a short duration.**

Cultures of A549 cells (Section 2.2.1) were seeded at  $5 \times 10^4$  cells/well of a six-well plate and incubated overnight prior to transfection with either pG4-mCEFI-soLux (Figure 3.2C) or pG4-hCEFI-soLux (Figure 3.2D) (Section 2.2.4). Five wells per condition were evaluated from 6 hours to 3 days after transfection (Section 2.2.5). The results of three independent experiments are shown here. Levels of transgene activity were assessed by luminometry (Section 2.6.1). Data are shown as mean  $\pm$  SD. Statistical differences between conditions were analysed by two-way ANOVA with Bonferroni's multiple comparison post-hoc tests:

\*\*\*  $p < 0.001$  and \*  $p < 0.05$ .

### 3.2.3. Determining LA-4 transfection conditions

The A549 cell line is of human origin; therefore, a murine lung cell line was investigated to determine whether the use of mouse cells could recapitulate results obtained *in vivo* (Hyde et al., 2008). An experiment was performed to identify whether transfection of pG4-hCEFI-soLux (Figure 3.2D) complexed with PEI using high or low density cells at point of transfection ( $5 \times 10^5$  and  $5 \times 10^4$  cells/well, respectively) could lead to detectable luciferase activity when measured by standard luminometry. The results (Figure 3.6) showed that luciferase activity was detectable 24 hours after transfection in both high- and low-density cultures, although the mean activity was approximately 2-fold higher in the low-density culture. Therefore, the low-density condition was chosen for the LA-4 time-course experiment.

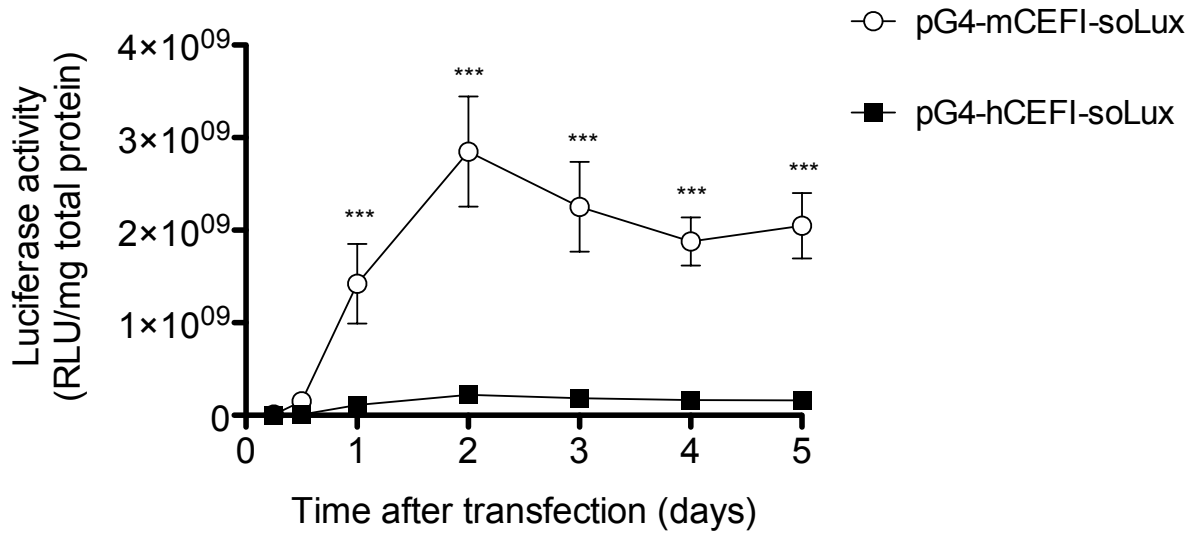


**Figure 3.6. Comparison of luciferase activity after transfection of low- and high-density LA-4 cells with pG4-hCEFI-soLux.**

Flasks containing LA-4 cells were inoculated at  $3 \times 10^6$  cells/flask until 90 % confluent (Section 2.2.2) and seeded at either low- or high-density in a six-well plate ( $5 \times 10^4$  and  $5 \times 10^5$  cells/well, respectively) (Section 2.2.4). Cells were incubated overnight prior to transfection with pG4-hCEFI-soLux (Figure 3.2D) complexed with PEI (Section 2.2.4). Six wells per condition were harvested 24 hours after transfection (Section 2.2.5). Levels of transgene activity were assessed by luminometry (Section 2.6.1). Data are shown as median  $\pm$  range. Statistical differences between conditions were analysed by Kruskal-Wallis followed by Dunn's multiple comparison post-hoc tests: \*  $p < 0.05$ .

#### 3.2.4. Time-course study using in LA-4 cells

Using the cell density for transfection established in Section 3.2.3 ( $5 \times 10^4$  cells/well), a time-course experiment was performed to compare transgene activity in LA-4 cells after transfection with pG4-mCEFI-soLux (Figure 3.2C) or pG4-hCEFI-soLux (Figure 3.2D) using low density cells at transfection. The results (Figure 3.7) showed that peak luciferase activity was detected at 2 days after transfection with both plasmids, but was 13-fold higher for pG4-mCEFI-soLux than pG4-hCEFI-soLux. In addition, the levels of transgene activity remained relatively unchanged from day 2 to day 5. Overall, the results obtained from the LA-4 time-course experiment did not recapitulate the mouse *in vivo* data (Hyde et al., 2008).



**Figure 3.7. Time-course of luciferase activity after transfection of low-confluence LA-4 cells with pG4-mCEFI-soLux or pG4-hCEFI-soLux.**

Cultures of LA-4 cells (Section 2.2.2) were seeded at  $5 \times 10^4$  cells per well of a six-well plate and incubated overnight prior to transfection with either pG4-mCEFI-soLux (Figure 3.2C) or pG4-hCEFI-soLux (Figure 3.2D) complexed with PEI (Section 2.2.4). Five wells per condition were evaluated from 6 hours to 3 days after transfection (Section 2.2.5). The results of two independent experiments are shown here. Levels of transgene activity were assessed by luminometry (Section 2.6.1). Data are shown as mean  $\pm$  SD. Statistical differences between conditions were analysed by two-way ANOVA with Bonferroni's multiple comparison post-hoc tests: \*\*\*  $p < 0.001$ .

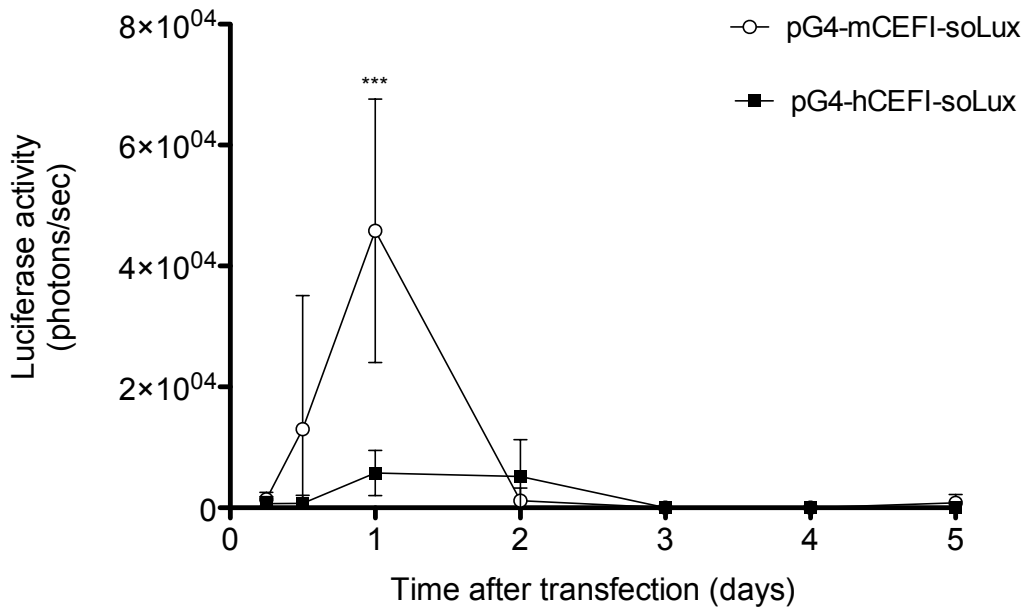
### 3.2.5. Establishing transfection conditions for ALI cultures

Primary human bronchial epithelial cells grown at air-liquid interface (ALI cultures, Figure 3.1) were investigated as a possible cell culture model. Data from previous work in the laboratory to compare transgene activity driven by the hCEFI or mCEFI enhancer/promoter sequences following transfection with PEI suggested that PEI was too variable to use as a reliable GTA in ALI cultures, leading to excessive cell damage (pers. comm. S. Sumner-Jones). Thus, Lipofectamine, with the most reproducible results, was used as the GTA of choice for further experiments.

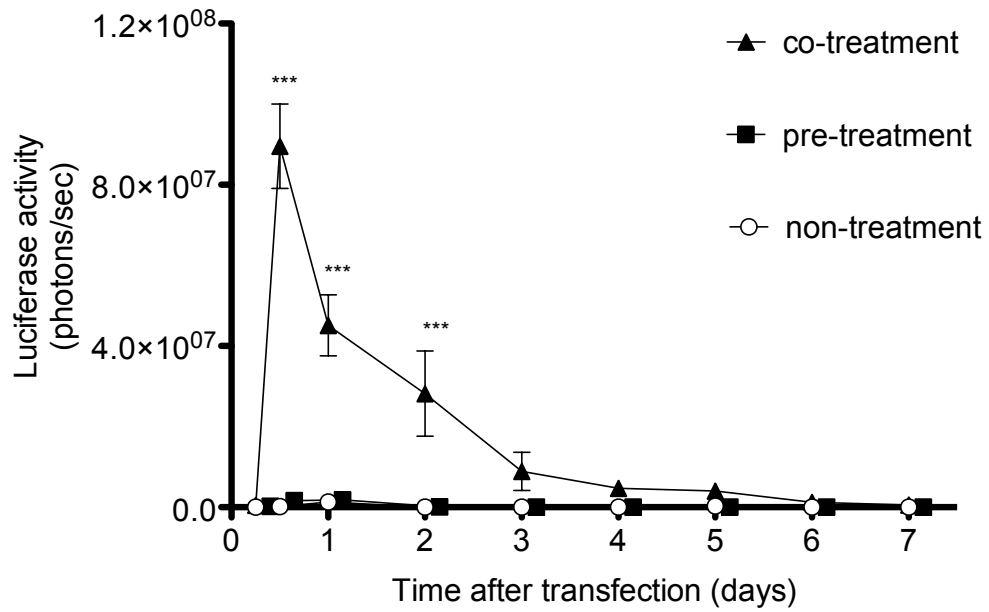
Transgene activity in ALI cultures after transfection was assessed using bioluminescence imaging (BLI) (Section 2.2.7). As seen in Figure 3.8A, luciferase activity mediated by pG4-mCEFI-soLux (Figure 3.2C) was higher than pG4-hCEFI-soLux (Figure 3.2D). However, bioluminescence values were close to the limit of detection and luciferase activity fell to baseline levels by 3 days. In order to increase detectable reporter activity, the effect of a tight junction modulator on transfection efficiency was investigated using pG4-mCEFI-soLux (Section 2.2.6). Ethylene glycol tetraacetic acid (EGTA) was applied apically to ALI cultures used previously, either as a pre-treatment prior to transfection, or co-treatment within the transfection mix. The peak level of activity observed at 12 hours was 63-fold higher and increased by 4 days compared with non-treated or pre-treated cultures (Figure 3.8B). In order to determine if this result was a function of a greater number of transfected cells, or higher transgene expression in a similar number of transfected cells, a further experiment was performed using a plasmid containing the GFP reporter gene. Cultures of ALI cells were transfected with the pEGFP-N1 plasmid (Figure 3.2B), either with or without EGTA in the transfection mix. The cells were observed using fluorescence microscopy (Section 2.6.4) at 6 and 24 hours after transfection. The number of transfected cells positive for GFP at 6 hours after transfection in both conditions was very low and fluorescence was very faint (data not shown). However, at 24 hours after transfection, both the number and brightness of cells

transfected with EGTA was greater than those without EGTA (Figure 3.9). These results suggest that the increased levels of transgene activity may be due to both increased number of transfected cells and higher expression levels in transfected cells when the basolateral membrane is accessible due to tight junction disruption.

A



B



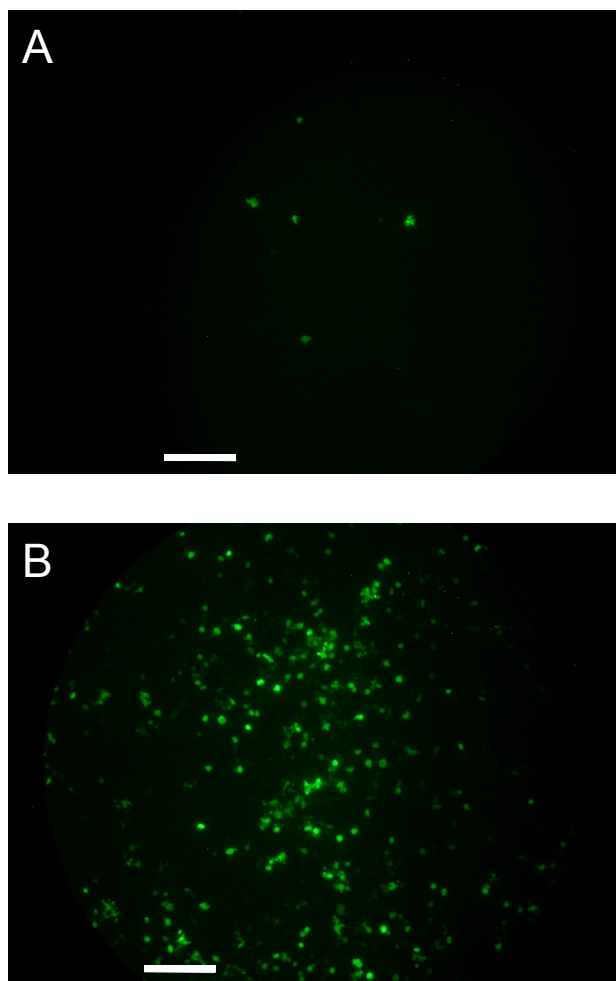
---

**Figure 3.8. Comparison of luciferase activity after transfection of ALI cultures with pG4-mCEFI-soLux or pG4-hCEFI-soLux.**

Cultures of ALI cells (Section 2.2.3) were transfected using the standard protocol (n = 3 per plasmid; Section 2.2.6). Luciferase activity was assessed using BLI (Section 2.2.7) at indicated time-points until activity was below background (untreated control, n=1).

**A:** The difference in activity between the pG4-mCEFI-soLux (Figure 3.2C) and pG4-hCEFI-soLux (Figure 3.2D) was statistically significant at 1 day after transfection: \*\*\*  $p < 0.001$  (two-way ANOVA with Bonferroni's multiple comparison post-hoc tests). Data are shown as mean  $\pm$  SD.

**B:** ALI cultures were transfected with pG4-mCEFI-soLux with EGTA included before transfection (pre-treatment, n = 3), or included in the transfection mix (co-treatment, n = 3), or not included (non-treatment, n = 1). Data are shown as mean  $\pm$  SD. The pre-treatment data points have been shifted for clarity. Statistical differences between conditions were analysed by two-way ANOVA with Bonferroni's multiple comparison post-hoc tests: \*\*\*  $p < 0.001$ .

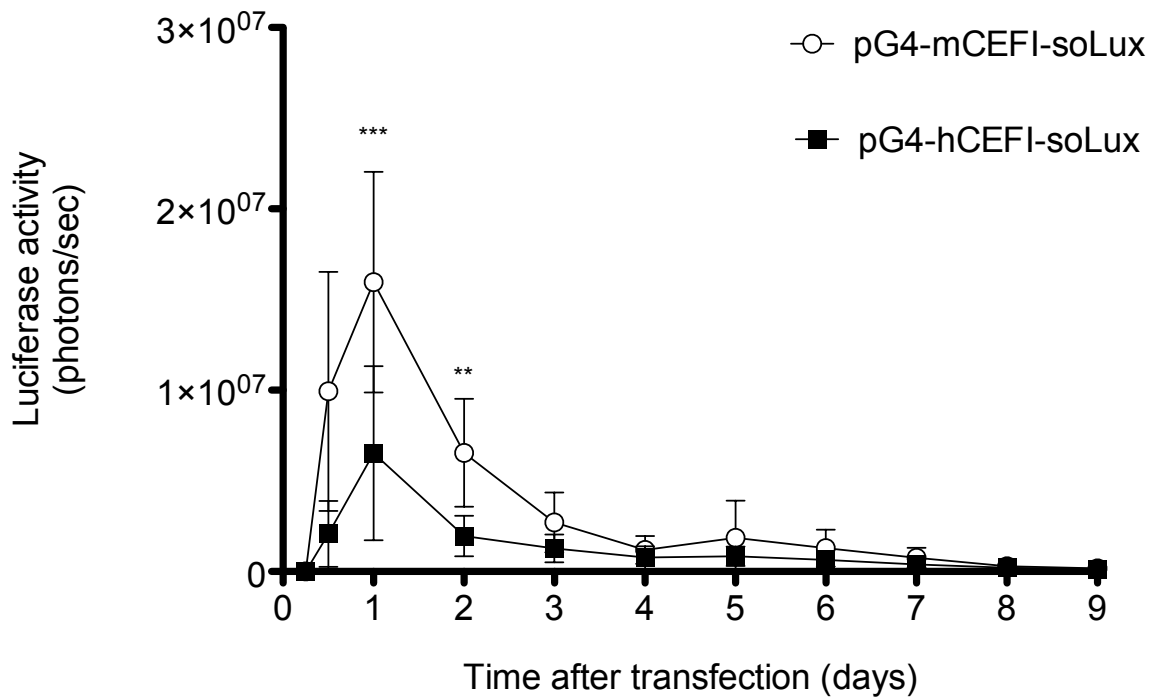


**Figure 3.9. Comparison of GFP activity after transfection of ALI cultures with pEGFP-N1.**

ALI cultures (Section 2.2.3) were transfected with pEGFP-N1 (Figure 3.2B) (Section 2.2.6) using (A) standard protocol ( $n = 3$ ) or (B) including EGTA in the transfection mix ( $n = 3$ ). Fluorescent microscopy imaging (Section 2.6.4; 10x magnification) was performed 24 hours after transfection. The image in Panel A represents the best-case sample observed from all three inserts after standard transfection (without EGTA) while the image in Panel B is a representative sample seen in every field of view of all three inserts treated with EGTA. White bars represent 100  $\mu\text{m}$ .

### 3.2.6. Comparison of luciferase activity in ALI cultures

New ALI cultures were transfected with pG4-mCEFI-soLux (Figure 3.2C) or pG4-hCEFI-soLux (Figure 3.2D) in the presence of 10 mM EGTA (Section 2.2.6, co-treatment protocol). Bioluminescence imaging (Section 2.2.7) was performed regularly until luminescence values fell below background (values obtained from a non-treated culture). Despite the increased level and duration of luciferase activity compared with previous experiments (Figure 3.8), neither plasmid directed persistent transgene activity; all cultures were below background level by 9 days after transfection (Figure 3.10). As observed in the A549 cell line studies (Section 3.2.2), activity in cells transfected with pG4-mCEFI-soLux was generally higher at early time-points than those transfected with pG4-hCEFI-soLux. Thus, the results previously observed in the mouse lung (Hyde et al., 2008) were not recapitulated in this model.



**Figure 3.10. Time-course of luciferase activity in ALI cultures after transfection with pG4-mCEFI-soLux or pG4-hCEFI-soLux in the presence of 10 mM EGTA.**

ALI cultures (Section 2.2.3) were transfected with either pG4-mCEFI-soLux (Figure 3.2C) or pG4-hCEFI-soLux (Figure 3.2D) in the presence of 10 mM EGTA ( $n = 7$ , Section 2.2.6). Luciferase activity was assessed using BLI (Section 2.2.7) at indicated time-points until activity was below background (untreated control,  $n = 1$ ). Statistical differences between conditions were analysed by two-way ANOVA with Bonferroni's multiple comparison post-hoc tests: \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

### 3.3. Discussion

Studies in this chapter investigated the use of cell culture models of airway cells to determine if the transgene activity observed after transient transfection of pG4-hCEFI-soLux and pG4-mCEFI-soLux could reflect observations in the mouse lung (Hyde et al., 2008). None of the cell culture systems tested here mimicked the long-term duration of pG4-hCEFI-soLux transgene activity and the rapid decrease of activity after transfection with pG4-mCEFI-soLux under the conditions tested. The results obtained from the cell lines is perhaps not surprising. Little correlation has been found between the predictive quality of non-viral transfection in cell lines and gene transfer *in vivo* (Lee et al., 1996; Pringle et al., 2012b). One could question whether the use of a human cell line, such as A549 cells, to recapitulate a result observed in the mouse lung is appropriate. Therefore, the murine lung cell line LA-4 was tested. The results (Figure 3.7) showed that transgene persistence from both plasmids seemed likely. This may be due to the expression of different transcription factors in the murine cell line when compared with A549 cells. Irrespective of these differences, recapitulation of the *in vivo* result was not observed. Further investigation into possible differences was not pursued here, but microarray gene expression assays could be performed to identify differential gene expression profiles between these two cell lines.

Cell culture may fail to provide an adequate model to predict or recapitulate *in vivo* results due to a number of reasons. Cell lines typically comprise a clonal population of genetically identical cells, which although advantageous in minimising experimental variability means that expression of specific cytokines, growth factors or transcription factors may be absent (Cohn and Adler, 1991). Furthermore, cells that are targeted in aerosol gene transfer for CF are ciliated airway epithelial cells, whereas A549 and LA-4 cell lines are derived from type II alveolar cells. It is possible that alternative cell lines derived from other cell types in the proximal lung would yield different results. One example is the NuLi cell line, which was

derived from normal human bronchial airway cells (Zabner et al., 2003). However, this line of investigation was not pursued.

Some of the complexity of the *in vivo* airway can be observed in ALI cell cultures (Table 3.1) specifically with regard to the formation of tight junctions, beating cilia and formation of mucus. A study comparing the transcriptome of ALI cultures and freshly-sampled human airway epithelial cells showed many similarities, although some differences were evident, possibly explained by the heterogeneity of cell types *in vivo* (Dvorak et al., 2011). ALI cultures can suffer from low transfection efficiency in the absence of a tight junction modulator (Chu et al., 1999), but one might argue that a low rate of transfection is also representative of the *in vivo* situation. Transfection efficiencies were increased in the presence of EGTA to disrupt tight junctions (Figure 3.8), causing the basolateral cell surface receptors to be exposed and accessible to uptake of non-viral vectors (Chu et al., 1999). The relative increase of transgene activity after disruption of tight junctions (63-fold, Figure 3.8) is in line with studies performed by Chu *et al.* where a 55-fold increase in activity of secreted alkaline phosphatase (SEAP) was observed after treatment of ALI cultures with EGTA when compared with non-treated cultures (Chu et al., 1999). However, in the studies described here (Figure 3.10), there was no sustained transgene activity observed to recapitulate results in the mouse lung after aerosolisation of pG4-hCEFI-soLux (Hyde et al., 2008). Interestingly, a study using the secreted Gaussia luciferase transgene driven by the hCEFI enhancer/promoter resulted in sustained levels of transgene activity for up to 27 days after transfection (Griesenbach et al., 2011). However, attempts to repeat these results were not successful (pers. comm. I. Pringle). One possibility for this discrepancy may have been due to the fact that a different batch of ALI cultures from a different donor were used or that the use of a secreted reporter gene may yield different results (Tannous, 2009). Variability between donors has been reported to be problematic when using ALI cultures (BeruBe et al., 2009), which could lead to results that may not necessarily be reproducible when a different donor source is used. Further studies using ALI cultures from a number of different donors

would be needed to investigate this. However, the cost of sufficient ALI cultures required to achieve statistical power prohibited this approach.

Ultimately, the reasons for the lack of transgene duration after non-viral transfection in cell lines or human primary ALI cultures were unclear. In order to further investigate the ability of the hCEFI enhancer/promoter to mediate persistent transgene expression, a different approach was undertaken, described in Chapter 4.

## Chapter 4: Investigation of lung transgene activity using bioluminescence imaging

### 4.1. Introduction

The studies described in Chapter 3 demonstrated that cell culture models of human and mouse lung cells did not recapitulate the transgene activity profiles observed *in vivo* after aerosol delivery of lipoplexes containing pG4-hCEFI-soLux or pG4-mCEFI-soLux (Hyde et al., 2008). A number of reservations were highlighted in Section 3.1 regarding the use of the mouse aerosol model to investigate transgene activity. However, due to the lack of a suitable cell culture model, the *in vivo* model was re-visited to understand the reasons for transgene activity.

One commonly utilised method for detecting luciferase activity following gene transfer to the mouse lung requires the sacrifice of the animal. This involves the homogenisation of the lung tissue and the addition of D-Luciferin substrate to the lung lysate in a luminometer to measure the amount of light emitted from the sample; the amount of light is reported as a value relative to the amount of protein in the sample. However, luciferase activity can also be measured by repeatable *in vivo* BLI; this non-destructive assay offers the advantage that data following a single treatment can be collected over time from one small group of animals. Additionally, transgene activity can be tracked in individual animals throughout the experiment, which is useful when monitoring the duration of transgene expression.

Bioluminescence imaging in living, adult animal models was first investigated by Contag *et al.*, using external detection of internal sources of light to assess the temporal and spatial expression of mammalian genes *in vivo*. With the delivery of D-Luciferin, *in vivo* luciferase activity could be non-invasively visualised in tissues of living laboratory animal models after gene delivery (Contag et al., 1997). Despite light scattering due to organs and fur, significant amounts of emitted light escapes and can still be detected externally (Contag and

Bachmann, 2002). Bioluminescence imaging is now widely used for detecting transgene expression after gene transfer in a variety of organs primarily because BLI analysis is fast, sensitive, and eliminates tissue harvesting (Rettig et al., 2006). The sensitivity of BLI is dependent on a number of factors, including the number of cells expressing the reporter gene and the efficiency of the promoter (as reviewed by Sato et al., 2004). Further studies have been performed to develop methods to increase detectable luminescence. For example, delivery of substrate via the intranasal route led to a two-fold increase in signal compared with traditional intraperitoneal injection when non-viral vectors were instilled to the mouse lung (Buckley et al., 2008), possibly due to a higher local concentration of the substrate in transfected lung epithelium.

Published studies using bioluminescence after nasal instillation have utilised pDNA vectors containing strong viral promoters, capable of expressing high levels of luciferase activity for detection. Griesenbach *et al.* confirmed that intranasal administration of D-Luciferin could result in higher luminescence compared with intraperitoneal delivery after aerosol dosing of GL67A complexed with pCIKLux (a plasmid containing the strong hCMV promoter) (Griesenbach et al., 2009). Aerosols using PEI as the gene transfer agent have also been shown to yield detectable levels of luminescence after dosing using pCpG-Luc, a plasmid containing the mCMV promoter (Geiger et al., 2010).

To investigate the reasons for differences in transgene activity level and persistence in the mouse lung after aerosol delivery of plasmids, deletions of the hC and mC regulatory sequences were prepared. Plasmids containing a variety of deletions were constructed (Section 2.3), delivered to the lungs of mice using aerosol delivery (Section 2.5.6) and luciferase activity was measured using BLI (Section 2.6.3). The aim was to determine if specific regions of the hC and mC enhancer sequences are responsible for the transgene activity profiles observed *in vivo* (Hyde et al., 2008).

The aims of the work presented in this chapter included: (i) to assess the feasibility of using BLI to detect *in vivo* transgene activity after instillation and aerosol delivery of formulations containing CpG-free pG4-mCEFI-soLux (Figure 4.1A) and pG4-hCEFI-soLux (Figure 4.1B); (ii) to compare the sensitivity of the measurement of luciferase activity obtained with BLI compared with standard luminometry in lung lysates; (iii) to investigate the effects of deletions of the hC enhancer (Figure 4.4) on luciferase activity using BLI and standard luminometry; and (iv) to investigate the effects of deletions of the mC enhancer (Figure 4.5) on luciferase activity using standard luminometry.

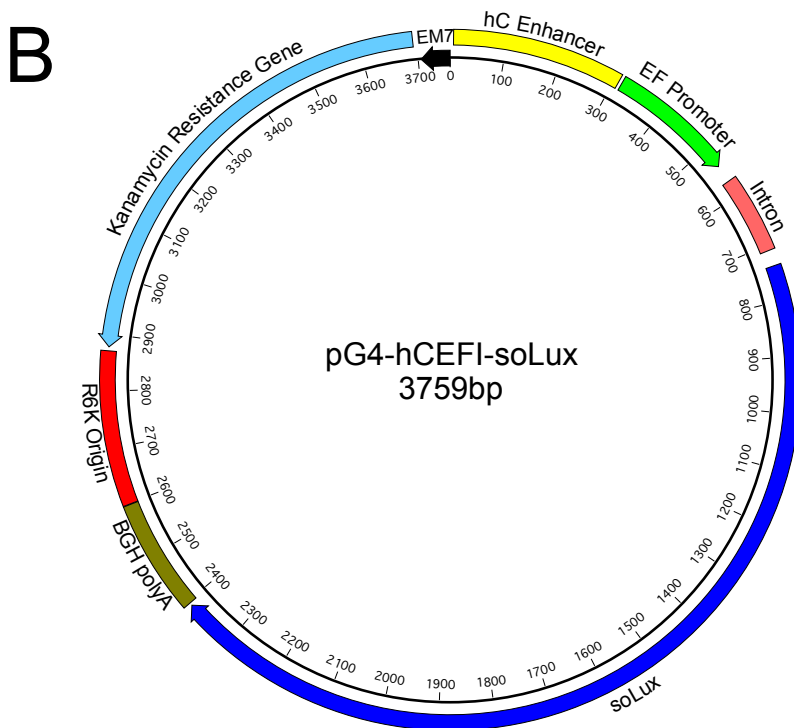
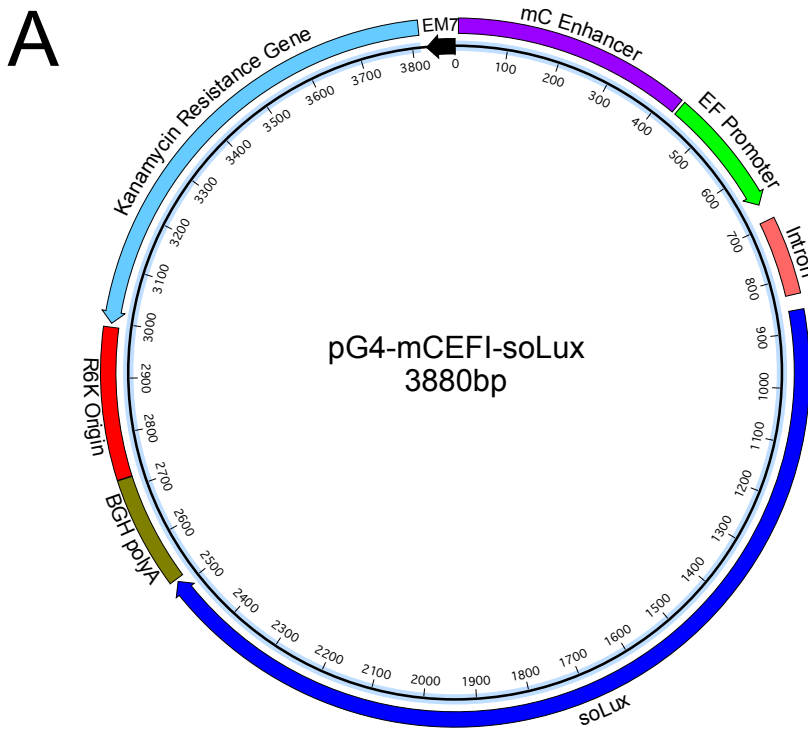
## 4.2. Results

### 4.2.1. Using BLI to detect luciferase activity after dosing with pG4-mCEFI-soLux

Gene delivery to the lung may be achieved via several administration routes, including intravenous injection, aerosol delivery or nasal instillation. The studies in this chapter have focused on aerosol delivery of PEI polyplexes and nasal instillation with lipoplexes of GL67A. Lung aerosol delivery is perhaps the delivery route most relevant for taking a formulation forward to the clinic. Both PEI and GL67A formulations yield similar levels of transgene activity in the mouse lung after aerosol delivery (Bazzani, 2011), but pragmatic reasons dictated the use of PEI for aerosol delivery; approximately 2 mg of pDNA is required for robust and detectable levels of gene transfer when PEI is used as the transfer agent (Gautam et al., 2000; Pringle et al., 2012a), whereas approximately 26 mg of pDNA is used for aerosol delivery with GL67A (Hyde et al., 2008). For nasal instillation, the use of PEI has previously been shown to result in increased levels of pro-inflammatory cytokines (Section 1.2.11) (Davies et al., 2012; Davies et al., 2008); therefore, GL67A liposomes were used for delivery of pDNA by nasal instillation, as it was previously shown to mediate gene transfer (Lee et al., 1996) with only a mild degree of toxicity which resolved within 14 days after dosing (Scheule et al., 1997) (Section 1.2.12). It has previously been shown that nasal instillation of pG4-mCEFI-soLux lipoplexes leads to higher transgene activity levels at early timepoints relative to aerosol dosing of polyplexes (pers. comm. I. Pringle); therefore, a formulation consisting of pG4-mCEFI-soLux (Figure 4.1A) complexed with GL67A was first used to assess whether BLI could be used to detect luciferase activity from CpG-free plasmids.

After nasal instillation of pG4-mCEFI-soLux/GL67A lipoplexes, a time-course experiment was performed to evaluate luciferase activity using BLI (Figure 4.2A) and standard luminometry (Figure 4.2B). These results showed that detection of transgene activity mediated by a synthetic CpG-free enhancer/promoter was possible using BLI. Furthermore, BLI yielded

similar expression profiles when compared with results obtained from standard luminometry. Luciferase activity was higher in mice that received the intranasal formulation (solid symbols) when compared to the aerosol formulation (open symbols). Transgene activity was lower at day 14 relative to day 1 values irrespective of the dosing method or the assay used. For example, when assessing luciferase expression by BLI, luciferase activity at day 14 was 1.6 % of day 1 values after instillation and 17.6 % of day 1 values after aerosol dosing. When standard luminometry was used, luciferase activity at day 14 was 1.4 % of day 1 values after instillation and 2.6 % of day 1 values after aerosol dosing. Although transgene activity profiles were similar between assays, standard luminometry was more sensitive. For example, luciferase activity from the aerosolised polyplexes at day 14 was 1.6-fold higher than naïve mice using BLI, but 15-fold higher when standard luminometry was used.

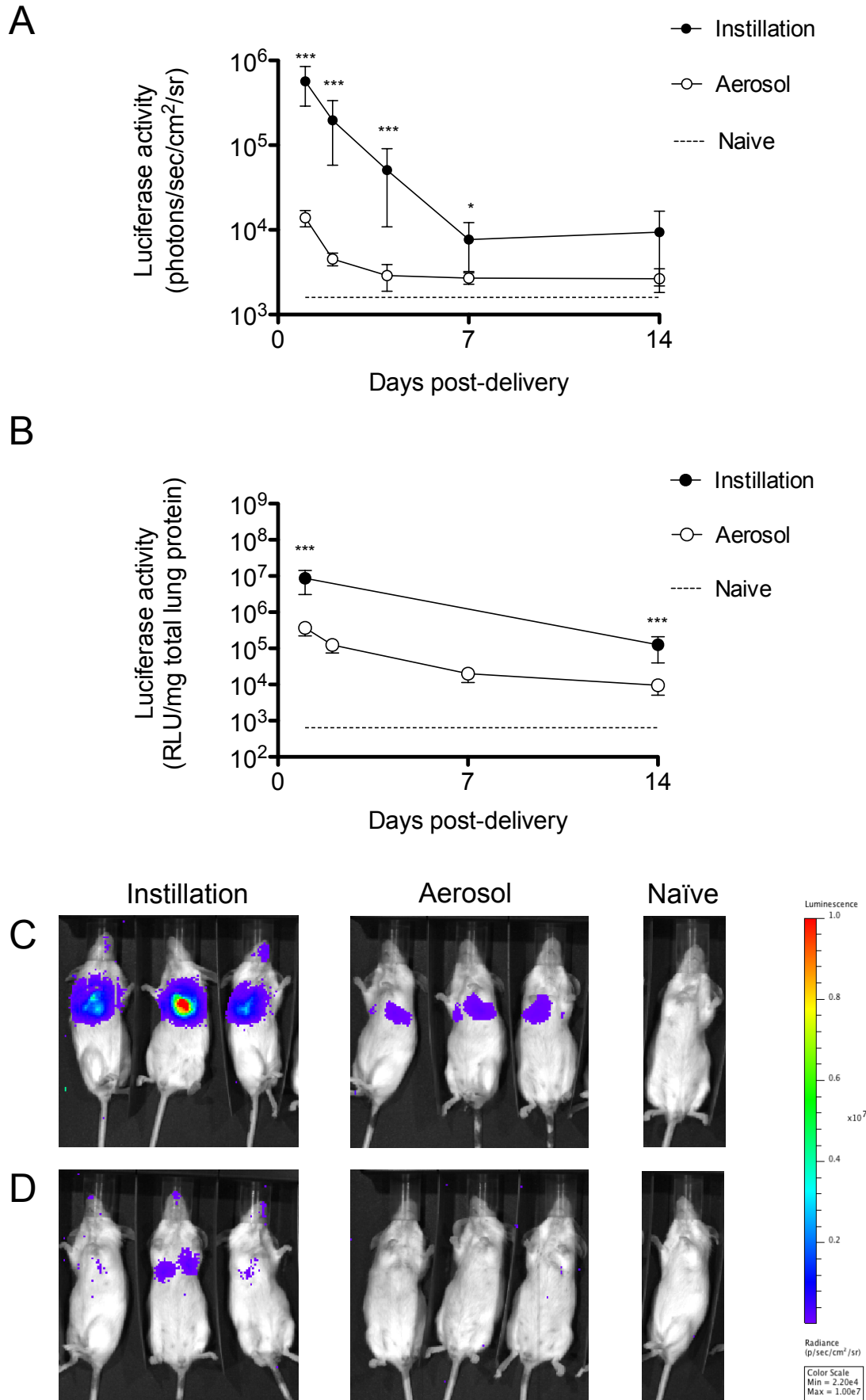


**Figure 4.1. Plasmids used in BLI studies.**

**A:** The pG4-mCEFI-soLux plasmid (Hyde et al., 2008; Table 2.1) is completely devoid of CpGs and contains the synthetic version of the mCMV (mC) enhancer, the human elongation factor 1 $\alpha$  (EF) promoter, intron, firefly luciferase (soLux), bovine growth hormone (BGH) polyA, R6K origin of replication and the kanamycin resistance gene.

**B:** The pG4-hCEFI-soLux plasmid (Table 2.1) is identical to pG4-mCEFI-soLux except that it contains the synthetic version of the hCMV (hC) enhancer rather than the mC enhancer. Both CpG-free plasmids were sourced from the Gene Medicine Group, University of Oxford (Hyde et al., 2008).

Size in bp is indicated beneath the plasmid name.

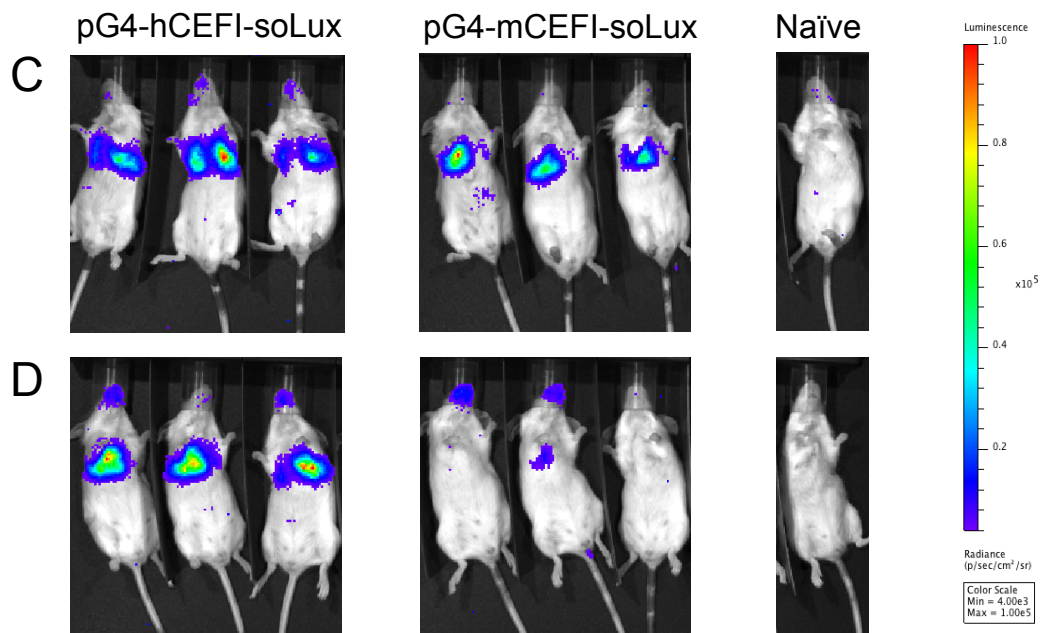
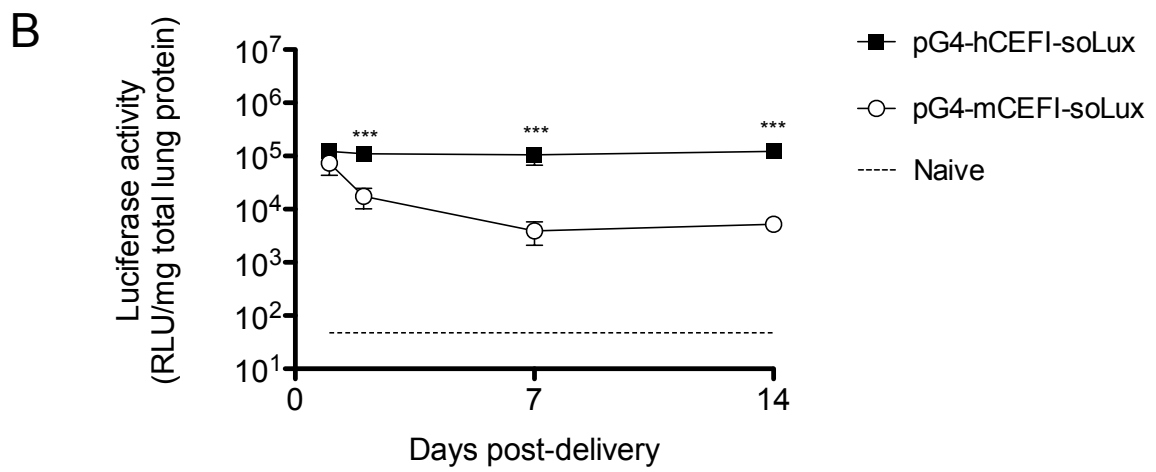
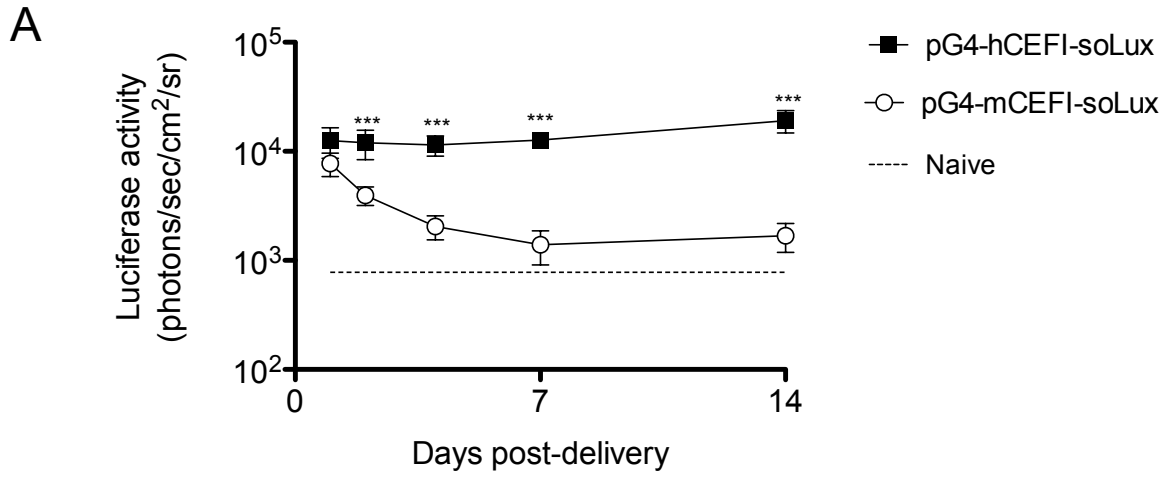


**Figure 4.2. Comparison of luciferase activity after dosing of pG4-mCEFI-soLux using (A) BLI or (B) standard luminometry.**

Female BALB/c mice were dosed with pG4-mCEFI-soLux (Figure 4.1A) by nasal instillation of 80 µg of pDNA complexed to GL67A (Section 2.5.3, n = 11) or aerosol delivery of 10 mL of 0.2 mg/mL pDNA complexed to PEI (Section 2.5.4, n = 30). Lung firefly luciferase activity was measured at the time-points indicated using (A) BLI (Section 2.6.3, n = 6) or (B) standard luminometry (Section 2.6.1, n = 5-6 per time-point). Data shown are mean ± SD. Statistical differences between conditions were analysed by two-way ANOVA with Bonferroni's multiple comparison post-hoc tests: \*\*\*  $p < 0.001$ , \*  $p < 0.05$ . Representative BLI images at (C) 1 day and (D) 14 days post-treatment are shown for three instilled animals, three aerosolised animals and a naïve mouse (no treatment). The colour bar represents relative light intensity from low (purple) to high (red).

#### 4.2.2. Comparison of luciferase activity of pG4-mCEFI-soLux and pG4-hCEFI-soLux

A time-course experiment was performed to compare luciferase activity following aerosol dosing of PEI complexes (Section 2.5.6) of pG4-hCEFI-soLux or pG4-mCEFI-soLux, measured using BLI (Section 2.6.3). Figure 4.3A shows that luciferase activity directed by the hCEFI enhancer/promoter is significantly higher than the mCEFI enhancer/promoter at day 2 until day 14 ( $p < 0.001$  at each time-point). The pattern of expression mediated by each plasmid was similar irrespective of using BLI or standard luminometry (Figure 4.3B); luciferase activity from pG4-hCEFI-soLux persisted at similar levels for the duration of the experiment whereas expression from pG4-mCEFI-soLux fell rapidly. Although the expression profiles were similar between assays (see Section 4.2.1), standard luminometry was more sensitive. Luciferase activity from pG4-mCEFI-soLux at day 14 was approximately 100-fold higher than naïve mice using luminometry, but only 2-fold higher when BLI was used, and luciferase activity from pG4-hCEFI-soLux at day 14 was approximately 2600-fold over naïve mice using luminometry, but only 20-fold higher when BLI was used. Importantly, the general reporter gene activity profiles observed in this experiment confirmed that BLI was suitable for measuring luciferase activity from both low- (e.g. pG4-mCEFI-soLux at day 14) and high-expressing (e.g. pG4-hCEFI-soLux at day 14) formulations. Furthermore, both assays demonstrate that the initial high transgene activity levels mediated by the mCEFI enhancer/promoter are not sustained.



**Figure 4.3. Comparison of luciferase activity after aerosolisation of pG4-mCEFI-soLux or pG4-hCEFI-soLux using (A) BLI or (B) standard luminometry.**

Female BALB/c mice (n = 30 per group) were dosed with 10 mL PEI polyplexes containing 2 mg of pG4-mCEFI-soLux (Figure 4.1A) or pG4-hCEFI-soLux (Figure 4.1B) by aerosol delivery (Section 2.5.6). Lung firefly luciferase activity was measured at the time-points indicated using (A) BLI (Section 2.6.3, n = 6) or (B) standard luminometry (Section 2.6.1, n = 6 per time-point). Data are shown as mean  $\pm$  SD. Statistical differences between conditions were analysed by two-way ANOVA with Bonferroni's multiple comparison post-hoc tests: \*\*\*  $p < 0.001$ . Representative BLI images at (C) 1 day and (D) 14 days post-treatment are shown for both plasmid treatments and a naïve mouse (no treatment). The colour bar represents relative light intensity from low (purple) to high (red).

### 4.2.3. Construction of plasmids encompassing enhancer deletions

Deletions encompassing approximately 70 bp starting from the 5'-end of the hC (Figure 4.4) or mC (Figure 4.5) enhancers were designed and constructed using standard techniques (Section 2.3). All plasmids contained the CpG-free versions of the human elongation factor 1 $\alpha$  (EF) promoter and firefly luciferase (soLux) transgene (Hyde et al., 2008).

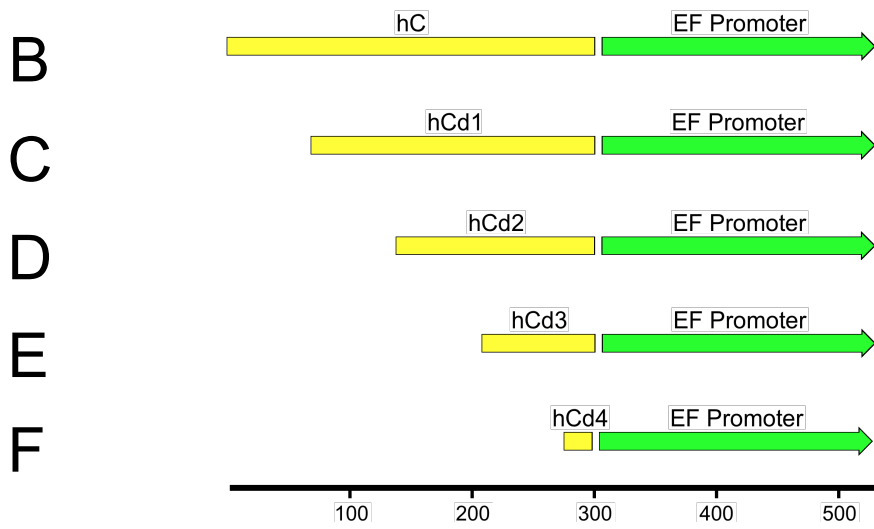
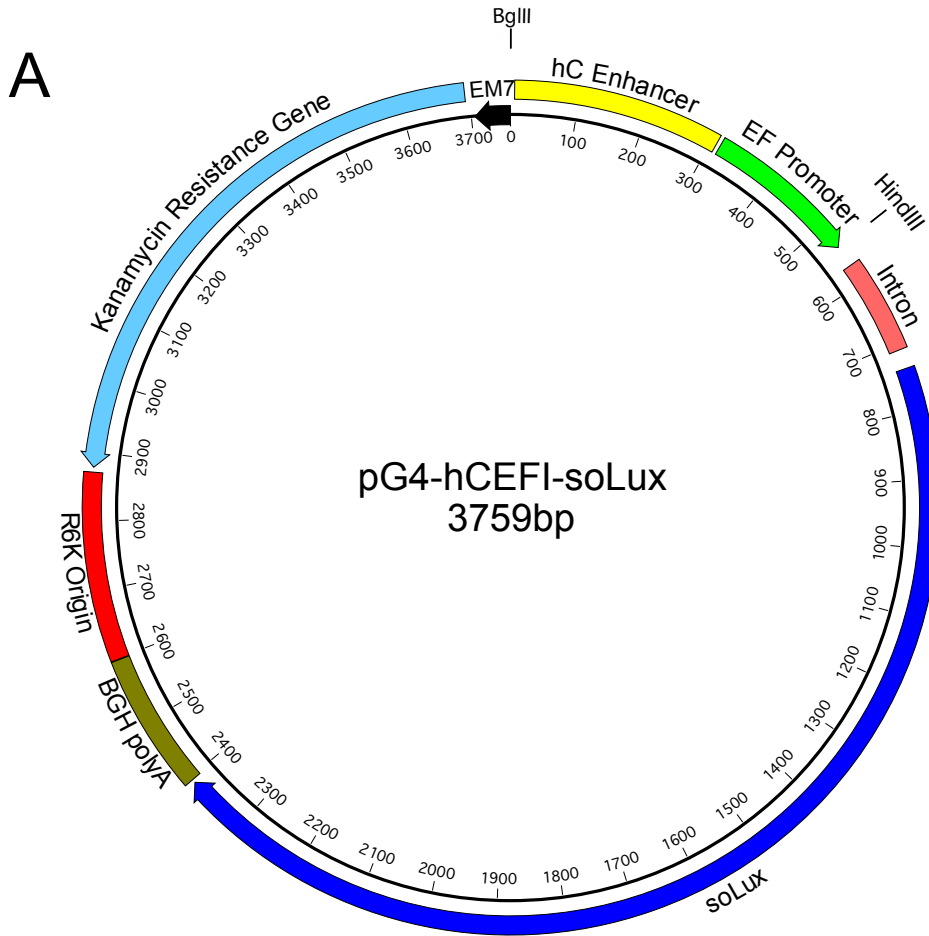
For each deletion variant, the fragment containing the CpG-free backbone, intron and the soLux transgene was obtained by digesting pG5-hCEFI-soLux (Table 2.1) with the restriction enzymes *Bgl*III and *Hind*III (Section 2.3.3), followed by gel electrophoresis to size-fractionate the fragments (Section 2.3.2) and subsequent extraction and purification from the gel (Section 2.3.4). The new hC (Figure 4.4C through F) and mC (Figure 4.5C through G) enhancer variants were constructed for these studies by PCR amplification of the desired fragments (Section 2.3.1). The new enhancer/promoter fragments were ligated to the extracted fragment at the *Bgl*III and *Hind*III sites (as shown in Figure 4.4A or Figure 4.5A) (Section 2.3.5). Following transformation of *E. coli* GT115 cells (Section 2.4.2), several colonies from each transformation were used to inoculate small-scale cultures (5 mL; Section 2.4.3), from which pDNA was purified (Section 2.4.4). Prior to large-scale (400 mL; Section 2.4.3) endotoxin-free plasmid production (Section 2.4.4), predominant monomeric plasmid topology was verified by agarose gel electrophoresis (Section 2.3.2) and the identity of the pDNA construct was confirmed by restriction enzyme digests (Section 2.3.3) followed by gel electrophoresis (data not shown). The new plasmid constructs (Table 4.1) were tested for their expression profile following aerosol delivery of polyplexes (Section 2.5.6) to the mouse lung. A plasmid containing no enhancer (pG4-EFI-soLux, Figure 4.6) was used as a negative control.

Table 4.1. Overview of plasmid name and enhancer length.

Plasmid name	Enhancer length	Figure reference
pG4-hCEFI-soLux <sup>1</sup>	302 bp	4.4A and B
pG5-hCd1EFI-soLux	233 bp	4.4C
pG5-hCd2EFI-soLux	163 bp	4.4D
pG5-hCd3EFI-soLux	93 bp	4.4E
pG5-hCd4EFI-soLux	23 bp	4.4F
pG4-mCEFI-soLux <sup>1</sup>	423 bp	4.5A and B
pG5-mCd1EFI-soLux	354 bp	4.5C
pG5-mCd2EFI-soLux	297 bp	4.5D
pG5-mCd3EFI-soLux	214 bp	4.5E
pG5-mCd4EFI-soLux	144 bp	4.5F
pG5-mCd5EFI-soLux	74 bp	4.5G
pG4-EFI-soLux <sup>1</sup>	0 bp	4.6

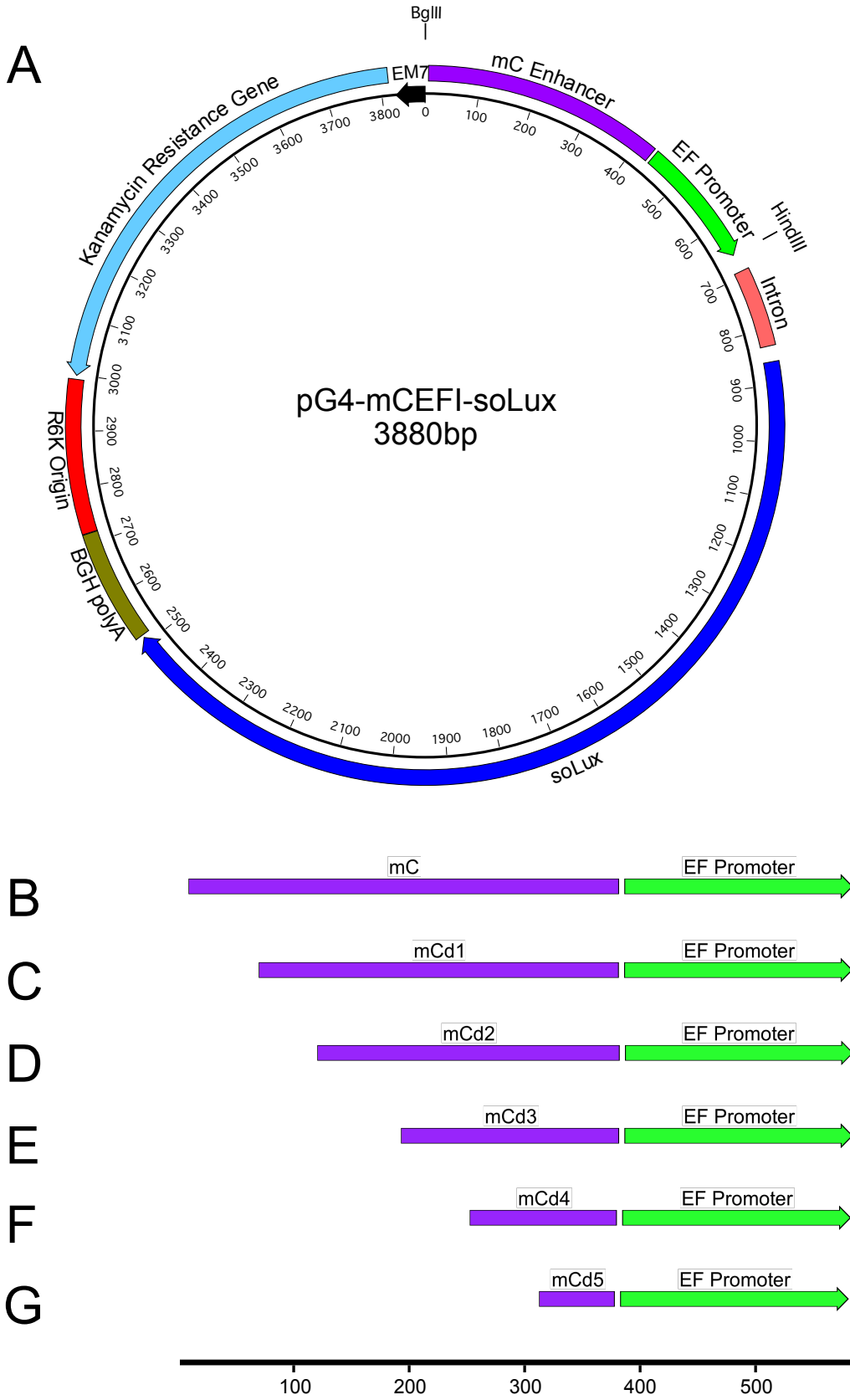
<sup>1</sup>Hyde *et al.*, 2008

The table lists the pDNA used in the enhancer deletion studies presented in this chapter. The plasmids were named according to the enhancer/promoter combination (where 'd' signifies deletion), and the transgene (soLux).



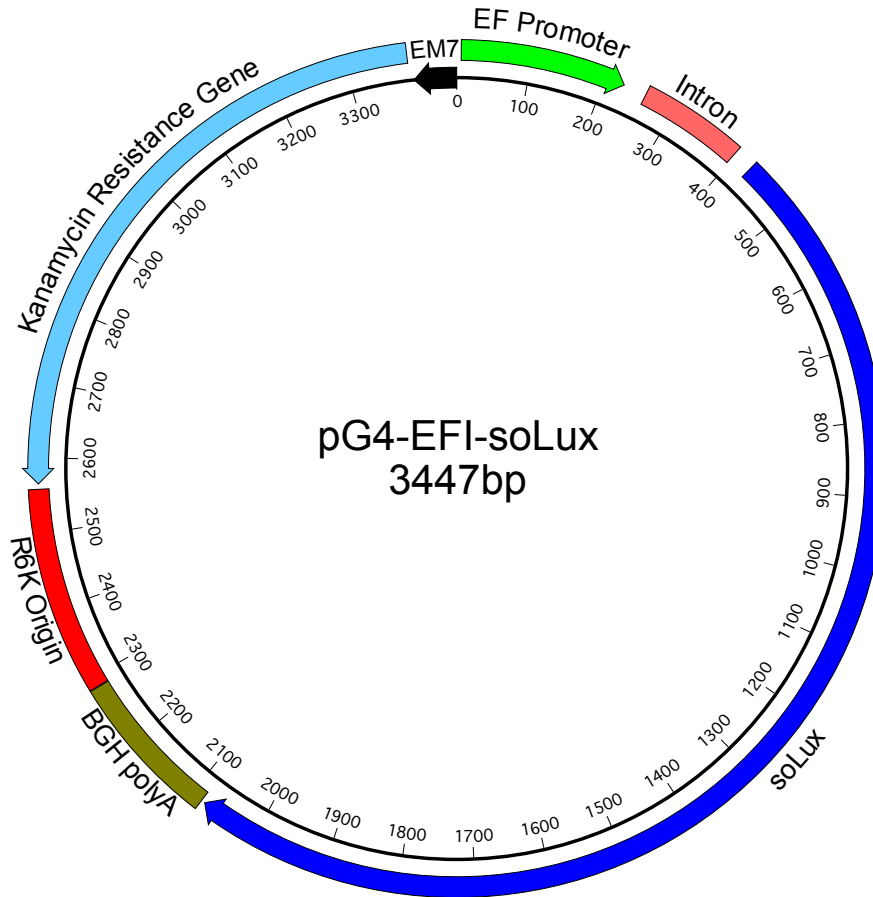
**Figure 4.4. Diagram of plasmid DNA constructs containing hC enhancer variants.**

Each CpG-free plasmid contains the synthetic version of the human elongation factor 1 $\alpha$  (EF) promoter (light green), intron (pink), firefly luciferase (soLux, dark blue), BGH polyA (dark green), R6K origin of replication (red) and the kanamycin resistance gene (light blue). The new plasmids were constructed by ligation of the relevant enhancer/promoter fragment at the *Bgl*II and *Hind*III sites as shown in Panel A. Each plasmid contains a different version of the synthetic CpG-free hCMV (hC) enhancer (C to F; yellow). A ruler is included as a visual aid to indicate the length of the enhancer/promoter region inserted between the *Bgl*II and *Hind*III restriction sites. The length of each enhancer is listed in Table 4.1.



**Figure 4.5. Diagram of plasmid DNA constructs containing mC enhancer variants.**

Each CpG-free plasmid contains the synthetic version of the human elongation factor 1 $\alpha$  (EF) promoter (light green), intron (pink), firefly luciferase (soLux, dark blue), BGH polyA (dark green), R6K origin of replication (red) and the kanamycin resistance gene (light blue). The new plasmids were constructed by ligation of the relevant enhancer/promoter fragment at the *Bgl*II and *Hind*III sites as shown in Panel A. Each plasmid contains a different version of the synthetic CpG-free mCMV (mC) enhancer (C to G; purple) A ruler is included as a visual aid to indicate the length of the enhancer/promoter region inserted between the *Bgl*II and *Hind*III restriction sites. The length of each enhancer is listed in Table 4.1.



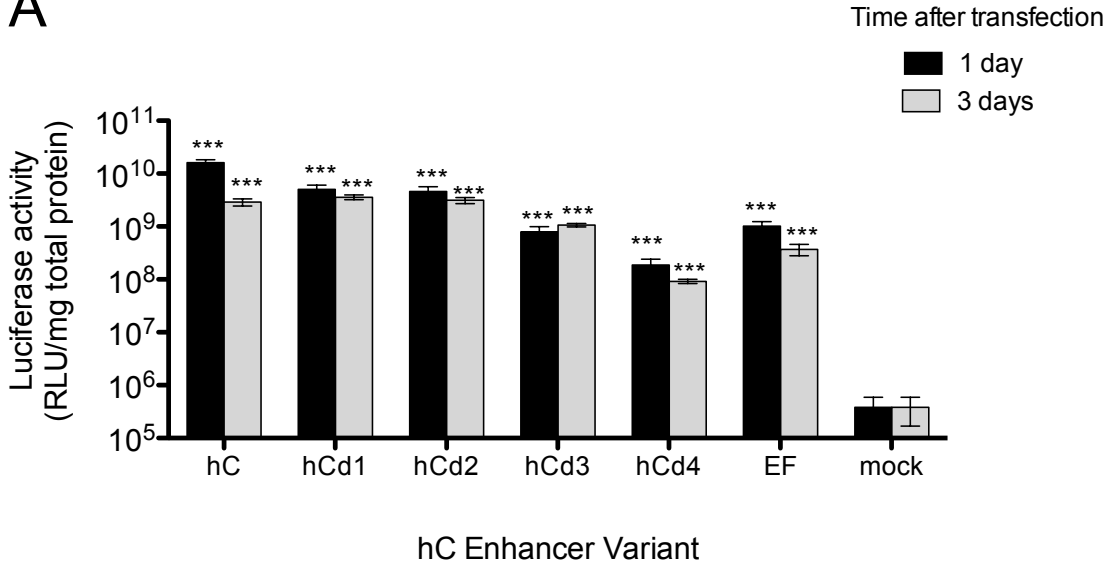
**Figure 4.6. Diagram of the plasmid DNA construct containing no enhancer.**

The CpG-free plasmid pG4-EFI-soLux contains the synthetic version of the human elongation factor 1 $\alpha$  (EF) promoter (light green), intron (pink), firefly luciferase (soLux, dark blue), BGH polyA (dark green), R6K origin of replication (red) and the kanamycin resistance gene (light blue).

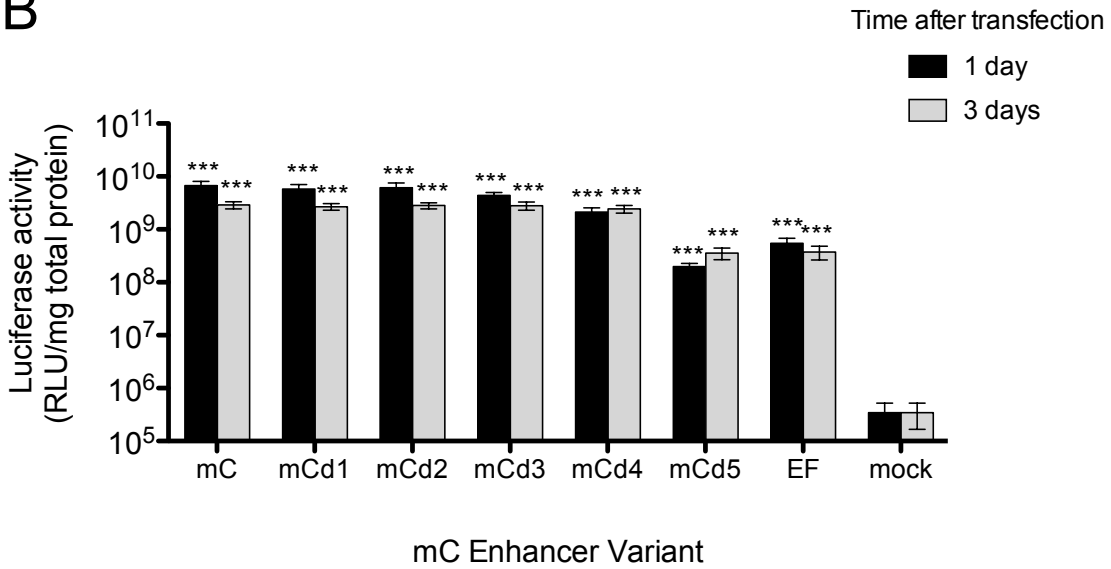
#### 4.2.4. Luciferase activity in A549 cells from plasmids containing enhancer variants

The plasmids listed in Table 4.1 were tested in A549 cells prior to dosing into mice to demonstrate that the new enhancer constructs could direct measurable luciferase activity when compared with mock-treated cells. Cells were transiently transfected with PEI polyplexes (Section 2.2.4) and luciferase activity was measured by standard luminometry (Section 2.6.1). As seen in Figure 4.7, all plasmids directed measurable levels of luciferase activity at both 1 day and 3 days after transfection. Transgene activity mediated by the plasmids containing the hC enhancer variants (Figure 4.7A) ranged from approximately 510- to 45,000-fold higher over the mock-treated cells at 1 day after transfection. At 3 days after transfection, this difference decreased to approximately 250- to 10,000-fold higher than the mock-treated cells. The plasmids containing the mC enhancer variants (Figure 4.7B) similarly directed measurable luciferase levels when compared with mock-treated cells, ranging from approximately 600- to 20,000-fold higher at 1 day after transfection, and 1,000- to 8,300-fold higher at day 3. Although A549 time-course experiments may not be indicative of luciferase activity in the mouse lung after aerosol dosing with PEI polyplexes (Section 3.2.2), these results demonstrate that measurable levels of luciferase can be detected and therefore, an *in vivo* study was warranted.

**A**



**B**



**Figure 4.7. A time-course experiment comparing luciferase activity in A549 cells after transfection with plasmids containing deletions of the (A) hC enhancer or (B) mC enhancer.**

Cultures of A549 cells (Section 2.2.1) were seeded at  $5 \times 10^4$  cells per well of a six-well plate and incubated overnight prior to transfection with the plasmids containing either (A) hC enhancer variant (Figure 4.4) or (B) mC enhancer variant (Figure 4.5) plasmids using PEI (Section 2.2.4). Further plasmid details can be found in Table 4.1. Five wells per condition were evaluated at the time-points indicated after transfection (Section 2.2.5). Luciferase activity was measured in cell lysates using standard luminometry methods (Section 2.6.1). Data are shown as mean  $\pm$  SD. Statistical significant was determined for values compared with mock-treated cells by one-way ANOVA analysis followed by Bonferroni's multiple comparison post-hoc tests; \*\*\*  $p < 0.001$ .

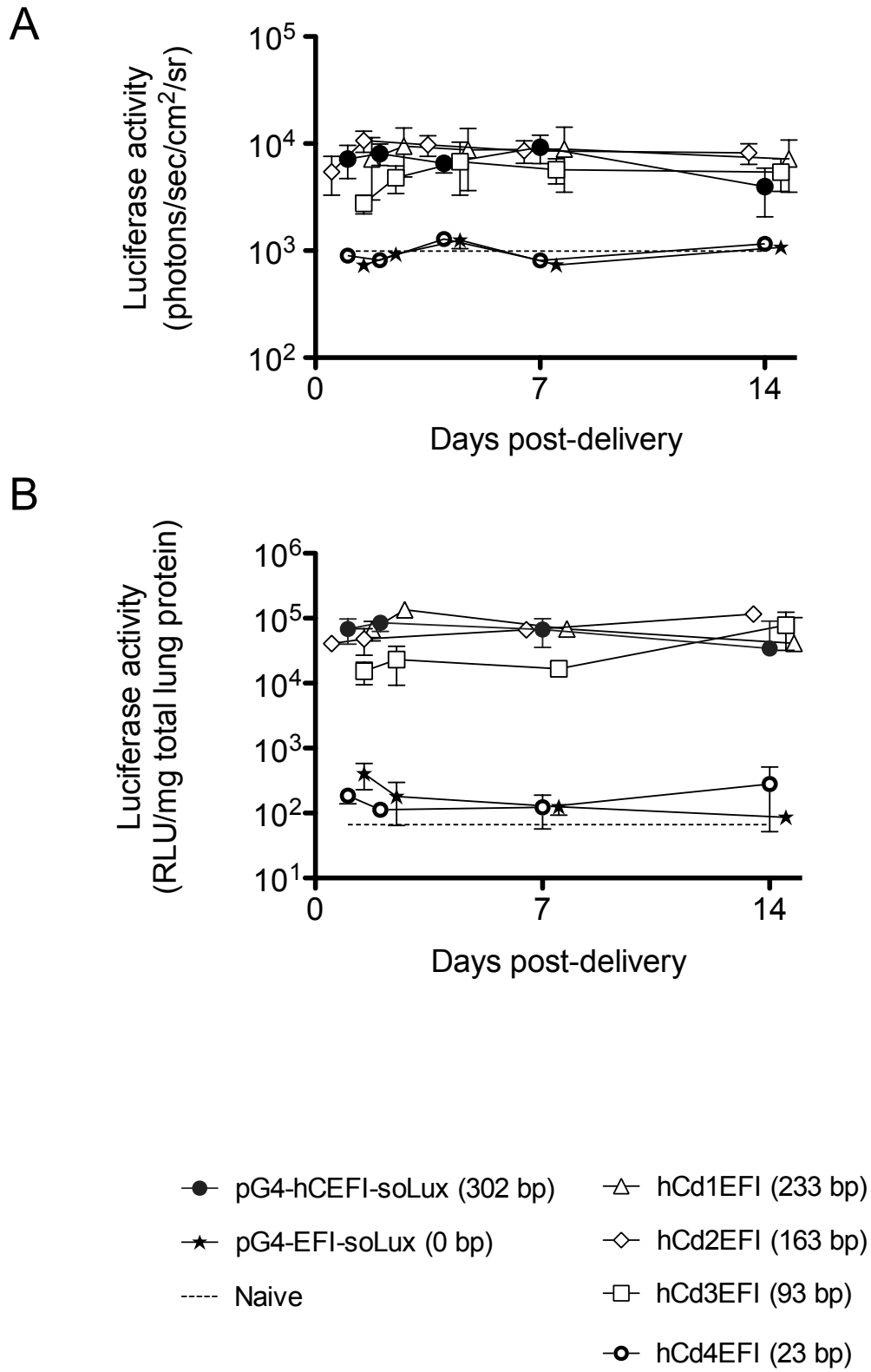
#### 4.2.5. Time-course studies of plasmids containing deletions of the hC enhancer in the mouse lung

An *in vivo* study was performed to compare luciferase activity after aerosol delivery of PEI complexes of hC-derived plasmids (Figure 4.4) and the negative control plasmid pG4-EFI-soLux (Figure 4.6) using bioluminescence imaging (BLI) (Section 2.6.3) and standard luminometry (Section 2.6.1). As expected, luciferase activity mediated by the hC enhancer as assessed by BLI resulted in high levels (approximately 65-fold over pG4-EFI-soLux) and remained high for the duration of the experiment (Figure 4.8A). Reporter gene activity mediated by the hCd4 enhancer variant was statistically different from pG4-hCEFI-soLux ( $p < 0.001$  at days 1, 2, 4 and 7,  $p < 0.05$  at day 14), but similar to pG4-EFI-soLux and naïve mice (no treatment) at all time-points. The three other enhancer variants tested (hCd1, hCd2 and hCd3) yielded similar activity levels to pG4-hCEFI-soLux, except for hCd3, which was significantly lower at day 1 ( $p < 0.05$ ) (Table 4.2A). Importantly, the plasmid containing the hCd3 enhancer variant was not statistically different to pG4-hCEFI-soLux at day 14.

Standard luminometry was also used to measure luciferase activity. Mice were sacrificed and the lungs were removed and homogenised (Section 2.5.8) and luciferase activity measured in lung lysates (Section 2.6.1) (Figure 4.8B). Overall, the reporter gene activity profiles were similar to data obtained using BLI (Figure 4.8A). Luciferase activity mediated by the hCd1 enhancer variant was similar to activity observed from pG4-hCEFI-soLux, while the other variants resulted in statistically significant differences (Table 4.2B). The plasmid containing the hCd3 enhancer variant resulted in a statistically significant difference at day 14 ( $p < 0.01$ ); however, this difference was due to higher luciferase activity mediated by the hCd3 enhancer. Transgene activity mediated by the hCd4 enhancer was similar to pG4-EFI-soLux and naïve mice (no treatment). As expected from results presented earlier (Figure 4.2 and Figure 4.3), the dynamic range of the standard luminometry assay was greater than that observed from BLI. For example, values at day 14 for pG4-hCEFI-soLux

using BLI were 23-fold over the values obtained from naïve mice, but 514-fold above the values obtained from naïve mice using standard luminometry. Despite this difference, the profiles were very similar, again demonstrating the utility of BLI when evaluating differences between high- and low-expressing plasmids.

The results of these studies in the mouse lung show that up to 209 bp of the hC enhancer can be removed (as in hCd3) without affecting transgene activity. Removal of a further 70 bp (as in hCd4) resulted in transgene activity levels similar to the negative control plasmid, pG4-EFI-soLux. These results suggest that this 70 bp region may be necessary and sufficient for high-level persistent transgene expression in the mouse lung.



**Figure 4.8. Luciferase activity after aerosol administration of plasmids containing deletions of the hC enhancer using (A) BLI or (B) standard luminometry.**

Female BALB/c mice (n = 18 per group) were exposed to 10 mL aerosols containing 2 mg of the respective plasmid complexed with PEI (Section 2.5.6). Firefly luciferase activity was analysed at the time-points indicated using (A) BLI (Section 2.6.3, n = 6) or (B) standard luminometry methods (Section 2.6.1, n = 3 per time-point). The enhancer length is listed in brackets. Data are shown as mean  $\pm$  SD. Data points have been shifted for clarity. For plasmid details, see Table 4.1, Figure 4.4 and Figure 4.6. For clarity, statistical analysis is presented separately in Table 4.2.

Table 4.2. Statistical information for Figure 4.8.

A

Plasmid name	Day 1	Day 2	Day 4	Day 7	Day 14
pG5-hd1CEFI-soLux	ns	ns	ns	ns	ns
pG5-hd2CEFI-soLux	ns	ns	ns	ns	ns
pG5-hd3CEFI-soLux	$p < 0.05$	ns	ns	ns	ns
pG5-hd4CEFI-soLux	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.01$
pG4-EFI-soLux	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$

B

Plasmid name	Day 1	Day 2	Day 7	Day 14
pG5-hCd1EFI-soLux	ns	ns	ns	ns
pG5-hCd2EFI-soLux	ns	ns	ns	$p < 0.001$
pG5-hCd3EFI-soLux	ns	ns	ns	$p < 0.01$
pG5-hCd4EFI-soLux	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
pG4-EFI-soLux	$p < 0.01$	$p < 0.001$	$p < 0.001$	$p < 0.001$

Statistical significance was determined for values compared with pG4-hCEFI-soLux by two-way ANOVA followed by Bonferroni's multiple comparison post-hoc tests for results using (A) BLI or (B) standard luminometry; ns =  $p > 0.05$ . For plasmid information, see Table 4.1, Figure 4.4 and Figure 4.6.

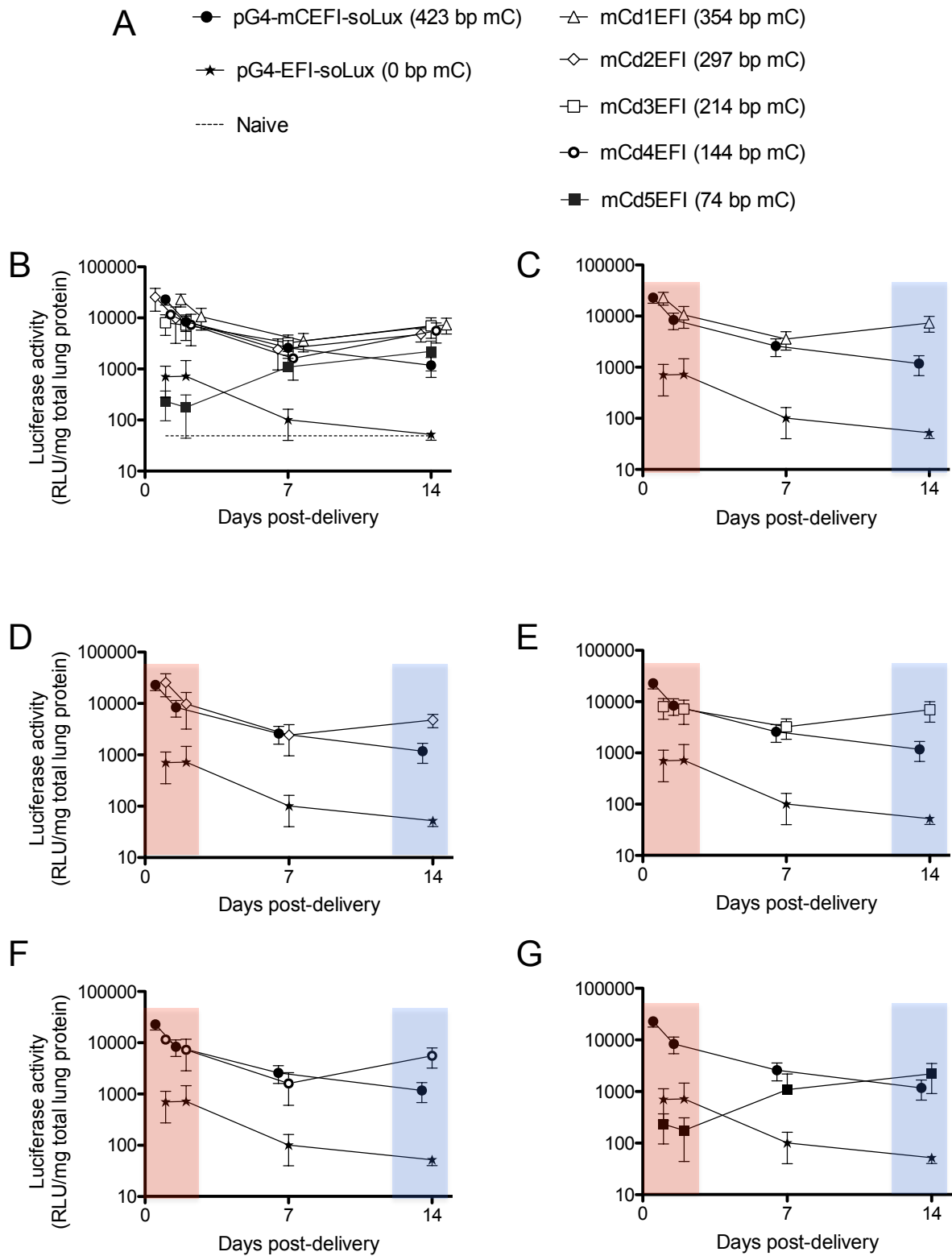
#### 4.2.6. Time-course studies of plasmids containing deletions of the mC enhancer in the mouse lung

To further understand how the mC enhancer sequence affects transgene activity in the lungs of mice, an *in vivo* study was performed to compare luciferase activity after aerosol delivery of PEI complexes of mC-derived plasmids (Figure 4.5) and the negative control plasmid, pG4-EFI-soLux (Figure 4.6). Bioluminescence imaging was not performed for this set of *in vivo* experiments due to the limited dynamic range of the assay when transgene activity is relatively low, as was expected for the mC enhancer at later time-points (Figure 4.3A); therefore, standard luminometry was used to measure the expression profile from these plasmids.

Luciferase activity mediated by the mC enhancer at day 14 fell approximately 17-fold when compared with peak values at day 1 (Figure 4.9B). However, a number of the mC enhancer variants exhibited higher luciferase activity values at day 14 than the full-length mC enhancer. For example, the mCd1 variant, deleting 69 bp of the mC enhancer (Figure 4.9C), yielded luciferase values similar to mC at days 1, 2 and 7 ( $p > 0.05$ , Table 4.3), but was approximately 6-fold higher at day 14 (highlighted in blue areas) ( $p < 0.001$ , Figure 4.9C). Similar results were observed with mCd2 (4-fold higher,  $p < 0.001$ , Figure 4.9D), mCd3 (6-fold higher,  $p < 0.001$ , Figure 4.9E), mCd4 (5-fold higher,  $p < 0.001$ , Figure 4.9F), and mCd5 (2-fold higher,  $p > 0.05$ , Figure 4.9G).

Intriguingly, a deletion of 209 bp (as in mCd3) led to stable transgene activity levels (Figure 4.9E) at early time-points (highlighted in the red areas), rather than the approximate 3-fold reduction of luciferase activity observed with mC (Figure 4.9B), mCd1 (Figure 4.9C) and mCd2 (Figure 4.9D). Furthermore, despite lower luciferase activity at early time-points mediated by mCd5 (Figure 4.9G, red area,  $p < 0.001$ ), the values observed at day 14 (blue area) were very similar to the reporter gene level mediated by mC ( $p > 0.05$ ).

The results of this study show that a large portion of the mC enhancer (up to 349 bp, as in mCd5) can be removed without leading to a detrimental effect on transgene activity at day 14. However, luciferase activity observed at early time-points was affected by the deletions.



**Figure 4.9. Luciferase activity after aerosol administration of enhancer variants of pG4-mCEFI-soLux variants using standard luminometry.**

Female BALB/c mice (n = 24 per group) were exposed to 10 mL aerosols containing 2 mg of the respective plasmid complexed with PEI (Section 2.5.6). Firefly luciferase activity was analysed at the time-points indicated using standard luminometry (Section 2.6.1, n = 6 per time-point). Enhancer length is listed in brackets. Data are shown as mean  $\pm$  SD. Data points have been shifted for clarity. The red areas highlight the differences in transgene activity between days 1 and 2, while the blue areas highlight the differences between the plasmids at day 14. For plasmid details, see Table 4.1, Figure 4.5 and Figure 4.6. Statistical analysis is presented separately in Table 4.3.

**A:** Figure legend for all subsequent graphs for Figure 4.9.

**B:** Transgene activity data for all plasmids tested. The data presented in this panel have been extracted and partitioned to compare luciferase activity from pG4-mCEFI-soLux, pG4-EFI-soLux and each individual mC enhancer variant, as follows:

**C:** mCd1EFI

**D:** mCd2EFI

**E:** mCd3EFI

**F:** mCd4EFI

**G:** mCd5EFI

**Table 4.3. Statistical information for Figure 4.9.**

Plasmid name	Day 1	Day 2	Day 7	Day 14
pG5-mCd1EFI-soLux	ns	ns	ns	$p < 0.001$
pG5-mCd2EFI-soLux	ns	ns	ns	$p < 0.001$
pG5-mCd3EFI-soLux	$p < 0.01$	ns	ns	$p < 0.001$
pG5-mCd4EFI-soLux	ns	ns	ns	$p < 0.001$
pG5-mCd5EFI-soLux	$p < 0.001$	$p < 0.001$	$p < 0.001$	ns
pG4-EFI-soLux	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$

Statistical significance was determined for values compared with pG4-mCEFI-soLux by two-way ANOVA followed by Bonferroni's multiple comparison post-hoc tests; ns =  $p > 0.05$ .

For plasmid information, see Table 4.1, Figure 4.5 and Figure 4.6.

### 4.3. Discussion

#### 4.3.1. Use of bioluminescence imaging after delivery of non-viral vectors to the lung

Studies in this chapter investigated the use of BLI to measure luciferase activity after delivery of non-viral formulations to the mouse lung. Published studies have shown good detection of luciferase activity using BLI, but only with strong viral promoters (Geiger et al., 2010; Griesenbach et al., 2009). Intranasal instillation of GL67A complexes of pG4-mCEFI-soLux resulted in high levels of luciferase activity when compared with aerosol administration of PEI polyplexes (Section 4.2.2). This may partly be due to the fact that the instillation protocol utilised in these studies delivers 80  $\mu\text{g}$  of pDNA whereas an average aerosol deposits only approximately 0.2 % of the total nebulised dose (i.e. estimated lung deposition of 4  $\mu\text{g}$  pDNA/mouse) (Koshkina et al., 2003). Encouragingly, the expression profile obtained by BLI was similar to the data obtained using standard luminometry regardless of the method of plasmid administration (Figure 4.2). Detection of persistence of reporter gene activity after aerosol administration of both pG4-hCEFI-soLux and pG4-mCEFI-soLux was also possible using BLI. Again, the resulting expression profiles from BLI were similar to those obtained from standard luminometry (Figure 4.3).

#### 4.3.2. Dynamic range of BLI compared with standard luminometry

The data presented in this chapter demonstrated the feasibility of using BLI as a method for detecting luciferase activity in the mouse lung after aerosol administration of CpG-free non-viral formulations, a result that has not been reported previously. Despite the similarities in expression profiles, the sensitivity of BLI was lower than standard luminometry (Sections 4.2.1 and 4.2.2). This may be due to a number of issues, including: (i) organ depth, (ii) anaesthetic usage and (iii) the relative efficiencies of each assay. The transmission efficiency of light emitted by firefly luciferase after addition of luciferin *in vivo* in a deep organ such as the lung has been estimated to be approximately 2.5 % (Zhao et al., 2005). This may be due to photon scattering at the air/alveolar junction, therefore leading to a reduction in the

signal intensity detected at the surface (Zhao et al., 2005). Additionally, animals must be anaesthetised for the duration of the image-capture process. The imaging system used in these studies delivers isoflurane through nose cones to the animal to ensure the mouse remains still in order to allow accurate detection of light from a particular area. It has been shown that isoflurane binds at the same active site as the luciferin substrate and thus, the anaesthetic may decrease the capacity of luciferin to bind to the luciferase enzyme (Zhang et al., 2001). Therefore, the possibility cannot be excluded that luciferase activity was somewhat reduced by isoflurane. The level of anaesthetic was kept at levels as low as possible for induction of anaesthesia (4 %) and maintenance during imaging (3 %) to minimise animal movement during the imaging period. In terms of assay efficiency of standard luminometry, the whole lung is removed from the animal and total luciferase activity is detected in the lysate regardless of which lung cells have been transfected. Comparing this method to BLI where the lung remains intact *in situ*, where the substrate must enter the transfected cells, and where photons must pass through layers of tissue, skin and fur, the probability of detecting the full signal from luciferase activity is much higher when using standard luminometry. Despite these obstacles, robust persistent luciferase signal was detected after aerosol delivery of the CpG-free plasmids described in this chapter (Figure 4.3).

Conversely, a number of factors that were inherent in the experimental design helped to obtain a strong bioluminescent signal from the mouse lung, which may have mitigated a number of the drawbacks listed above. These factors included the use of white mice as the animal model and the use of firefly luciferase as the transgene. Light is transmitted more efficiently through the tissues of white or hairless mice because melanin absorbs substantial amounts of light and light is scattered more by dark fur (Contag and Bachmann, 2002). It has also been shown that the light signal from firefly luciferase is less likely to be attenuated due to absorption by haemoglobin when compared with *Renilla* or beetle luciferases (Zhao et al., 2005). The traditional route of administration of the D-Luciferin substrate is typically via the

intraperitoneal cavity. However, the studies in this chapter delivered the substrate via the intranasal route due to the published finding (Buckley et al., 2008) that luciferase activity after intranasal administration to the mouse lung was two-fold higher using BLI when compared with the intraperitoneal route. This is presumably due to the fact that the substrate reaches a higher local concentration which does not limit light emission (Lee et al., 2003).

#### **4.3.3. Variant upstream regulatory elements and their effects on transgene activity**

The development of CpG-free plasmids, and in particular, the synthetic novel hCEFI enhancer/promoter, has been critical in developing a non-viral gene transfer formulation suitable for the treatment of chronic airway disease, and in particular, cystic fibrosis. Overall, the differences in lung transgene activity observed after aerosol delivery of plasmids containing the full-length hC or mC (Figure 4.3) enhancer can be described in terms of both the level and the persistence of reporter gene activity. Generating a better understanding of the reasons for these differences may aid the development of novel enhancer regulatory sequences which could be applied to both viral and non-viral gene transfer approaches.

Plasmids containing deletions of the enhancer sequences were constructed, modifying the length of the upstream regulatory enhancer element, in order to identify how transcription is affected by the presence or absence of particular sequences. After aerosol delivery of the plasmid/PEI polyplexes to the mouse lung, the minimal sequence elements required in the hC (Figure 4.8) and mC (Figure 4.9) enhancers affecting persistence were identified; here, persistence is defined as maintenance of transgene activity when compared with day 1 values. In the case of hC, the 3' 93 bp variant (as seen in hCd3) resulted in similar transgene activity at day 14 versus day 1 (Figure 4.8,  $p > 0.05$ ). Additionally, the reporter gene levels were at the same high level as the full-length hC enhancer. Similarly, the mCd3 enhancer variant (Figure 4.9E) was able to maintain day 1 luciferase activity for the duration of the study ( $p > 0.05$ ). These results indicate that persistence is possible with variants of either the hC or mC enhancer. There may be a specific sequence in the hC enhancer that confers the

high-level transgene activity observed when compared with the mC enhancer. However, transgene activity directed by enhancer variant hCd3 was approximately 1-log greater than the luciferase activity mediated by the highest-expressing mC enhancer variant (mCd3). Therefore, derivatives of the mC enhancer may have limited utility for airway gene transfer protocols.

Interestingly, the level of transgene activity at day 14 mediated by the mC enhancer variants (Figure 4.9, blue areas) is similar to, or higher than, activity mediated by the full-length mC enhancer regardless of the size of the deletion or whether luciferase activity was relatively low at early time-points (as per mCd5, Figure 4.9G). This observation may indicate that the sequence required for overall level of activity is not dependent on the presence of the 5'-end of the mC enhancer, but that the upstream sequence may be exerting a negative influence on transgene activity at early (Figure 4.9, red areas) or late (Figure 4.9, blue areas) time-points. It would be interesting to determine luciferase activity at later time-points (> 14 days), as expression mediated by the mCd5 variant may continue its apparent upward trend (Figure 4.9G). If the reporter gene levels from mCd5 could eventually achieve the high-level of activity mediated by the hC enhancer, this result would indicate another CpG-free enhancer sequence that could be used for directing transgene high-level activity in the mouse lung. However, this line of study was not pursued.

### 4.3.4. Conclusions

The studies in this chapter have demonstrated that BLI can be used to follow luciferase activity after instillation and aerosol delivery of non-viral vectors. The use of BLI leads to a reduction in the usage of mice and overall cost of the experiment and is an appropriate tool to examine vector expression profiles. While less sensitive than standard luminometry, BLI can still reliably differentiate between high- and low-expressing pDNA as is the case with pG4-hCEFI-soLux and pG4-mCEFI-soLux, respectively. Further studies have shown that by constructing deletion variants of the hC and mC enhancers, particular sequences can be

identified which may play a role in directing transgene activity in the mouse lung after gene delivery. One potential explanation for this observation is that the presence of particular transcription factor binding sites may play a role in the repression or activation of transcription of the transgene (see Section 1.4.4). Studies in Chapter 5 address this hypothesis.

## Chapter 5: Transcription factor binding site analysis of the hC and mC enhancer variants

### 5.1. Introduction

The studies in Chapter 4 identified variants of the hC enhancer that were able to mediate transgene activity at similar levels to the full-length hC enhancer after aerosol delivery of plasmid/PEI polyplexes to the mouse lung (Figure 4.8). Interestingly, only the final 93 bp of the 302 bp hC enhancer (as in hCd3) were required for directing persistent luciferase activity. A further observation was that a large portion of the mC enhancer could be removed without reducing the level of luciferase activity at day 14 (Figure 4.9). The identification of these minimal enhancer sequences have formed the basis for further investigations in order to understand the reasons for the differential hC and mC enhancer activity in the mouse lung.

Promoter and enhancer elements in plasmid constructs are critical for generating transgene expression (see Section 1.4.4). Enhancers are regulatory elements that act *in cis* and contain binding sites for transcription factors, which can be involved in repression or activation of transcription. Enhancers operate at a distance from a promoter, irrespective of position or orientation. The function of enhancers can be identified through chromatin-based assays such as chromatin immunoprecipitation (ChIP), which can evaluate the association of the enhancer with transcription factors (TFs) or specific epigenetic marks (reviewed in Lam et al., 2014). Due to the relatively low levels of recoverable pDNA from the lung after aerosol delivery, the use of standard laboratory techniques such as ChIP is highly challenging (pers. comm. I. Pringle). Furthermore, the pDNA isolated under such circumstances is likely to be a combination of both transcriptionally active and inactive pDNA molecules, and it may be difficult to correlate these with subsequent transgene activity. Instead, the identification of transcription factor binding sites (TFBSs) by computer-based methods was used.

A number of databases are available for TFBS analysis. For example, the MacVector software (MacVector, Cary, NC, USA) utilises a database maintained by the Institute for

Transcriptional Informatics, which predicts putative TFBSs based on consensus IUPAC (International Union of Pure and Applied Chemistry) strings and nucleotide ambiguity codes (Ghosh, 1990; Ghosh, 2000). This dataset encompasses all published literature on single TFBSs. An alternative method of detecting TFBSs is by using position weight matrices (PWMs). Using the PWM method can lead to more sensitive and precise TFBS predictions than the best IUPAC consensus sequence (Stormo, 2000). The Genomatix Software Suite (Genomatix Software, Munich, Germany) utilises the PWM approach and, additionally, refers to a number of libraries in the EIDorado database which represent the best of current knowledge in terms of specificity and sensitivity of the resulting PWM, rather than representing all available literature; thus, putative erroneous binding sites are minimised in the resultant output (Cartharius et al., 2005).

The aims of the work presented in this chapter were: (i) to identify putative TFBSs of the hC- and mC-derived enhancer variants; and (ii) to compare and contrast the different TFBSs to generate hypotheses to explain why the hC and mC enhancers result in different transcriptional activity profiles.

## 5.2. Results

### 5.2.1. Identification of predicted transcription factor binding sites using MacVector

The presence of TFBSs may play a role in the repression or activation of transcription of a transgene by recruiting factors that can modify chromatin structure and/or aid in the binding of general transcription factors to form the pre-initiation complex (see Section 1.4.4). The mammalian transcription factor database utilised by the MacVector software v12.5.1 (Ghosh, 1990; Ghosh, 2000) was used to predict TFBSs present in the hC and mC enhancers, and each of the enhancer variants used in the studies in Chapter 4 (Table 4.1). Overall, the hC and mC sequences are approximately 39 % identical (Figure 5.1). Table 5.1 shows the total number of individual binding sites, as well as the number of different binding sites. The full-length hC enhancer contains 215 total predicted TFBSs, comprising 104 different sites, while the mC enhancer contains 405 total binding sites, comprising 98 different sites. Of the different sites, 66 (49 %) are shared in the mC and hC enhancers; 31 (23 %) are unique to mC while 38 (28 %) are unique to hC. The names of the different sites as listed by the MacVector software are detailed in Table 5.2.

As described in Section 5.1, the database which MacVector uses for identification of TFBS is based on consensus IUPAC sequences, and in addition, represents all literature regarding a single binding site, in the absence of further experimentally-validated information. Therefore, alternative software was used for further investigations.

```

hC   1 -----GTTACATAACTTATGGTAAATGGCCTGCCTGGCT---GA 36
mC   1 GAGTCAATGGGAAAAACCCATTGGAGCCAAGTACTGACTCAATAGGGA 50
      * * * * * * * * * * * * * * * * * * * * * *

hC   37 CTGCCAATGACCCCTGCCA----ATGATGTCAATAATGATGTATGTT 82
mC   51 CTTTCCATTGGGTTTTGCCAGTACATAAGGTCAATAG-GGGGTGAGTCA 99
      ** ** * * * * * * * * * * * * * * * * * * * *

hC   83 CCATGTAATGCCAATAGGGACTT--TCCATTGATGTCAATGGGTGGAGTA 130
mC  100 ACAGGAAAGTCCCATTGGAGCCAAGTACATTGA-GTCAATAGG-GACTTT 147
      * * * * * * * * * * * * * * * * * * * * * *

hC   131 TTTATGGTAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAA- 179
mC   148 CCAATGGGTTTTGCCAGT--ACATAAGGTCAA-TGGGAGGTAAGCCAAT 194
      **** * * * * * * * * * * * * * * * * * * * *

hC   180 GTATGCCCCCTATTGATGTCAATGATGGTAAAT-----GGCCTGCCTGG 223
mC   195 GGGTTTTTCCCACTACTGACATGTATACTGAGTCATTAGGGACTTTCCAA 244
      * * * * * * * * * * * * * * * * * * * * *

hC   224 CA--TTATGCCAGTACATGACCTTAT--GGGACTTTCCTACTTGGCAGT 269
mC   245 TGGGTTTTGCCAGTACATAAGGTCAATAGGGGTGAATCAACAGGAAAGT 294
      ** * * * * * * * * * * * * * * * * * * * * *

hC   270 ACATCTATGTATTAGT-CATTGCTATTACCATGG----- 302
mC   295 CCCATTGGAGCCAAGTACTGA-GTCAATAGGGACTTTCATTGGGTTT 343
      * * * * * * * * * * * * * * * * * * * * *

hC   303 ----- 302
mC   344 TGCCCAGTACAAAAGGTCAATAGGGGGTGAAGTCAATGGGTTTTTCCATT 393

hC   303 ----- 302
mC   394 ATTGGCACATACATAAGGTCAATAGGGGTG 423

```

**Figure 5.1. Sequence alignment of the hC and mC enhancer sequences.**

The sequences of the full-length hC (302 bp) and mC (423 bp) enhancers are listed. Sequence identity between the two sequences is indicated (\*).

**Table 5.1. Number of potential transcription factor binding sites identified by MacVector.**

Enhancer name	Enhancer length (bp)	Total binding sites	Different binding sites
hC	302	215	104
hCd1	233	158	100
hCd2	163	87	58
hCd3	93	50	42
hCd4	23	10	10
mC	423	405	98
mCd1	354	335	95
mCd2	297	288	93
mCd3	214	207	87
mCd4	144	143	45
mCd5	74	59	44

The enhancer sequences were queried against the mammalian transcription factor database used by the MacVector software (v12.5.1). The results were manually tallied for total binding sites present within the different sequences. The output was changed to detail all sites by name, the results of which were manually tallied to identify the number of different binding sites present in the enhancers. See Figures 4.4 and 4.5 for plasmid diagrams.

**Table 5.2. Output names of potential transcription factor binding sites identified by MacVector.**

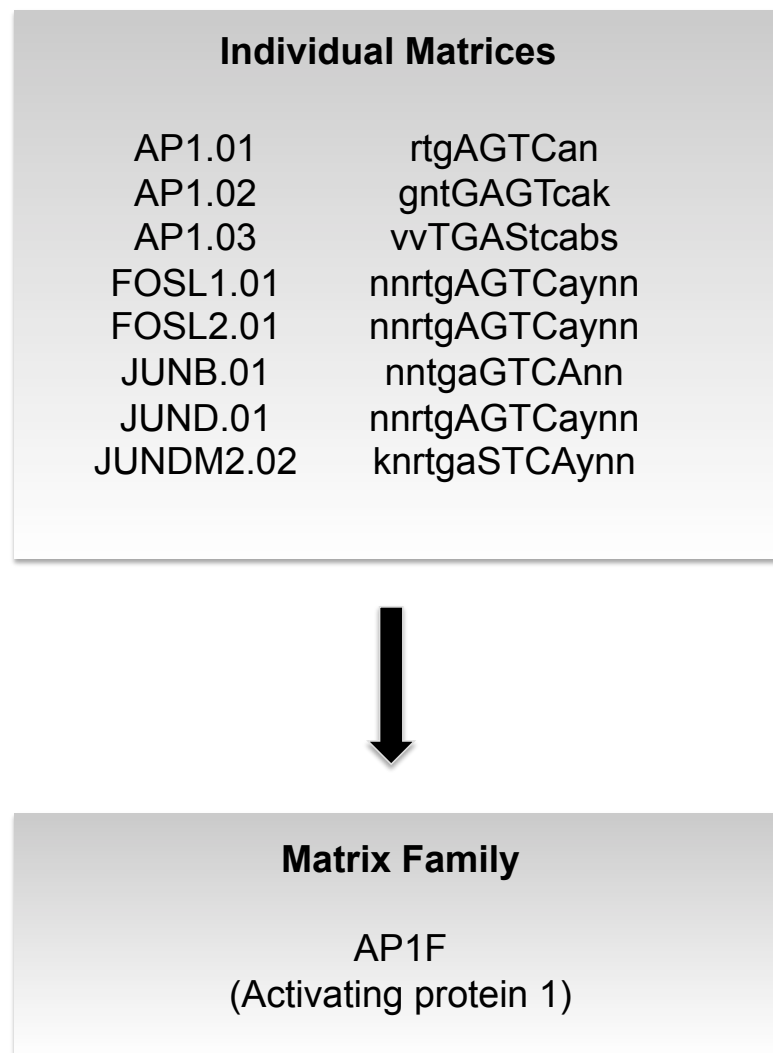
hC		Enhancer name		mC	
hC		hC and mC		mC	
AP-1-TBXAS1	IE1.1	A-MuLV_US1	IL-6-undefi	AP-1-PAI-2	Isl-1_site
AP-2_CS5	IL-9-CRE	a2(l)coll_U	IRF-2_RS	AP-1-thromb	IUF1-CT-II
AP-2_CS6	JCV_repeate	Ad-conserve	LBP-1_CS	AP-2_beta_C	LFA1-albumi
AP-4/E1247-	LF-A1_CS	AP-1_CS1	LBP-1_RS	AP1-TRE-4/T	NBRE
AP1-SV40.2	LVc_RS	AP-1_CS2	LyF/Ikaros	CArG_box_CS	NF-E2-conse
ARIIc	LVc-Mo-MuLV	AP-1_CS3	NF-kB_CS	CARG/SRE_si	NF-kB-IL-9
bHLH_CS	N-Oct-3-CS	AP-1_CS4	NF-Y-MHCII	CD14-TLS	NFI-NFI
c-Myb_CS1	PCI-uAP2	AP-1-IL-3	NF-Y*-tk	complement_	NGFI-B-CYP2
c-Myb_CS4	polymerase-	AP-3_RS	NFI_CS2	CT-I	OMP-undefin
CAC-bp-beta	prolactin-D	AP1-TRE0/C	NFkB_CS1	CTF/NF-1a	Sp1-erk1_(1
CF1_CS	pyT-Ag-CS	CAAT_site(1	NFkB_CS2	CTF/NF-1b	TCF-1_CS
CRE_(6)	SOX_CS	CACCC_site	NFkB_CS4	CTF/NF-I-t	TCF1-GATA-1
E2_US	SP-C	CACCC-box	NFkB_CS5	Erg-2-CS	TFIID-EIIa
E4BP4-conse	T-Ag-SV40.3	CBFaA-glob	NFkB-b2m	GR-intron-s	TRE.1
EBP1_RS	Thy-1-undef	CBP-MSV	NFkB-HCMV-	hFAST-1-ARE	TSE/TRE/DR4
EKLF-gamma-	TTF-TPO-2	CBP-tk	OSP1_CS	HNF-3_site	
gamma-globi	v-MCS	CCAAT_box	PEA1_CS		
HCMV-18_bp_	v-Myb-cycli	CCAAT-bf-h	PEA1_RS		
IE1-V	Z_box_(Zta)	CRFEII-lat	PEA3-uPA		
		EBP1_RS	PPAR-consen		
		enhancer_co	PTF1-consen		
		forkhead-fa	SV40.11		
		FOX_family_	SV40.13		
		gamma-IRE_C	SV40.16		
		GAS/SIE/APR	SV40.6		
		GMCSF_CS	T-Ag-SV40.4		
		HIV-kB	TCF/LEF_CS		
		HNF-6-CS2	TEF1-GTI*		
		hsp70.1	TGGCA-		
		hsp70.3	TonE_CS		
		hsp70.6	TonEBP-CS		
		IE1.2	UPE		
		Ig-undefine	WAP_US5		

	hC	hC and mC	mC
Total	38	66	31
% of total	28 %	49 %	23 %

The names of the different transcription factor binding sites, as per the MacVector software, present in the hC and/or the mC enhancers, are listed alphabetically. The number of total binding sites per category is listed and the contribution of each category to the overall number of different TFBSs has been calculated as a percentage of total TFBSs.

### 5.2.2. Use of the Genomatix software suite

As an alternative to the MacVector software for identifying putative TFBSs, the Genomatix software suite was investigated. Online programs within the Genomatix software suite were used, utilising the EIDorado database (v12-2013). These programs were MatInspector, Common TFs, ModellInspector, and Overrepresented TFBSs. Genomatix software utilises core similarity in order to identify putative TFBSs, whereby only matches to the core region (the four consecutive highest conserved positions) are considered. The output from the Genomatix programs is based on the concept of matrices. Figure 5.2 is a diagrammatic representation of the matrix concept. As an example, the MatInspector and ModellInspector programs are able to show results as individual matrices, representing the different transcription factors with highly similar binding sites, which are all known to bind to the TFBS identified, but cannot be distinguished computationally (Cartharius et al., 2005). The individual matrices are then compiled into a matrix family, which leads to a significantly condensed output (Cartharius et al., 2005).



**Figure 5.2. A schematic of the matrix concept used in the Genomatix software suite.**

The names of individual matrices are listed with their sequences. Nucleotides in capital letters denote the core sequence used by MatInspector. The data generated by MatInspector is based on similarity to the core sequence. In this example, the individual matrices in the AP-1 family are listed. The individual matrices can be compiled into a matrix family, which condenses the data generated by the Genomatix Software Suite.

### 5.2.2.1. MatInspector to identify putative binding sites

The MatInspector program was used to identify individual TFBSs and families of similar TFBSs. After selection of the desired enhancer or promoter sequence, the vertebrate and general core promoter elements were chosen as the matrix libraries to be queried (Matrix Library v9.1). Additionally, the user can manually cross-check the EIDorado database to confirm that the resultant matrix families and individual families are present in the mouse genome. The parameters that can be displayed are either individual matrices or matrix families and searches of the enhancer sequences were carried out using both options.

The number of individual matrices and matrix families are listed in Table 5.3. Of the different families, 41 (57 %) are present in both the mC and hC enhancers; 16 (22 %) are present only in mC while 15 (21 %) are present only in hC. The names of the different families as listed by the MatInspector program, and the enhancer in which they can be found, are listed in Table 5.4. Although the family concept reduces redundant matches, it can also remove important details of the individual binding sites. For example, the Activating Protein 1 family contains a number of family members (as per Figure 5.2): one is present in hC only (i.e. JUNDM2.02), three are present in mC only (AP1.01, AP1.02 and AP1.03) and two are present in both hC and mC (FOSL1.01 and JUND.01). Therefore, for further discrimination between the two enhancers, the individual matrix names have been listed in Table 5.5. There are a total of 179 individual matrices represented across the hC and mC enhancers; 47 (26 %) of the total are shared by hC and mC, 74 (41 %) are present only in hC while 58 (33 %) are present in mC only.

Upon mapping of the TFBSs to their location in the hC enhancer, five TFBSs were identified (Table 5.5 highlighted in the yellow boxes in the hC column) which are completely absent in the mC enhancer and present only once in the entirety of the hC enhancer (Figure 5.3). These five binding sites are absent in the hCd4 enhancer. The TF proteins that bind to

these sites are all expressed in the mouse lung (Table 5.6). Furthermore, the TFs are all involved in both negative and positive transcriptional regulation (Table 5.6).

Conversely, 13 TFBSs in the mC enhancer are present only in the 5' 126 bp of the mC enhancer (as in mCd1 and mCd2), and completely absent in the hC enhancer (Table 5.5, highlighted in the green boxes the mC column) (Figure 5.4). These 13 binding sites are involved in both positive and negative transcriptional regulation (Table 5.7). Of these sites and their related TF proteins, only six of the TFs are expressed in the mouse lung (Table 5.7).

**Table 5.3. Number of potential transcription factor binding sites identified by MatInspector.**

Enhancer name	Enhancer length (bp)	Individual Matrices		Matrix Families	
		Total	Different matrices	Total	Different families
hC	302	185	121	111	56
hCd1	233	129	96	75	49
hCd2	163	79	72	49	38
hCd3	93	35	32	21	18
hCd4	23	0	0	0	0
mC	423	332	105	175	56
mCd1	354	292	103	154	55
mCd2	297	231	97	120	51
mCd3	214	168	83	89	48
mCd4	144	104	61	56	36
mCd5	74	54	46	33	27

The different enhancer sequences were entered into the MatInspector program and the results were manually tallied for the total number of individual matrices and the different matrices present. Similarly, MatInspector yielded names for the matrix families, the results of which were manually tallied to identify the number of total families and the number of different families. Core similarity at 0.75 (default) was chosen and Genomatix-optimised thresholds for matrix similarity were used. See Figures 4.4 and 4.5 for plasmid diagrams.

**Table 5.4. Output names of potential families of transcription factor binding sites identified by MatInspector.**

Enhancer name			
hC	hC and mC		mC
ABDB	AP1F	NBRE	CIZF
CDXF	AP1R	NF1F	ETSF
E4FF	BRN5	NFAT	FAST
FXRE	BRNF	NFKB	FKHD
HAND	CAAT	NKXH	GLIF
LEFF	CART	NOLF	HOXH
MYBL	CLOX	NR2F	IKRS
PARF	CREB	OCT1	OAZF
SAL2	CTCF	PBXC	PAX5
SATB	DMRT	PERO	PDX1
SIXF	EREF	PLAG	SP1F
SNAP	GCMF	PRDF	THAP
STAT	GRAF	RBPF	YBXF
STEM	HBOX	RORA	YY1F
XBBF	HEAT	RXRF	ZF02
	HNF6	SORY	ZF10
	HOMF	SRFF	
	HOXC	TALE	
	HOXF	VTBP	
	IRXF	YY1F	
	LHXF		

	hC	hC and mC	mC
Total	15	41	16
% of all matrix families	21 %	57 %	22 %

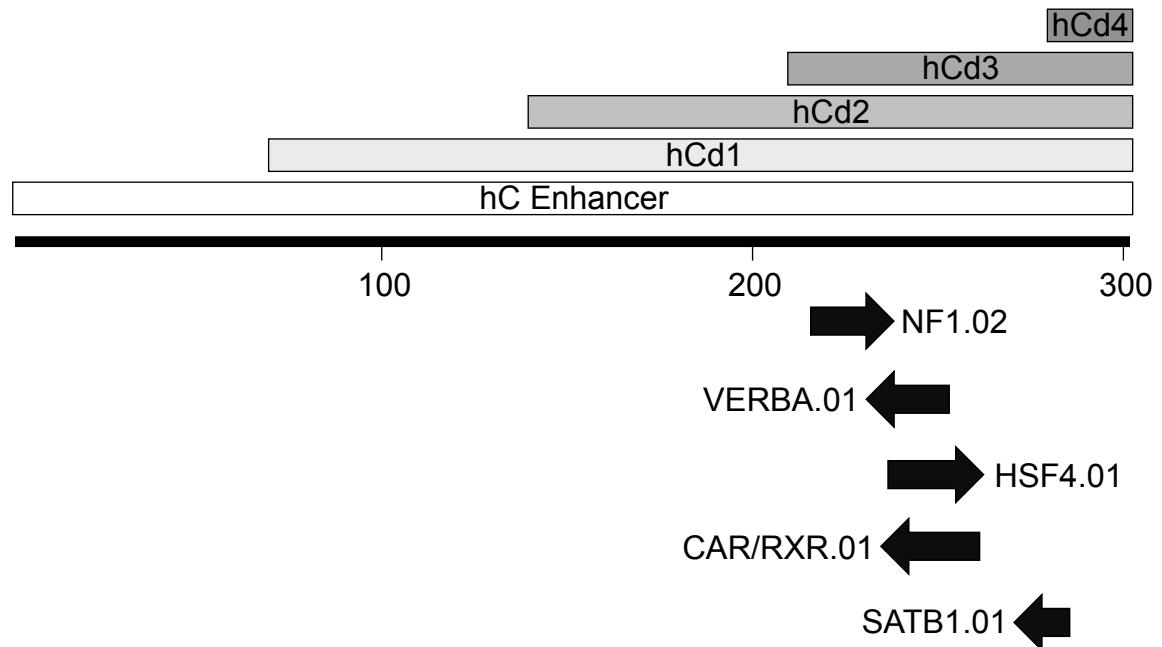
The names of the different matrix families, as per the MatInspector program, present in the hC and/or the mC enhancers, are listed alphabetically. The number of total matrix families per category is listed. In addition, the contribution of each category (hC only, mC only, or present in both hC and mC) to the overall number of different matrix families has been calculated as a percentage of total families.

**Table 5.5. Output names of potential transcription factor binding sites identified by MatInspector.**

hC		hC and mC	mC	
ARP1.01	NF1.04	ARE.02	AP1.01	MEIS1B.01
ATF.02	NKX25.01	ARE.03	AP1.02	<b>NANOG.01</b>
ATF1.01	NKX26.01	ATATA.01	AP1.03	NFE2.01
BRN3.01	OCT.01	BACH1.01	BACH2.01	NFE2.02
BRN3.03	OCT1.01	BARX2.01	<b>BRN4.01</b>	NFY.03
BRN5.01	OCT1.03	BRN2.01	BRN5.03	<b>NKX31.02</b>
BRN5.04	OCT1.04	CAAT.01	CDP.02	NMP4.01
<b>CAR/RXR.01</b>	OCT1.06	CREL.01	<b>COUP.02</b>	<b>OCT1.05</b>
CDPCR3HD.01	OCT2.01	CTCF.05	<b>DMRT7.01</b>	PAX5.04
CDX2.03	OCT3 4.01	ERR.01	ELK1.01	PBX1/MEIS1.01
COUP.01	OSNT.01	ESRRA.01	EMX2.01	PDX1.01
CREB.03	PBX/HOXA9.01	ESRRA.03	ERG.02	PPARG.03
CREB1.02	PBX1/MEIS1.02	ESRRB.01	EVX1.01	PRDM14.01
CREB2.01	POU2F3.01	FOSL1.01	FAST1.01	<b>REV-ERBA.02</b>
CTCF.03	PPAR/RXR.01	GCM1.03	FAST1.02	REV-ERBA.03
CTCF.04	PSE.02	GRE.02	FAST1.03	<b>ROAZ.01</b>
DBP.01	RAR/RXR.01	GRE.03	<b>FOXP2.01</b>	<b>RORA.01</b>
DMRT2.01	RBPJK.01	GSH1.01	<b>FREAC4.01</b>	RORA1.01
DMRT3.01	REX1.01	HBP1.02	GLIS2.01	SOX3.01
DMRT4.01	RFX1.01	HIVEP1.01	HLXB9.01	TAXCREB.02
E4F.01	SALL2.01	HNF6.01	HNF4.02	TCF11MAFG.01
ESRRA.02	<b>SATB1.01</b>	HOX1-3.01	HSF1.03	THAP1.01
FXRE.01	SIX1.01	IRX6.01	IK1.01	TIEG.01
HMGA.01	SIX6.01	ISL1.01	IK3.01	TR2.01
HMX3.01	SOX9.01	JUND.01	IRX2.01	VDR/RXR.06
HOX/PBX.01	STAT.01	MEIS1.03	IRX3.01	<b>XFD3.01</b>
HOXB3.01	TCF7L2.01	NBRE.01	IRX4.01	<b>YB1.01</b>
HOXB9.01	TEF.01	NF1.01	IRX5.01	YY2.02
HOXD8.01	TR2/TR4.01	NF1.03	JUNDM2.02	ZBP89.01
<b>HSF1.04</b>	TTF1.02	NFAT5.02		
JUNDM2.01	VDR/RXR.03	NFKAPPAB.01		
LEF1.03	VDR/RXR.05	NFKAPPAB.02		
LEF1.04	<b>VERBA.01</b>	NFKAPPAB65.01		
LXRE.02	VMYB.02	NFKAPPAB65.02		
MESP1/2.01	VMYB.04	NFY.01		
MYBL1.02	YY1.01	NFY.02		
<b>NF1.02</b>	ZBTB3.01	NFY.04		
		NFY.05		
		OLF1.01		
		PBX1/MEIS1.03		
		PHOX2.01		
		PLAG1.02		
		PRDM1.02		
		RAR/RXR.03		
		RBPJK.02		
		SRF.01		
		TGIF.01		

	hC	hC and mC	mC
Total	74	47	58
% total individual matrices	41 %	26 %	33 %

The names of the different individual matrices, as per the MatInspector program, present in the hC and/or the mC enhancers, are listed alphabetically. The number of total individual matrices per category is listed. In addition, the contribution of each category (hC only, mC only, or present in both hC and mC) to the overall number of individual matrices has been calculated as a percentage of total matrices. The sites in the hC column (highlighted in yellow) are uniquely present in the 3' 93 bp of the hC enhancer (as per hCd3) (Figure 5.3). The sites in the mC column (highlighted in green) are uniquely present in the 5' 126 bp of the mC enhancer (as per mCd1 and mCd2) (Figure 5.4). See Figures 4.4 and 4.5 for plasmid diagrams.



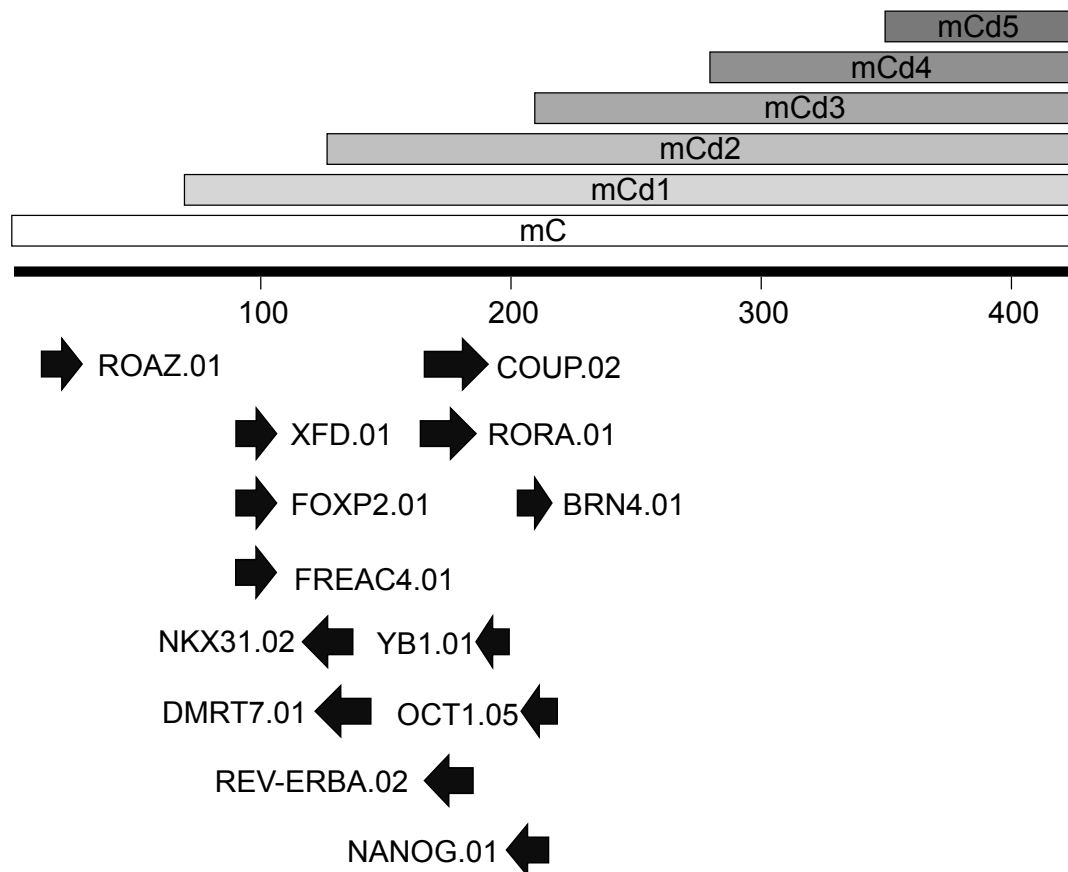
**Figure 5.3. Predicted transcription factor binding sites which occur once in the hC enhancer and are absent in the mC enhancer.**

Each hC enhancer variant is represented by the white and grey boxes, with a base pair ruler to indicate the length of each enhancer. The arrows represent the location of the five TFBSs (identified by MatInspector; Section 5.2.2.1) that are present once in the entirety of the hC enhancer which are also completely absent in the mC enhancer. These sites are: NF1.02 (nuclear factor 1), VERBA.01 (viral homolog of thyroid hormone receptor  $\alpha$  1), HSF1.04 (heat shock factor 1), CAR/RXR.01 (constitutive androstane receptor/retinoid X receptor heterodimer), and SATB1.01 (special AT-rich sequence-binding protein 1). See Table 5.6 for further information.

**Table 5.6. Detail of TFBSs uniquely present in the 3' 93 bp of the hC enhancer.**

TFBS name	Matrix information	Evidence of TF mRNA or protein in mouse adult lung	Positive and negative regulator of transcription (as per UniProt)?
NF1.02	nuclear factor 1	Yes (Chaudhry et al., 1997)	Yes
VERBA.01	viral homolog of thyroid hormone receptor $\alpha$ 1	Yes (Flamant and Samarut, 2003)	Yes
HSF1.04	heat shock factor 1	Yes (Gally et al., 2011)	Yes
CAR/RXR.01	constitutive androstane receptor/retinoid X receptor (heterodimer)	CAR – Yes (Petrick and Klaassen, 2007) RXR – Yes (Naltner et al., 2000)	Yes
SATB1.01	special AT-rich sequence-binding protein 1	Yes (Baguma-Nibasheka et al., 2007)	Yes

The names of the TFBS detailed in Table 5.5 (yellow) and Figure 5.3 are further explained. The TFBS name and matrix information (as per MatInspector), and published evidence of the relevant TF mRNA or protein in adult mouse lung, are listed. Evidence for positive or negative regulator of transcription was based on GeneOntology (GO) biological process annotations compiled in the Universal Protein Resource (UniProt) based on mouse entries for the relevant proteins (Huntley et al., 2015).



**Figure 5.4. Predicted transcription factor binding sites which occur once in the mC enhancer and are absent in the hC enhancer.**

Each mC enhancer variant is represented by the white and grey boxes, with a base pair ruler to indicate the length of each enhancer. The arrows represent the location of the 13 TFBSs (identified by MatInspector; Section 5.2.2.1) that are present once in the entirety of the mC enhancer which are also completely absent in the hC enhancer. These sites are: ROAZ.01 (rat zinc finger protein), COUP.02 (chicken ovalbumin upstream promoter; NR2F2), XFD3.01 (Xenopus forkhead domain factor-3; FoxA2a), RORA.01 (RAR-related orphan receptor alpha), FOXP2.01 (forkhead box protein P2), BRN4.01 (POU domain transcription factor brain 4), FREAC4.01 (forkhead related activator-4; FOXD1), NKX31.02 (NK3 homeobox 1), YB1.01 (Y box binding protein 1), DMRT7.01 (doublesex and mab-3 related transcription factor 7), OCT1.05 (octamer-binding factor 1), REV-ERBA.02 (orphan nuclear receptor rev-erb alpha), NANOG.01 (homeobox transcription factor Nanog). See Table 5.7 for further information.

**Table 5.7. Detail of TFBSs uniquely present in the 5' 126 bp of the mC enhancer.**

TFBS name	Matrix information	Evidence of TF mRNA or protein in mouse adult lung	Positive and negative regulator of transcription (as per UniProt)?
ROAZ.01	rat zinc finger protein; EBFAZ	Yes (Warming et al., 2004)	Yes
COUP.02	chicken ovalbumin upstream promoter; NR2F2	No; not present in differentiated epithelial cells (Bao et al., 2014)	Yes
XFD3.01	Xenopus for head domain factor 3; FOXA2	No; involved in embryonic lung development (Wan et al., 2005)	Yes
RORA.01	RAR-related orphan receptor alpha	Yes (Stapleton et al., 2005)	Yes
FOXP2.01	forkhead box protein P2	Yes (Shu et al., 2001)	Yes
BRN4.01	POU domain transcription factor brain 4; POU3F4	No; limited to the brain (Hara et al., 1992)	Positive only
FREAC4.01	forkhead related activator 4; FOXD1	No; involved in embryonic lung development (Hung et al., 2013)	Yes
NKX31.02	NK3 homeobox1	No; specific to urogenital system (Sciavolino et al., 1997)	Yes
YB1.01	Y box binding protein 1	Yes (Miwa et al., 2006)	Yes
DMRT7.01	doublesex and mab-3 related transcription factor 7	Not present (Hou et al., 2010)	Positive only
OCT1.05	octamer binding factor 1; POU2F1	Yes (Zhao et al., 2004a)	Yes
REV-ERBA.02	orphan nuclear receptor rev-erb alpha; NR1D1	Yes (Vasu et al., 2009)	Yes
NANOG.01	homeobox transcription factor Nanog	Not present in lung (Hart et al., 2004)	Yes

The names of the TFBS detailed in Table 5.5 (green) and Figure 5.4 are further explained. The TFBS name and matrix information (as per MatInspector), and published evidence of the relevant TF mRNA or protein in adult mouse lung, are listed. Evidence of positive or negative regulator of transcription was based on GeneOntology (GO) biological process annotations compiled in the Universal Protein Resource (UniProt) based on mouse entries for the relevant proteins (Huntley et al., 2015)

### 5.2.2.2. **ModellInspector to identify joint promoter modules**

Promoter-mediated transcription typically requires the joint activity of many transcription factors whose binding and activation are highly interdependent (Ravasi et al., 2010). Therefore, although the identification of single TFBSs may be useful in some cases, it is unlikely that a solitary binding site is responsible for activation or repression of transcription. To this end, the ModellInspector program queries the input sequence for the presence of so-called promoter modules, which are defined as functional units consisting of two or more TFBSs conserved in order and distance (Schmid et al., 2006), that have been experimentally verified to interact with each other to activate or repress a gene (Cartharius et al., 2005). The predicted promoter modules are based on a library comprising vertebrate-specific sequences, such that the results can be manually checked by the user by querying the EIDorado database to ensure that the promoter modules are represented in the mouse genome. Additionally, the nature of the interaction (i.e. positive effect on transcription) can also be manually verified against the library of MatInspector's promoter modules. Table 5.8 lists the predicted promoter modules in each of the hC and mC enhancer variants.

**Table 5.8. Output names of potential interacting transcription factor binding sites identified by ModellInspector.**

**A**

	Enhancer name				
	hC	hCd1	hCd2	hCd3	hCd4
ETSF_NFKB_02	ETSF_NFKB_02	ETSF_NFKB_02	ETSF_NFKB_02	ETSF_NFKB_02	n/a
NF1F_EBOX_01	NF1F_EBOX_01	NF1F_EBOX_01	NF1F_EBOX_01	NF1F_EBOX_01	
NFKB_CREB_01	NFKB_CREB_01	NFKB_CREB_01	NFKB_CREB_01		
PAX6_CDXF_01	PAX6_CDXF_01	PAX6_CDXF_01			
TALE_HOXC_01	TALE_HOXC_01				
NFAT_CREB_01					
NR2F_EREF_01					
CAAT_AP1F_01					
CREB_NFAT_NFAT_01					
Total	9	5	4	2	0

**B**

	Enhancer name					
	mC	mCd1	mCd2	mCd3	mCd4	mCd5
ETSF_NFKB_02	ETSF_NFKB_02	ETSF_NFKB_02	ETSF_NFKB_02	ETSF_NFKB_02	ETSF_NFKB_02	n/a
NFAT_AP1F_03	NFAT_AP1F_03	NFAT_AP1F_03	NFAT_AP1F_03	NFAT_AP1F_03	NFAT_AP1F_03	
NR2F_EREF_01	NR2F_EREF_01	NR2F_EREF_01	NR2F_EREF_01	NR2F_EREF_01	NR2F_EREF_01	
BRNF_RXRF_01	BRNF_RXRF_01	BRNF_RXRF_01	BRNF_RXRF_01	BRNF_RXRF_01		
KLFS_KLFS_NFKB_01	KLFS_KLFS_NFKB_01	KLFS_KLFS_NFKB_01	KLFS_KLFS_NFKB_01	KLFS_KLFS_NFKB_01		
CAAT_SREB_03	CAAT_SREB_03	CAAT_SREB_03				
OCT1_CAAT_02	OCT1_CAAT_02	OCT1_CAAT_02				
AP1F_NFKB_06	AP1F_NFKB_06					
NFAT_AP1R_02	NFAT_AP1R_02					
NFKB_AP1F_01						
NFKB_CREB_01						
Total	11	9	7	5	3	0

The names of the predicted promoter modules in the enhancers of the (A) hC- and (B) mC-derived plasmids, as identified by the program ModellInspector, are listed alphabetically. The total number of promoter modules present in each enhancer is listed in each column. All promoter modules have been checked to confirm that they have been experimentally verified to be active in mouse (i.e. *Mus musculus*).

### 5.2.2.3. Identification of overrepresented TFBSs

An additional function available within the Genomatix software suite is the analysis of the input enhancer and promoter sequence for the presence of overrepresented binding sites compared with the number of occurrences in the genomic regions or promoters of a specific species. Overrepresented binding sites may constitute a potential source of competition for available transcription factors, possibly leading to weak transcriptional activation. The Overrepresented TFBSs program was used (Matrix Library v9.1) and compared to a *Mus musculus* genomic and promoter background (NCBI build 38). As for MatInspector, both the 'individual matrices' and 'matrix families' options were chosen. The number of overrepresented individual matrices in the enhancer variants, and the EF promoter, are listed in Table 5.9. The top two overrepresented TFBSs for each sequence are listed in Table 5.10.

A further comparison of each enhancer and promoter combination identified overrepresented families of TFBSs that are shared with both the EF promoter and each respective hC- and mC-derived enhancer variant. Table 5.11 and Table 5.12 list the names of the matrix families found in the EF promoter and the hC- or mC-derived enhancers, respectively. The matrix families that are overrepresented are highlighted in bold. The EF promoter and hC enhancer share 18 TFBS families, whereby the EF promoter and mC enhancer share 21 TFBS families. As expected, the number of overrepresented families of TFBSs decreases as the enhancer sequence is reduced. However, an unexpected finding was the high number of overrepresented TFBS families in mC (nine, Table 5.12) when compared with hC (one, Table 5.11). For the purposes of this section, a 'high' number of overrepresented binding sites or families is defined as five or more.

**Table 5.9. Number of overrepresented TFBSs in the EF promoter, and hC- and mC-derived enhancers.**

Sequence name	Genome	Promoters
EF	20	12
hC	42	46
hCd1	27	34
hCd2	19	14
hCd3	24	25
hCd4	0	0
mC	51	56
mCd1	54	59
mCd2	48	50
mCd3	35	36
mCd4	37	33
mCd5	36	38

The different promoter and enhancer sequences were entered into the Overrepresented TFBSs program and run against a *Mus musculus* genomic and promoter background. All overrepresented individual matrices with a Z-score of  $\geq 2$  were manually tallied. A Z-score is attributed to the relative over- or under-representation of the matrix in the sequence relative to the input sequence length. The Z-score is the number of standard deviations that a given data point is from the population mean. A minimum Z-score of 1.96 (approximate  $p$  value = 0.05) was used as the cut-off for inclusion in further analysis. The  $p$  value is an indication of statistical significance, rather than biological significance. See Figures 4.4 and 4.5 for plasmid diagrams.

**Table 5.10. The top two overrepresented binding sites found in the EF promoter and all hC- and mC-derived enhancers.**

Overrepresented transcription factor binding sites										
1						2				
Sequence name	Matrix name	Total matches	Number expected	Overrep.	Z-score	Matrix name	Total matches	Number expected	Overrep.	Z-score
EF	ZKSCAN3.01	2	0.05	37.2	6.24	KKLF.01	3	0.16	18.67	5.84
hC	NFKAPPAB65.02	4	0.06	62.59	13.59	JUNDM2.01	4	0.09	44.06	11.32
hCd1	NFKAPPAB65.02	4	0.05	77.16	15.15	NF1.03	3	0.11	27.1	7.18
hCd2	NF1.02	2	0.03	77.33	9.17	NF1.03	3	0.08	39.32	8.78
hCd3	NF1.02	2	0.01	135.54	12.23	NFKAPPAB65.02	2	0.02	101.62	10.55
hCd4	n/a					n/a				
mC	NFKAPPAB65.02	16	0.09	178.73	51.51	AP1.03	12	0.15	80.41	29.39
mCd1	NFKAPPAB65.02	12	0.07	160.18	41.75	AP1.03	10	0.12	80.07	26.53
mCd2	NFKAPPAB65.02	11	0.06	175.01	41.64	JUNDM2.02	9	0.13	67.15	22.86
mCd3	NFKAPPAB65.02	7	0.05	154.56	30.33	AP1.03	8	0.08	105.96	27.03
mCd4	NFKAPPAB65.02	5	0.03	164.07	25.61	AP1.03	4	0.05	78.74	25.61
mCd5	NFKAPPAB65.02	2	0.02	127.71	11.86	AP1.03	2	0.03	76.61	9.12

The promoter and enhancer sequences were entered into the Overrepresented TFBSs program and run against a *Mus musculus* genomic and promoter background. The top two overrepresented binding sites (indicated as 1 and 2) for the EF promoter and each enhancer variant used in the studies in this chapter are listed, along with the total matches in the sequence, the number of expected matches against a mouse genomic background in a sequence of the same length (rounded to the nearest hundredth), the overrepresentation of that binding site (total matches divided by number expected), and the Z-score, whereby a value  $\geq 1.96$  is equivalent to  $p \leq 0.05$ . See Figures 4.4 and 4.5 for plasmid diagrams.

Table 5.11. Matrix families represented in the EF promoter and hC enhancer variants.

Matrix family	Promoter or enhancer name					
	EF	hC	hCd1	hCd2	hCd3	hCd4
CTCF	<b>2</b>	1	1	1		
EREF	1	2	1	1	1	
GCMF	<b>2</b>	1	1	1		
HAND	1	1	1	1		
HBOX	2	1				
HOMF	1	2	1	1		
NKXH	2	1	1	1		
NOLF	1	1	1	1	1	
NR2F	1	2				
PERO	1	1				
PRDF	1	1	1			
RBPF	1	1	1			
RXRF	1	<b>5</b>	<b>3</b>	<b>3</b>	<b>2</b>	
SAL2	<b>2</b>	1	1			
SORY	1	3	3	1		
STAT	1	2	1			
VTBP	2	1	1			
YY1F	1	2	1	1		
Total families overrepresented	3	1	1	1	1	0

The names of the different matrix families, as per the MatInspector program, present in the EF promoter and the hC enhancer variants are listed alphabetically. The number of times the family appears in each sequence is listed. The numbers shown in bold signify that the family is overrepresented in the sequence when compared against a *Mus musculus* genomic or promoter background. The total number of overrepresented families is shown at the bottom of the table. See Figures 4.4 and 4.5 for plasmid diagrams.

Table 5.12. Matrix families represented in the EF promoter and mC enhancer variants.

Matrix family	Promoter or enhancer name						
	EF	mC	mCd1	mCd2	mCd3	mCd4	mCd5
CIZF	1	<b>3</b>	<b>2</b>	<b>2</b>	1	<b>1</b>	<b>1</b>
CTCF	<b>2</b>	<b>2</b>	<b>2</b>	1	1	1	1
EREF	1	<b>4</b>	<b>4</b>	<b>3</b>	<b>2</b>	1	1
FKHD	1	1	1				
GCMF	<b>2</b>	<b>3</b>	<b>3</b>	<b>2</b>	<b>2</b>	1	1
GLIF	1	2	2	1	1	1	1
HBOX	2	3	3	3	3	1	1
HOMF	1	1	1	1	1	1	1
NKXH	2	1	1				
NOLF	1	<b>4</b>	<b>3</b>	<b>2</b>	<b>2</b>	1	
NR2F	1	<b>8</b>	<b>7</b>	<b>6</b>	<b>3</b>	2	1
PAX5	<b>2</b>	<b>3</b>	2	1	1	1	
PERO	1	1	1	1	1	1	
PRDF	1	2	1	1	1	1	
RBPF	1	<b>3</b>	<b>2</b>	<b>2</b>	1	1	1
RXRF	1	3	3	2	1	1	1
SORY	1	2	2	1	1	1	
SP1F	1	2	2	1	1	1	1
VTBP	2	4	4	3	2	1	1
YY1F	1	<b>4</b>	<b>4</b>	<b>4</b>	<b>2</b>	<b>2</b>	<b>2</b>
ZF02	<b>7</b>	2	2	1	1	1	1
Total families overrepresented	4	9	8	7	5	2	2

The names of the different matrix families, as per the MatInspector program, present in the EF promoter and the mC enhancer variants are listed alphabetically. The number of times the family appears in each sequence is listed. The numbers shown in bold signify that the family is overrepresented in the sequence when compared against a *Mus musculus* genomic or promoter background. The total number of overrepresented families is shown at the bottom of the table. See Figures 4.4 and 4.5 for plasmid diagrams.

### **5.3. Discussion**

The identification and comparison of TFBSs in enhancer sequences can give a better understanding of how enhancer/promoter sequences mediate high or low levels of transgene transcription in the lung, information that might contribute to the design of new gene therapy vectors. The studies in Chapter 4 focused on a functional approach, whereby enhancer variants were constructed, administered to mice, and subsequent luciferase activity was measured. By combining the findings of Chapter 4 with the computer-based TFBS analysis in this chapter, a number of potentially important TFBSs have been identified. In addition to understanding the differences in the number and type of TFBSs in the hC and mC enhancers, the use of the Genomatix software has allowed for in-depth analysis of these enhancer sequences.

#### **5.3.1. Transcription factor binding site analysis**

The sequence similarity between the hC and mC enhancers is approximately 39 % (Figure 5.1). However, the similarity of TFBS between the two enhancers is approximately 49 % although this value depends on the program used. Irrespective of the software used, the mC enhancer contains more overall TFBSs than hC (332 versus 185 as identified by MatInspector; Table 5.3). However, the hC enhancer encompasses a higher number of different binding sites (121 different matrices compared to 105 in the mC enhancer). This analysis indicates that the sheer number of TFBSs is not sufficient to mediate high-level or persistent transgene activity but that perhaps the higher number of different binding sites may play a role.

#### **5.3.2. Identification of sites in the hC enhancer which may mediate high levels of transgene activity**

An advantage of using the Genomatix software suite over the transcription factor database utilised by MacVector is that Genomatix compiles individual binding sites into matrix families to reduce redundancy in the output of the binding site names. When evaluating the differences in matrix families between the hC and mC enhancers (Table 5.4), and where the

matrix families lie in relation to the enhancer variants, it transpires that the SATB family is absent in the mC enhancer. Interestingly, the SATB site is present only once in the hC enhancer and it is located within the final 93 bp of the hC enhancer (as in hCd3) - the smallest hC enhancer variant to mediate high-level persistent transgene activity similar to the full-length hC enhancer (Figure 4.8). Furthermore, this binding site is absent in hCd4, which yields reporter gene levels similar to the negative control plasmid, pG4-EFI-soLux (Figure 4.8). The SATB family is described by Genomatix as a family of special AT-rich sequence binding proteins able to bind to scaffold/matrix attachment regions (S/MAR); further analysis indicates that the specific TFBS associated present in the hC enhancer is SATB1.01, which binds the SATB1 protein. The SATB1 protein has been identified as a component of the nuclear matrix (Gotzmann et al., 2000), and is expressed in the mouse lung (Baguma-Nibasheka et al., 2007). Non-viral vectors have previously been designed to incorporate S/MAR sequences in order to prolong transgene activity *in vivo* (Argyros et al., 2008; Bode et al., 2000; Piechaczek et al., 1999) (Section 1.3.1). Therefore, the identification of this SATB binding site in the hC enhancer may represent a sequence that could play a role in persistent transgene activity. Although further studies would be required to more clearly understand the role of this site in the hC enhancer, this *in silico* analysis has identified a putatively important TFBS that can form the basis of further investigations.

A disadvantage of using the matrix family approach is that detail can be lost in attempting to identify individual TFBSs that may play a role in mediating transgene activity. Therefore, individual matrices (Table 5.5) were assessed for their location in the hC enhancer. Five TFBSs were identified (see Section 5.2.2.1) that are uniquely present in the 3' 93 bp of hC (as per hCd3) (Figure 5.3) and given their location of these sites within the hC enhancer, it is plausible that the combination of these binding sites mediates the high-level activity observed with the hC and hCd3 enhancers. The combination of these sites as interacting units was not identified by ModellInspector (Table 5.8). However, this does not preclude the fact that a combinatorial effect may play a role in the activation of high levels of transcription observed

in the plasmid containing the hCd3 enhancer. Further investigations into the potential roles of these binding sites are explored in Chapter 6.

### 5.3.3. Overrepresented binding sites as a source of competition

The results in Section 4.2.6 indicate that persistence of transgene activity was possible with deletions of the mC enhancer. For example, after a deletion of 209 bp (as in mCd3), stable luciferase activity was observed between day 1 and day 2 (Figure 4.9E); this is in contrast to the initial decrease in transgene activity observed with the larger mC enhancer variants (as in mCd1 and mCd2, Figures 4.9C and D). Additionally, the smallest mC enhancer variant, encompassing the 3' 74 bp of mC (as in mCd5), was able to mediate similar day 14 values when compared to mC (Figure 4.9G). Together, these data suggest that as the mC variants become smaller a hitherto unidentified element is being deleted, thereby abolishing the initial decrease in transgene activity.

Another possible aspect is the presence of overrepresented TFBSs. As detailed in Table 5.3, the mC enhancer contains a larger number of TFBSs per 100 bp than hC while the hC enhancer contains a greater diversity of binding sites. This is further reflected in the number of overrepresented TFBSs present in mC versus hC. As shown in Table 5.9, mC has 51 overrepresented TFBSs when compared with a mouse genomic background while hC contains 42 overrepresented sites. The relatively large number of overrepresented binding sites in the mC enhancer compared with the hC enhancer may indicate a source of inter-enhancer competition for available transcription factors.

An additional parameter that can be evaluated along this theme is the number of TFBS families the enhancer and promoter have in common, and assessing how many of these shared families are overrepresented. While hC and EF have 18 TFBS families in common, only one of these binding sites (RXRF) is overrepresented in hC, while three are overrepresented in the EF promoter (CTCF, GCMF and SAL2) (Table 5.11). The mC

enhancer and EF promoter have 21 TFBS families in common; four of these families are overrepresented in the EF promoter (CTCF, GCMF, PAX5 and ZF02) but nine are overrepresented in the mC enhancer (CIZF, CTCF, EREF, GCMF, NOLF, NR2F, PAX5, RBPF and YY1F) (Table 5.12). Additionally, three of the families are overrepresented in both EF and mC (CTCF, GCMF and PAX5). When portions of the mC enhancer are deleted, the decrease in transgene activity observed between day 1 and day 14 is reduced, and the initial decrease in activity observed at early time-points is also curtailed (Figure 4.9). These relative improvements in transgene activity coincide with a decrease in the number of shared overrepresented binding sites in the mC enhancer variants and EF (Table 5.12). The full-length mC enhancer has nine overrepresented TFBS families, mCd3 only has five overrepresented families and mCd5 only has two overrepresented TFBS families. Therefore, in addition to inter-enhancer competition for available transcription factors described above, these results may indicate that enhancer-promoter competition for TF proteins may lead to lack of persistence observed with the full-length mC enhancer. Further evidence for this possibility is suggested by the low number of overrepresented TFBS families in hC (Table 5.11) and its ability to mediate high-level activity (Figure 4.8). The high number of shared overrepresented binding sites in the mC enhancer may act as a sink, whereby TF proteins are pulled towards the mC enhancer and away from the EF promoter, which could result in weak transcriptional activity at later time-points (i.e. day 14).

The perturbation of one transcriptional unit by another linked *in cis* is generally referred to as transcriptional interference (TI) (Eszterhas et al., 2002), although TI can cover a variety of related terms such as: promoter interference, promoter occlusion, promoter competition, sequestering or squelching. In this discussion, TI refers to a negative effect on transcriptional activity. One of the mechanisms proposed for TI includes competition for *cis*- or *trans*-acting factors including enhancers or their binding proteins, co-activators or general transcription factors (Eszterhas et al., 2002). Also, published evidence indicates that TF binding to repetitive DNA can act as a direct negative regulator of transcriptional activity (Liu et al.,

2007). To date, the direct effect of overrepresented binding sites in enhancers and promoters on the decrease of transcriptional activity from plasmid vectors has not been addressed in the literature. While a clear mechanism to explain TI has not been identified, perturbations of transcription cannot always be anticipated in advance, and should be assessed empirically when gene transfer vectors contain more than one promoter or regulatory element (Curtin et al., 2008). Indeed, the functional studies presented in Chapter 4 have been useful in providing a framework from which to further explore the potential sources of TF competition described in this chapter.

#### **5.3.4. Conclusions**

*In silico* analysis has identified variations in TFBSs in the mC and hC enhancer variants constructed in Chapter 4. A model was developed in which transcriptional interference between the variant mC enhancers and the EF promoter may explain differing transcriptional activity. Interestingly, a number of TFBSs were identified in the hCd3 enhancer that may be crucial for mediating sustained, high-level transgene activity in the mouse lung. This finding forms the basis of further investigations.

## Chapter 6: Effect of deletions of the minimal hC enhancer on luciferase activity in the mouse lung

### 6.1. Introduction

The results from Chapter 4 showed that the 93 bp minimal hC enhancer (hCd3) is sufficient to mediate high-level and persistent transgene activity after aerosol delivery to the mouse lung (Section 4.2.5; Figure 4.8). The studies in this chapter assess a number of additional enhancer variants. These were generated in response to the observation that while the 93 bp minimal hC enhancer (hCd3) mediates high-level luciferase activity, the deletion of an additional 70 bp (as in hCd4) drastically reduces activity to levels similar to the negative control plasmid (pG4-EFI-soLux) containing no enhancer (Section 4.2.5; Figure 4.8). Therefore, it was hypothesised that this 70 bp region in the hCd3 enhancer contains regulatory sequences that are necessary and sufficient to mediate high, persistent levels of reporter gene activity in the mouse lung after aerosol dosing. A new enhancer variant was generated to encompass this 70 bp region, which formed the basis of further studies in this chapter.

A previous study (Pringle et al., 2012b) investigated the possibility of exploiting the persistence of transgene activity observed from hCEFI-containing plasmids in combination with the high levels mediated at early time-points with the mCEFI enhancer/promoter sequence. To this end, dual enhancers were constructed by using the hC and mC sequences (either as hCmC or mChC) (Pringle et al., 2012b). Using the mouse aerosol model, it was found that the enhancer immediately upstream of the EF promoter exerted the dominant effect on transgene activity (Pringle et al., 2012b). That is, in the case of using the hCmC enhancer combination, the transcriptional activity was essentially identical to that of a construct containing only mC. Conversely, the use of the mChC dual enhancer was indistinguishable from using the hC enhancer alone. The reasons for these results are unknown, but the authors suggested that the direct interaction of the proximal enhancer with the promoter may be critical for the dominant effect observed (Pringle et al., 2012b).

The aims of the work in this chapter were to: (i) evaluate whether a dual enhancer containing mC and the minimal 93 bp hC enhancer (mChCd3, Figure 6.1B) would yield similar transgene activity to the mChC enhancer (Pringle et al., 2012b); and (ii) determine transgene activity from new enhancer variants derived from the 93 bp minimal hC enhancer (hCd3).

## **6.2. Results**

### **6.2.1. Plasmid vectors used in Chapter 6 studies**

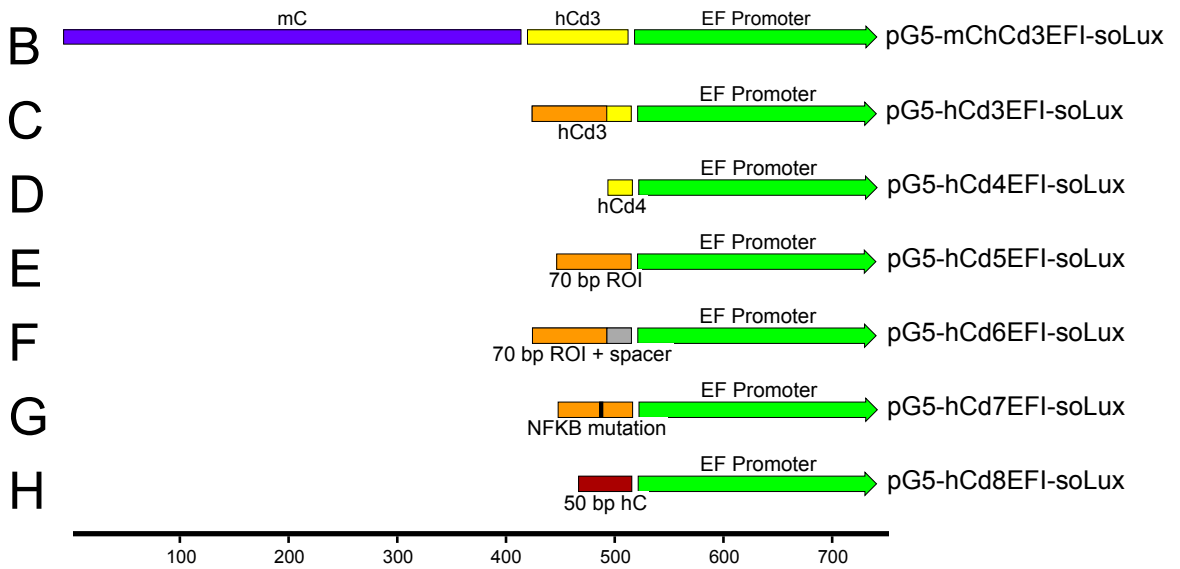
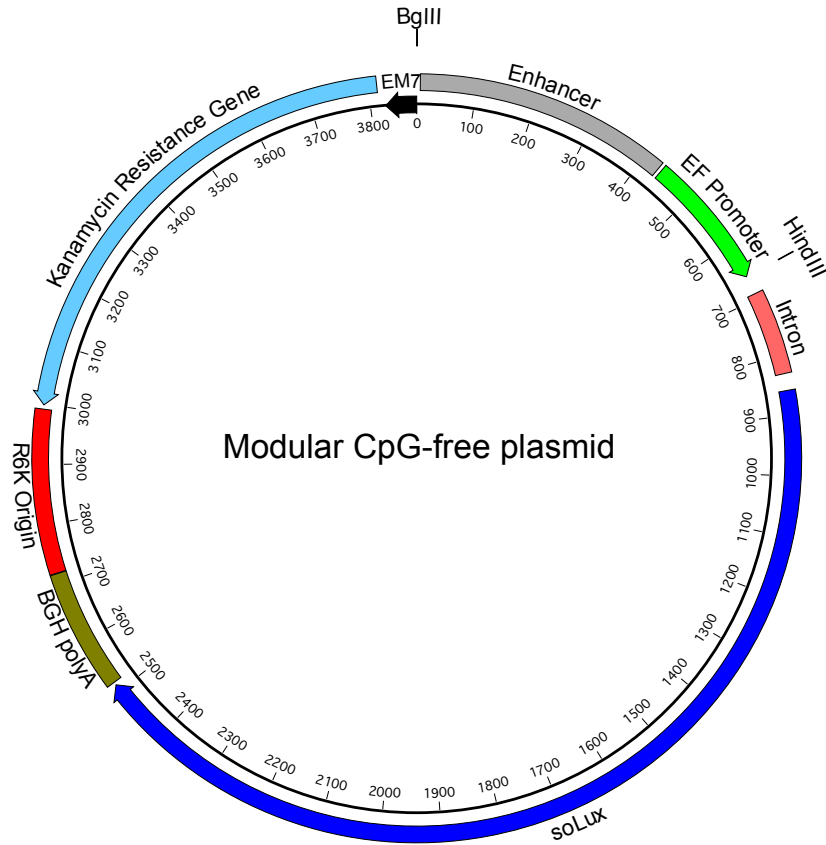
All constructs utilised in these studies in this chapter contained the CpG-free versions of the human elongation factor 1 $\alpha$  (EF) promoter and firefly luciferase (soLux) transgene (Hyde et al., 2008). The new enhancer variants (Table 6.1 and Figure 6.1) were constructed for these studies by PCR amplification of the desired fragments (Section 2.3.1) and subsequently ligated into an appropriate plasmid fragment as described in Section 4.2.3.

**Table 6.1. Overview of plasmid name, original source and enhancer length.**

Plasmid name	Enhancer length	Description	Figure 6.1 reference
pG5-mChCd3EFI-soLux	423 mC + 93 bp hC	Full-length mC enhancer placed before the hCd3 enhancer	B
pG5-hCd3EFI-soLux	93 bp	Minimal enhancer previously identified which mediates transgene activity similar to full-length hC (pG4-hCEFI-soLux)	C
pG5-hCd4EFI-soLux	23 bp	Enhancer previously identified which mediates transgene activity similar to plasmid without an enhancer (pG4-EFI-soLux)	D
pG5-hCd5EFI-soLux	70 bp	Contains the 5' 70 bp of the hCd3 enhancer; denoted as the 'region of interest' (ROI) in these studies	E
pG5-hCd6EFI-soLux	93 bp	Contains the 5' 70 bp ROI with an additional 23 bp 'spacer' region at the 3' end	F
pG5-hCd7EFI-soLux	70 bp	Contains the 5' 70 bp ROI, but the NF $\kappa$ B site has been mutated	G
pG5-hCd8EFI-soLux	50 bp	Encompasses a 5' 20 bp deletion of the 70 bp ROI; that is, contains the 3' 50 bp of the ROI	H

The table lists the pDNA used in the enhancer deletion studies presented in this chapter. The plasmid name, with the enhancer/promoter combination is detailed, along with the enhancer length. See Figure 6.1 for diagrammatical representations of the enhancer/promoter regions.

A

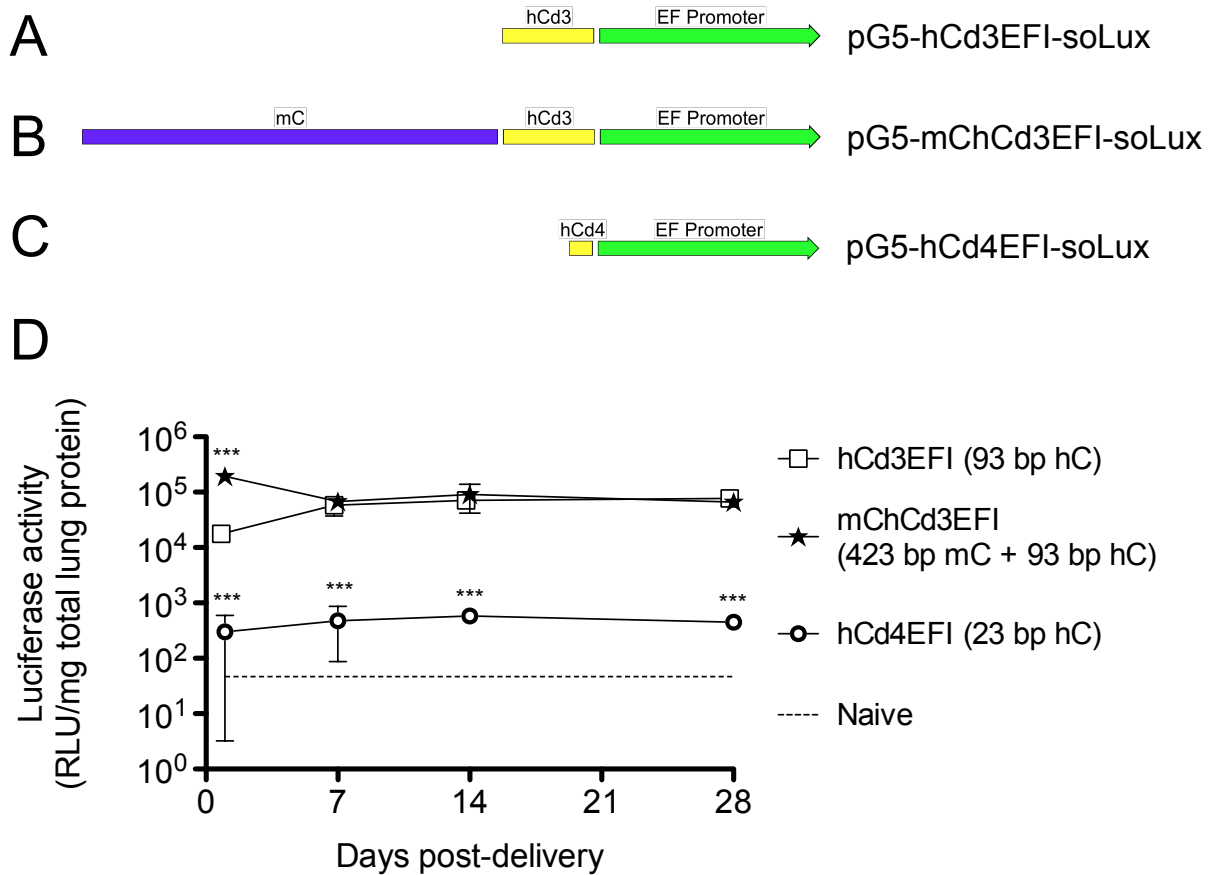


**Figure 6.1. Enhancer/promoter regions of plasmid constructs utilised in the studies in Chapter 6.**

Each CpG-free plasmid contains the synthetic version of the human elongation factor 1 $\alpha$  (EF) promoter (light green), intron (pink), firefly luciferase (soLux, dark blue), BGH polyA (dark green), R6K origin of replication (red) and the kanamycin resistance gene (light blue). Each plasmid contains a different version of an enhancer (B to H). The 5' 70 bp of the hCd3 enhancer (C) is highlighted in orange to denote the ROI that was used as the basis for further investigations. A ruler is included as a visual aid to indicate the length of the enhancer/promoter region inserted between the *Bgl*III and *Hind*III restriction sites. The length of each enhancer is listed in Table 6.1.

### 6.2.2. Effect of the mC enhancer placed upstream of the minimal hC enhancer

The result reported by Pringle *et al.* showed that when a plasmid with a dual enhancer containing the mC and hC sequences in tandem was delivered to the mouse lung, luciferase activity mirrored that of the enhancer most proximal to the EF promoter (Pringle *et al.*, 2012b). In order to assess whether the minimal hC enhancer (as in hCd3) would yield similar results to the full-length hC enhancer in the context of a synthetic murine/human dual enhancer (Pringle *et al.*, 2012b), a new dual murine/human enhancer was constructed by placing the mC enhancer at the 5'-end of the hCd3 enhancer (Figure 6.2B). This dual enhancer was compared with the single hCd3 enhancer (Figure 6.2A), and the plasmid containing the hCd4 enhancer was administered as a negative control (Figure 6.2C). Luciferase activity mediated by the dual enhancer was approximately 11-fold higher at day 1 than the hCd3 enhancer ( $p < 0.001$ ) (Figure 6.2D). The reporter gene levels dropped by day 7 to similar levels compared to the hCd3 enhancer ( $p > 0.05$ ), and remained at this level for the remainder of the experiment. This result is in agreement with that obtained by Pringle *et al.* and demonstrates that the enhancer closest to the EF promoter can mediate activity similar to a single enhancer even when a relatively short enhancer sequence is used.



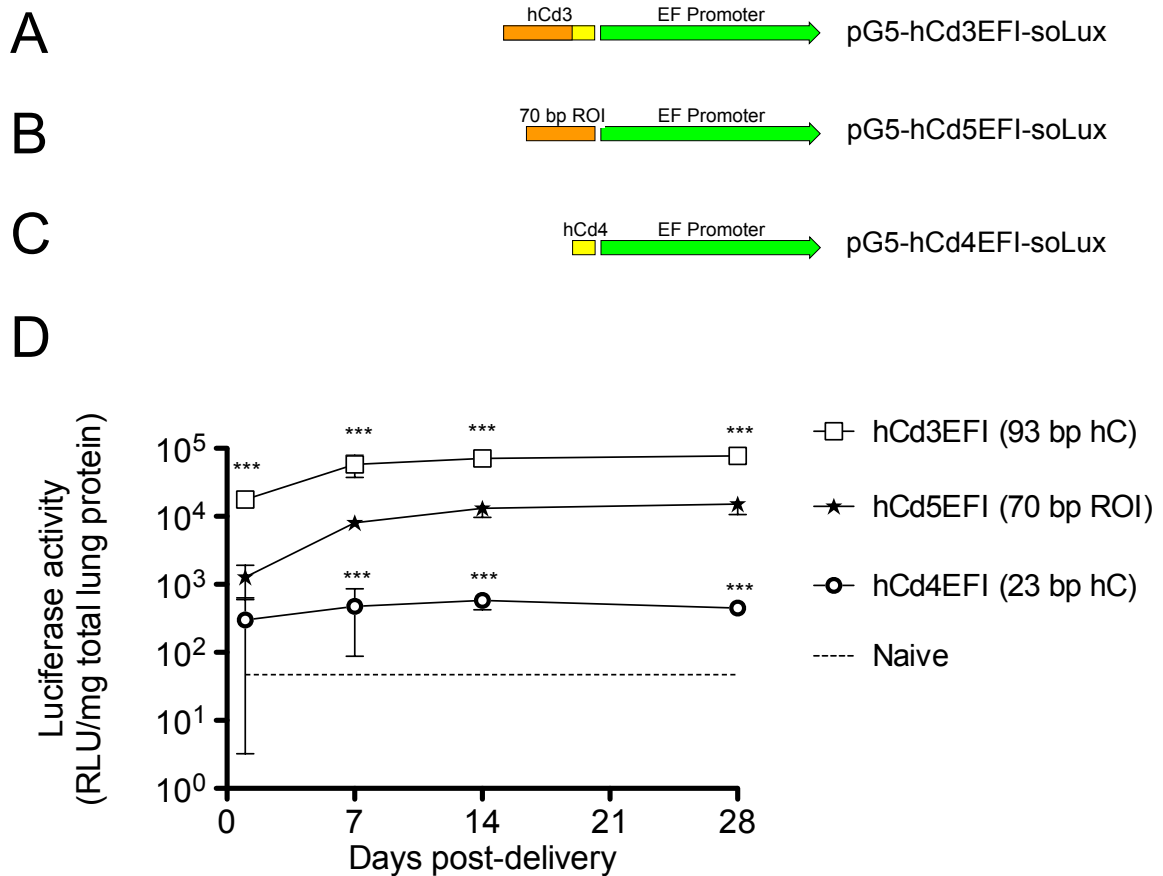
**Figure 6.2.** Comparison of luciferase activity after aerosol administration of plasmids containing the (A) hCd3 enhancer, (B) full-length mC and hCd3 dual enhancer and (C) hCd4 enhancer.

Female BALB/c mice (n = 16 per group) were exposed to 10 mL treatment aerosols containing 2 mg of the respective plasmid construct with enhancer variants (A, B, C) complexed with PEI (Section 2.5.6).

**D:** Firefly luciferase activity was analysed at the time-points indicated using standard luminometry (Section 2.6.1, n = 4 per time-point). The enhancer length is listed in brackets. Data are shown as mean  $\pm$  SD. Significant differences to values obtained from hCd3 at specific time-points were determined using two-way ANOVA followed by Bonferroni's post-hoc tests; \*\*\*  $p < 0.001$ . For plasmid details, see Table 6.1 and Figure 6.1.

### 6.2.3. Luciferase activity from the 70 bp region of interest

The previous studies have highlighted the importance of a 70 bp region in the hC enhancer located between the hCd3 and hCd4 enhancer variants (Figure 4.8). Thus, a new enhancer variant was constructed which contained only the 70 bp region of interest (ROI), termed hCd5 (Figure 6.3B). As before, the new enhancer variant was compared against the hCd3 enhancer (Figure 6.3A), while the plasmid containing hCd4 enhancer (Figure 6.3C) was administered as a negative control. As shown in Figure 6.3D, the hCd5 enhancer yielded intermediate levels of transgene activity ( $p < 0.001$ ) when compared with levels obtained by the plasmid containing the hCd3 or hCd4 enhancers. At the end of the experiment, activity mediated by hCd5 was approximately five-fold lower than the hCd3 enhancer. However, despite this reduction, reporter gene activity mediated by the hCd5 enhancer at day 14 was approximately double that observed at a similar time-point from the best mC enhancer variant (mCd3, Figure 4.9E), which means that the relative level of transgene activity still remained high. This result suggests that although the inclusion of the 3' 23 bp of the hCd3 enhancer is required for absolute maximal level of transgene activity in this model, the 70 bp ROI in the hCd5 enhancer is able to mediate high levels of luciferase activity when compared with the mC enhancer.



**Figure 6.3. Comparison of luciferase activity after aerosol administration of plasmids containing the (A) hCd3 enhancer, (B) hCd5 enhancer containing the 70 bp region of interest and (C) hCd4 enhancer.**

Female BALB/c mice (n = 16 per group) were exposed to 10 mL treatment aerosols containing 2 mg of the respective plasmid construct with enhancer variants (A, B, C) complexed with PEI (Section 2.5.6).

**D:** Firefly luciferase activity was analysed at the time-points indicated using standard luminometry (Section 2.6.1, n = 4 per time-point). The enhancer length is listed in brackets. Data are shown as mean ± SD. Significant differences to values obtained from hCd5 at specific time-points were determined using two-way ANOVA followed by Bonferroni's post-hoc tests; \*\*\*  $p < 0.001$ . For plasmid details, see Table 6.1 and Figure 6.1.

#### 6.2.4. Quantification of GT115 genomic DNA contamination

Studies by Bazzani *et al.* showed that contamination of plasmid preparations with residual bacterial chromosomal DNA could lead to increased inflammation (Bazzani *et al.*, 2011) and reduced transgene activity following gene transfer to the mouse lung (R. Bazzani, pers. comm.). Thus, one possible reason for reduced luciferase activity with plasmids containing enhancer variants (Section 6.2.3) could be the presence of residual levels of chromosomal DNA from the host bacterial strain GT115 (Section 2.4.2). To rule out this as the cause of the five-fold decrease in expression observed between plasmids containing the hCd3 and hCd5 enhancer variants (Figure 6.3D), bacterial GT115 chromosomal DNA (chrDNA) was quantified in the plasmid preparations used in Section 6.2.3 using a GT115-specific quantitative PCR assay (Section 2.7.3) (Bazzani *et al.*, 2011). Overall, levels of GT115 chrDNA contamination were similar between plasmid preparations and did not correlate with transgene activity levels at day 28 (Table 6.2). In fact, the plasmid preparation with the highest level of chrDNA contamination was pG5-hCd3EFI-soLux (6.1 %), which yielded the highest luciferase activity values in this study (Figure 6.3) and the plasmid preparation with the lowest chrDNA contamination was pG5-hCd5EFI-soLux (4.3 %). The GT115 chrDNA contamination of pG4-hCd4EFI-soLux (containing the 23 bp hC enhancer) was approximately 4.8 %. Therefore, the decrease in reporter gene activity observed when using the hCd5 enhancer when compared with the hCd3 enhancer is unlikely to be due to increased levels of GT115 chrDNA contamination.

**Table 6.2. Summary of bacterial chrDNA present in pDNA preparations.**

Plasmid name	% GT115 chrDNA contamination $\pm$ SD	Luciferase activity at day 28 $\pm$ SD (RLU/mg)
pG5-hCd3EFI-soLux	6.1 % $\pm$ 0.1 %	77,270 $\pm$ 17,069
pG5-hCd4EFI-soLux	4.8 % $\pm$ 0.2 %	450 $\pm$ 87
pG5-hCd5EFI-soLux	4.3 % $\pm$ 0.1	15,143 $\pm$ 4,530

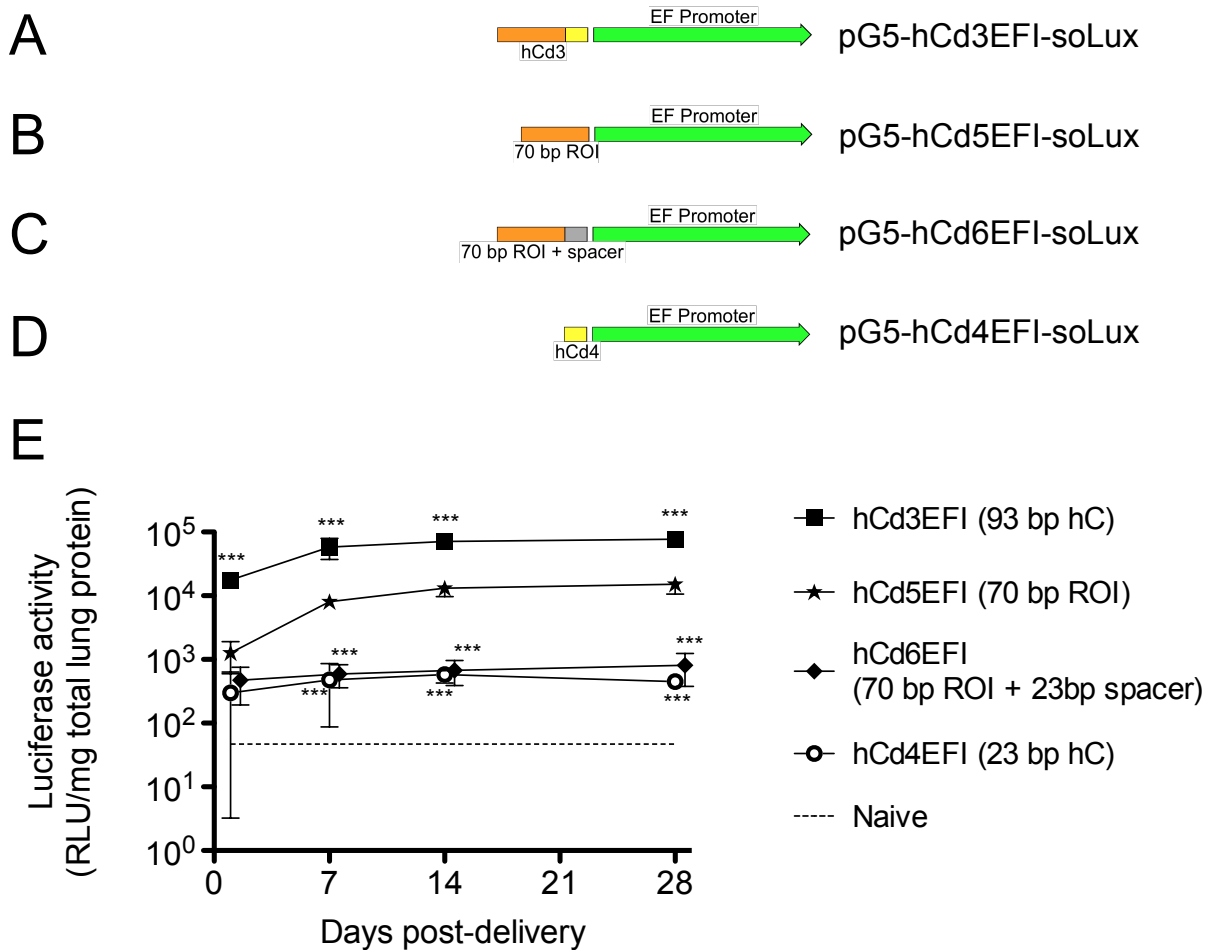
The amount of bacterial chrDNA present in pDNA preparations were determined by the GT115-specific quantitative PCR assay (Section 2.7.3) and reported as the percentage of chrDNA present in each pDNA preparation. Luciferase activity at day 28 is reported as numerical values (from data in Figure 6.3D).

### 6.2.5. Investigation of the effect of a spacer region

The result obtained in Section 6.2.3 suggests that the last 23 bp of the hCd3 enhancer has a role in mediating higher transgene activity when compared with the plasmid containing only the region of interest (as in hCd5). However, the 23 bp sequence on its own (as in hCd4) yielded transgene activity that is no different from the negative control plasmid (pG4-EFI-soLux) at all time-points (Figure 4.8). One explanation is that it is possible that the 23 bp sequence contains one or more TFBSs that act in concert with the remaining binding sites in the enhancer, or those in the EF promoter. An alternative explanation is that the 23 bp sequence is mediating stabilisation of a steric interaction between the hCd3 enhancer and the EF promoter. When this interaction is removed, as in the case of the hCd5 enhancer, the result after aerosol delivery is five-fold less reporter gene activity (Figure 6.3D).

To further understand the role of this 23 bp sequence in the hCd3 enhancer, a new enhancer variant (hCd6) was constructed. This variant contains the same 70 bp ROI as hCd3, but contains a different 23 bp region denoted as 'spacer' in Figure 6.4C. This region was identified by analysing the human alpha-2 pseudogene of the alpha-globin gene cluster with MacVector (GenBank accession number X03583.1) and selecting a 23 bp sequence which is CpG-free and contains no known TFBSs. Furthermore, both the A/T dinucleotide content and melting temperature of the hCd6 enhancer were similar to the hCd3 enhancer (data not shown). The new enhancer variant was compared against the hCd3 (Figure 6.4A) and hCd5 (Figure 6.4B) enhancers, while the plasmid containing the hCd4 enhancer (Figure 6.4D) was administered as a negative control. As shown in Figure 6.4E, the transgene activity from the hCd6 enhancer was similar to the negative control (hCd4 enhancer) at all time-points, and different from the plasmids containing the hCd5 enhancer at days 7, 14 and 28 after dosing ( $p < 0.001$ ). This result suggests that the role of the 3' 23 bp in the hCd3 enhancer is not simply one of a spacer effect; rather, it may contain requisite TFBSs for

mediating transgene activity, which on their own do not mediate transgene activity above the negative control (as in hCd4). Instead, the TFBSs in the 3' 23 bp may act in concert with the other binding sites present elsewhere in the construct.



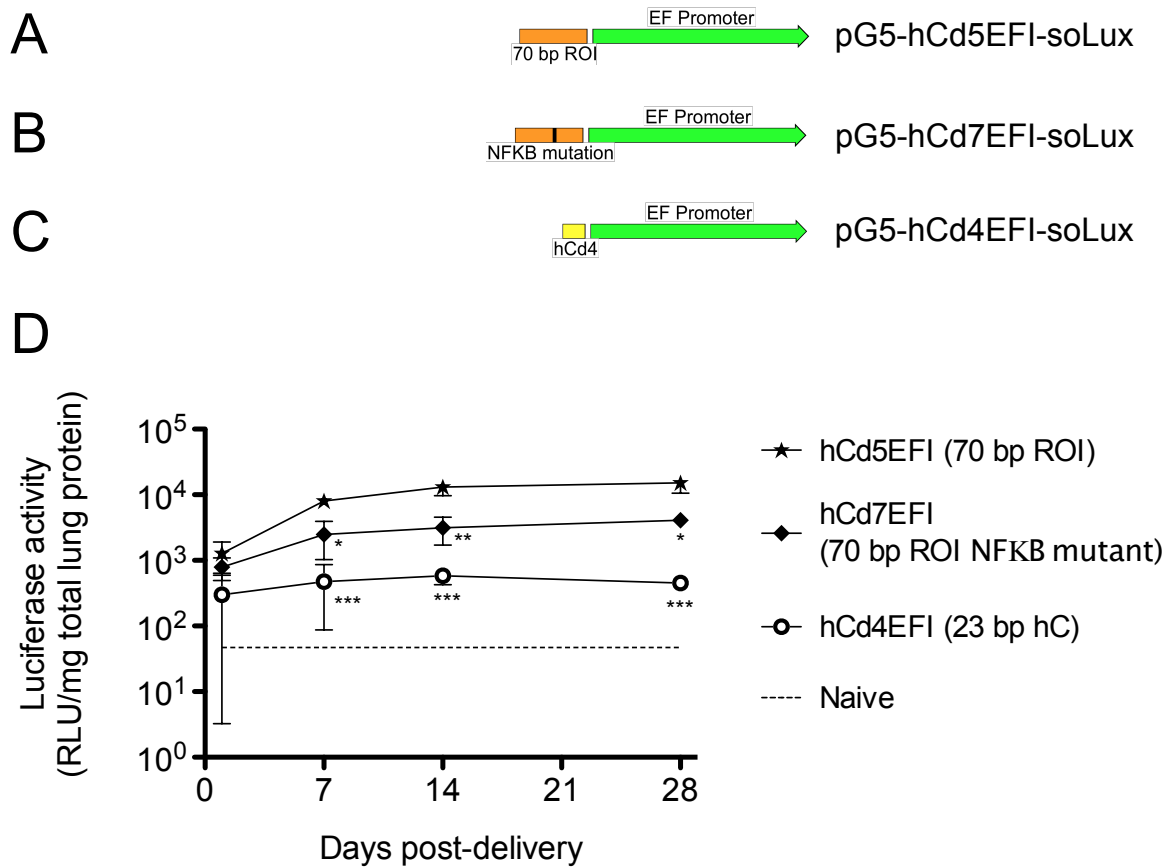
**Figure 6.4.** Comparison of luciferase activity after aerosol administration of plasmids containing the (A) hCd3 enhancer, (B) hCd5 enhancer containing the 70 bp region of interest, (C) hCd6 enhancer containing the 70 bp region of interest and a 23 bp spacer and (D) hCd4 enhancer.

Female BALB/c mice (n = 16 per group) were exposed to 10 mL treatment aerosols containing 2 mg of the respective plasmid construct with enhancer variants (A, B, C, D) complexed with PEI (Section 2.5.6).

**E:** Firefly luciferase activity was analysed at the time-points indicated using standard luminometry (Section 2.6.1, n = 4 per time-point). The enhancer length is listed in brackets. Data are shown as mean  $\pm$  SD. Data points have been shifted for clarity. Significant differences to values obtained from hCd5 at specific time-points were determined using two-way ANOVA followed by Bonferroni's post-hoc tests; \*\*\*  $p < 0.001$ . For plasmid details, see Table 6.1 and Figure 6.1.

### 6.2.6. Luciferase activity after mutation of the NF $\kappa$ B binding site

The plasmid containing the hCd5 enhancer resulted in lower transgene activity than the hCd3 enhancer, but it was still relatively high compared with plasmids containing the mC enhancer (Section 6.2.3). Therefore it was hypothesised that the 70 bp ROI contains one or more TFBSs which may confer an advantage following aerosol dosing to the mouse lung. The 70 bp ROI contains a single NF $\kappa$ B TFBS. Considering the role of NF $\kappa$ B in transcriptional activation in a variety of tissues including the lung (reviewed in Christman et al., 2000), its potential importance in mediating transgene activity from plasmid vectors in the mouse lung was investigated. Therefore, the NF $\kappa$ B TFBS was mutated (Section 2.3.6) by changing the first 3 nucleotides of the site from GGG to TCT, which has previously been shown to completely prevent the binding of the NF $\kappa$ B transcription factor (Nabel and Baltimore, 1987). This mutation was contained in the new hCd7 enhancer variant (Figure 6.5B) and was compared with the hCd5 enhancer (Figure 6.5A), with the plasmid containing the hCd4 enhancer (Figure 6.5C) administered as a negative control. As shown in Figure 6.5D, the new enhancer variant mediated intermediate levels of transgene activity at days 7 ( $p < 0.05$ ), 14 ( $p < 0.01$ ) and 28 ( $p < 0.05$ ) after aerosol dosing when compared with transgene activity obtained by the plasmid containing the hCd5 enhancer. Luciferase activity mediated by the hCd7 enhancer at day 28 was approximately four-fold lower than transgene activity from the hCd5 enhancer, but approximately nine-fold higher than activity from the negative control plasmid containing the hCd4 enhancer. These results suggest that although the destruction of the NF $\kappa$ B binding site can affect transgene activity, it does not completely abolish activity mediated by this enhancer sequence.



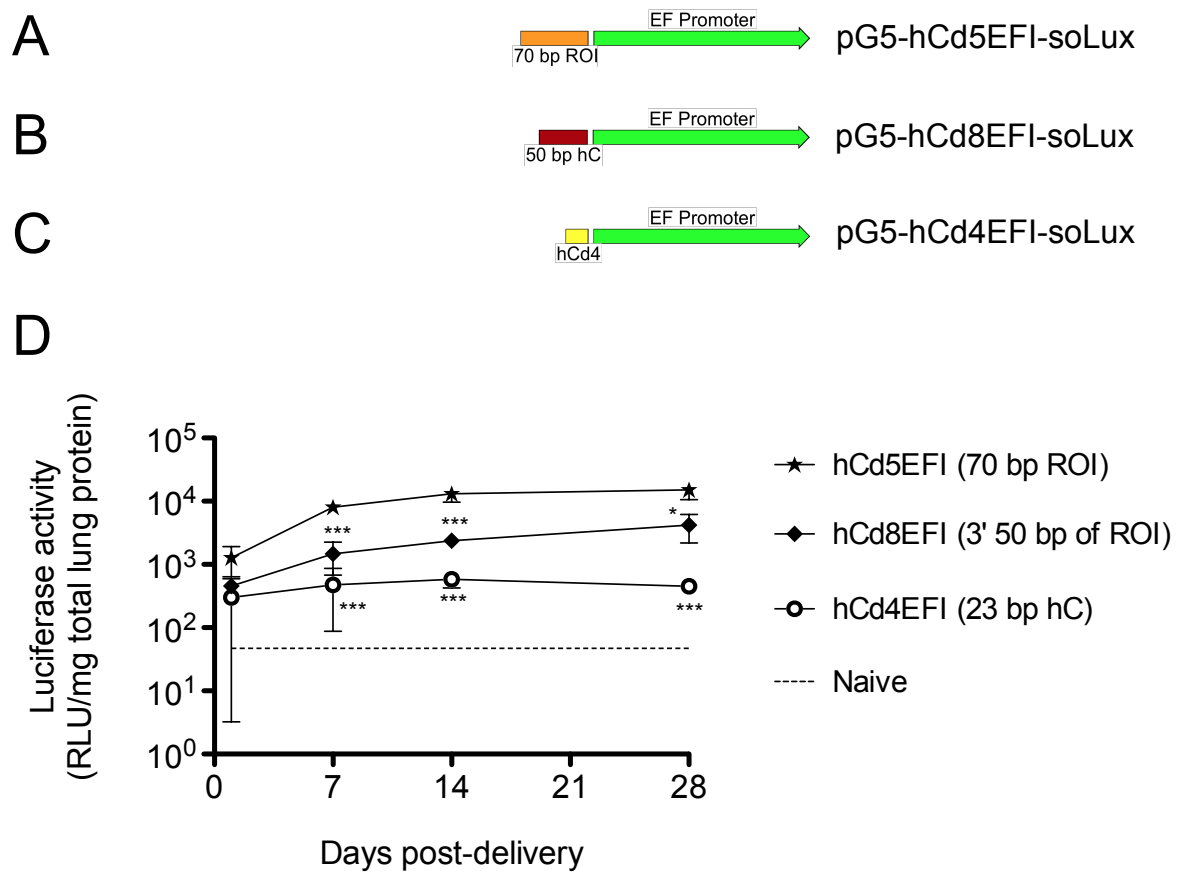
**Figure 6.5. Comparison of luciferase activity after aerosol administration of plasmids containing the (A) hCd5 enhancer containing the 70 bp region of interest, (B) hCd7 enhancer containing the 70 bp region of interest with a mutated NFκB binding site and (C) hCd4 enhancer.**

Female BALB/c mice (n = 16 per group) were exposed to 10 mL treatment aerosols containing 2 mg of the respective plasmid construct with enhancer variants (A, B, C) complexed with PEI (Section 2.5.6).

**D:** Firefly luciferase activity was analysed at the time-points indicated using standard luminometry (Section 2.6.1, n = 4 per time-point). The enhancer length is listed in brackets. Data are shown as mean ± SD. Significant differences to values obtained from hCd5 at specific time-points were determined using two-way ANOVA followed by Bonferroni's post-hoc tests; \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ . For plasmid details, see Table 6.1 and Figure 6.1.

### 6.2.7. Luciferase activity from a 50 bp enhancer

An additional modification of the hCd5 enhancer containing the 70 bp ROI was made to further investigate the importance of specific binding sites in this sequence. The 5' 20 bp of the hCd5 enhancer was deleted in order to generate the hCd8 enhancer, retaining just the 3' 50 bp of the 70 bp ROI (Figure 6.6B). The deleted region contains a number of TFBSs, including a unique NF1 site (Figure 5.3) which is present only once in the entire hC sequence and completely absent from the mC enhancer. The NF $\kappa$ B binding site was not destroyed by this deletion. The new enhancer variant was compared with the hCd5 enhancer (Figure 6.6A), while the plasmid containing the hCd4 enhancer was administered as a negative control (Figure 6.6C). As shown in Figure 6.6D, the hCd8 enhancer mediated intermediate levels of luciferase activity, which was statistically different on days 7, 14 ( $p < 0.001$ ) and 28 ( $p < 0.05$ ) after aerosol dosing when compared with transgene activity obtained by the plasmid containing the hCd5 enhancer. Luciferase activity mediated by the hCd8 enhancer at day 28 was approximately four-fold lower than transgene activity from the hCd5 enhancer, but approximately nine-fold higher than activity from the negative control plasmid containing the hCd4 enhancer. This result suggests that the removal of the 5' 20 bp may have an effect on the overall levels of transgene activity but does not completely suppress activity.



**Figure 6.6. Comparison of luciferase activity after aerosol administration of plasmids containing the (A) hCd5 enhancer containing the 70 bp region of interest, (B) hCd8 enhancer containing the 3' 50 bp of the 70 bp region of interest and (C) hCd4 enhancer.**

Female BALB/c mice ( $n = 16$  per group) were exposed to 10 mL treatment aerosols containing 2 mg of the respective plasmid construct with enhancer variants (A, B, C) complexed with PEI (Section 2.5.6).

**D:** Firefly luciferase activity was analysed at the time-points indicated using standard luminometry (Section 2.6.1,  $n = 4$  per time-point). The enhancer length is listed in brackets. Data are shown as mean  $\pm$  SD. Significant differences to values obtained from hCd5 at specific time-points were determined using two-way ANOVA followed by Bonferroni's post-hoc tests; \*\*\*  $p < 0.001$ , \*  $p < 0.05$ . For plasmid details, see Table 6.1 and Figure 6.1.

### 6.3. Discussion

The studies performed in this chapter have demonstrated that the minimal 93 bp hC enhancer is necessary for mediating high-level transgene activity similar to the full-length hC enhancer. Any further deletions or mutations of the minimal enhancer resulted in significantly lower levels of transgene activity after aerosol delivery to the mouse lung. However, there seems to be a combinatorial effect of binding sites on the conferral of high-level transgene activity. Of the deletions tested, none resulted in the complete abrogation of reporter gene activity in the mouse lung (Figures 6.3, 6.5 and 6.6). Despite a reduction in overall transgene activity with these deletions, it is difficult to pinpoint which binding site(s) may have been involved. For example, even in the case of the enhancer variant described as containing a single NF $\kappa$ B mutation (hCd7, Figure 6.5), additional TFBSs were destroyed (i.e. OLF1.01 and NFAT5.02) whereas others were created (i.e. RORGAMMA.01 and ERG.02) (MatInspector analysis; data not shown). Therefore, although the enhancer deletions may identify putatively important sequences that are required for high-level activity, the results must be interpreted with caution. These functional studies cannot easily differentiate between the multitude of interactions that may be occurring. Each combination of TFBSs may direct a discrete transcriptional profile and as such, the results must be considered within the context of that particular plasmid molecule.

The replacement of the original 3' 23 bp with that of a 'spacer' region resulted in luciferase activity similar to the negative control plasmid (Figure 6.4). This was unexpected and the reasons are unclear, especially considering that the plasmid containing the 70 bp ROI on its own (as in hCd5) yields relatively high levels of persistent transgene activity. One reason may be that the new spacer region introduced TFBSs that may bind TF proteins that are negative regulators of transcription. Although the analysis of the 23 bp spacer region by MacVector yielded no known TFBSs, the MatInspector program identified 11 TFBSs in the spacer region (data not shown). These binding sites may be responsible for the decrease in

transgene activity. Interestingly, analysis of the original 23 bp sequence (i.e. hCd4) by MacVector identified 10 TFBS (Table 5.1) whereas MatInspector did not identify any (Table 5.3). This result highlights the importance of cross-referencing a putative sequence with more than one TFBS analysis tool. Another possible reason for the drastic reduction in the level of transgene activity could be due to the inadvertent introduction of a sequence which may result in either steric hindrance or cooperative binding of TFBSs in the enhancer and promoter. Depending on whether the binding sites act as repressors or activators of transcriptional activity, the functional consequence of such structural changes may impact on overall transgene activity.

One particularly encouraging outcome in this chapter is the result obtained with the dual enhancer plasmid containing the full-length mC enhancer positioned 5' to the minimal 93 bp hC enhancer (Figure 6.2). These results are in agreement with the results obtained by Pringle *et al.* whereby the overall expression profile of the dual enhancer simply reflected the transgene activity associated with the single enhancer located immediately 5' of the promoter (Pringle *et al.*, 2012b). However, considering the short sequence of the minimal enhancer, it was surprising that such an enhancer could mediate high-level transgene activity in the presence of the mC enhancer, which consists of 423 bp. Interestingly, there are eight overrepresented families of TFBSs in common between the dual mChCd3 enhancer and the EF promoter. When comparing this result to the data obtained in Table 5.12, the number is similar to the number of shared overrepresented families in the mC, mCd1 and mCd2 enhancers which each mediate relatively low levels of transgene activity (Figure 4.9) compared with the mChCd3 enhancer. Thus, the number of overrepresented binding sites in the mChCd3 enhancer may not have an effect on the level of luciferase activity. The simplest explanation is that by placing the minimal 93 bp hC enhancer immediately before the EF promoter, the required interactions are still able to take place despite a relatively large upstream sequence which, in the context of this plasmid at least, may not play a role in transcription.

In summary, the results presented in this chapter have demonstrated that the minimal 93 bp hC enhancer cannot be further truncated without having a negative impact on transgene activity. Thus, the minimal 93 bp hC enhancer is necessary and sufficient to mediate high-level, sustained transgene activity in the mouse lung after aerosol delivery. The studies so far have focused on the detection of reporter gene activity in the lung. Additionally, all the variants have been evaluated in conjunction with the EF promoter. Therefore, there is scope to investigate the wider utility of the minimal enhancer in another organ, or with a different promoter. Furthermore, the short length of the minimal 93 bp hC enhancer could potentially be exploited in order to generate a plasmid containing a dual tandem copy of the minimal enhancer. These possibilities are further explored in Chapter 7.

---

## Chapter 7: Utility of the minimal hC enhancer

### 7.1. Introduction

The studies described so far have investigated luciferase reporter activity of plasmids containing different mC and hC enhancer variants in conjunction with the EF promoter. These enhancer and promoter sequences are based on the wild-type versions of the mCMV (Hyde et al., 2008) and hCMV (Yew et al., 2002) enhancers and the elongation factor 1 $\alpha$  (EF1 $\alpha$ ) promoter (Hyde et al., 2008), respectively. However, these sequences contain CpGs with potential to elicit a TLR9-dependent inflammatory response (Section 1.3.4). This can result in increased levels of pro-inflammatory cytokines, such as TNF $\alpha$  and IFN $\gamma$ , potentially involved in promoter attenuation of viral promoters (Li et al., 1999; Qin et al., 1997). Therefore, the CG dinucleotides present in the hCEFI and mCEFI enhancer and promoter sequences were previously changed to generate versions entirely free of CpGs, largely by changing the CG dinucleotides to TG (Hyde et al., 2008; Yew et al., 2002). The benefits of CpG depletion in the context of gene transfer have already been discussed (Section 1.3.4), including reduced inflammation, and prolonged transgene activity (Hodges et al., 2004; Hyde et al., 2008; Yew et al., 2002). One potential disadvantage is that since many TFBSs contain CpGs it is possible that some critical binding sites may have been destroyed, or new ones created, by the change of CG dinucleotides to alternative non-stimulating dinucleotide sequences.

The CpG-depletion approach has been successful for the development of the synthetic hCEFI enhancer/promoter (Hyde et al., 2008); however, the identification of naturally CpG-free enhancers and promoters could also be beneficial, particularly in establishing new promoters for alternative gene therapy applications. Pringle *et al.* designed a database to enable the rapid identification of naturally-occurring promoters which are CpG-free and identified the promoter for the murine fatty acid binding protein 4 (mFABP) gene to be most promising when combined with the hC enhancer (Pringle et al., 2012a). It may be informative

to combine the minimal 93 bp hC enhancer and the mFABP promoter to determine transcriptional activity after aerosol dosing to the mouse lung.

The use of CpG-depleted plasmids in non-viral gene transfer applications other than the lung has been investigated. Transgene activity in the liver may be affected by the presence of CpGs (Hodges et al., 2004; Yew et al., 2001; Yew et al., 2002) although in some studies this effect was minimal in the liver (Chen et al., 2008). Interestingly, the CpG-free hCEFI enhancer/promoter does not yield sustained levels of transgene activity in the liver after hydrodynamic tail vein (HTV) injection in BALB/c mice (Pringle et al., 2010), but high-level transgene activity was possible in the livers of immuno-deficient mice (Magnusson et al., 2011). Therefore, immuno-deficient mice may be a suitable model in which to assess the utility of the minimal 93 bp hC enhancer after HTV delivery to the liver.

The studies described in this chapter were designed to explore the utility of the minimal 93 bp hC enhancer in a variety of ways. Thus far, the enhancer variants described have been operably linked to the EF promoter; therefore, the ability of the minimal hC enhancer to direct persistent transgene activity with an alternative CpG-free promoter was assessed. Secondly, the effect of the minimal 93 bp hC enhancer on transgene expression in the liver was investigated, specifically in the mouse liver following HTV injection. Finally, the potential for the minimal 93 bp hC enhancer to affect transgene activity when placed as a tandem dual enhancer was evaluated following aerosol delivery to the mouse lung.

## **7.2. Results**

### **7.2.1. Plasmid vectors used in Chapter 7 studies**

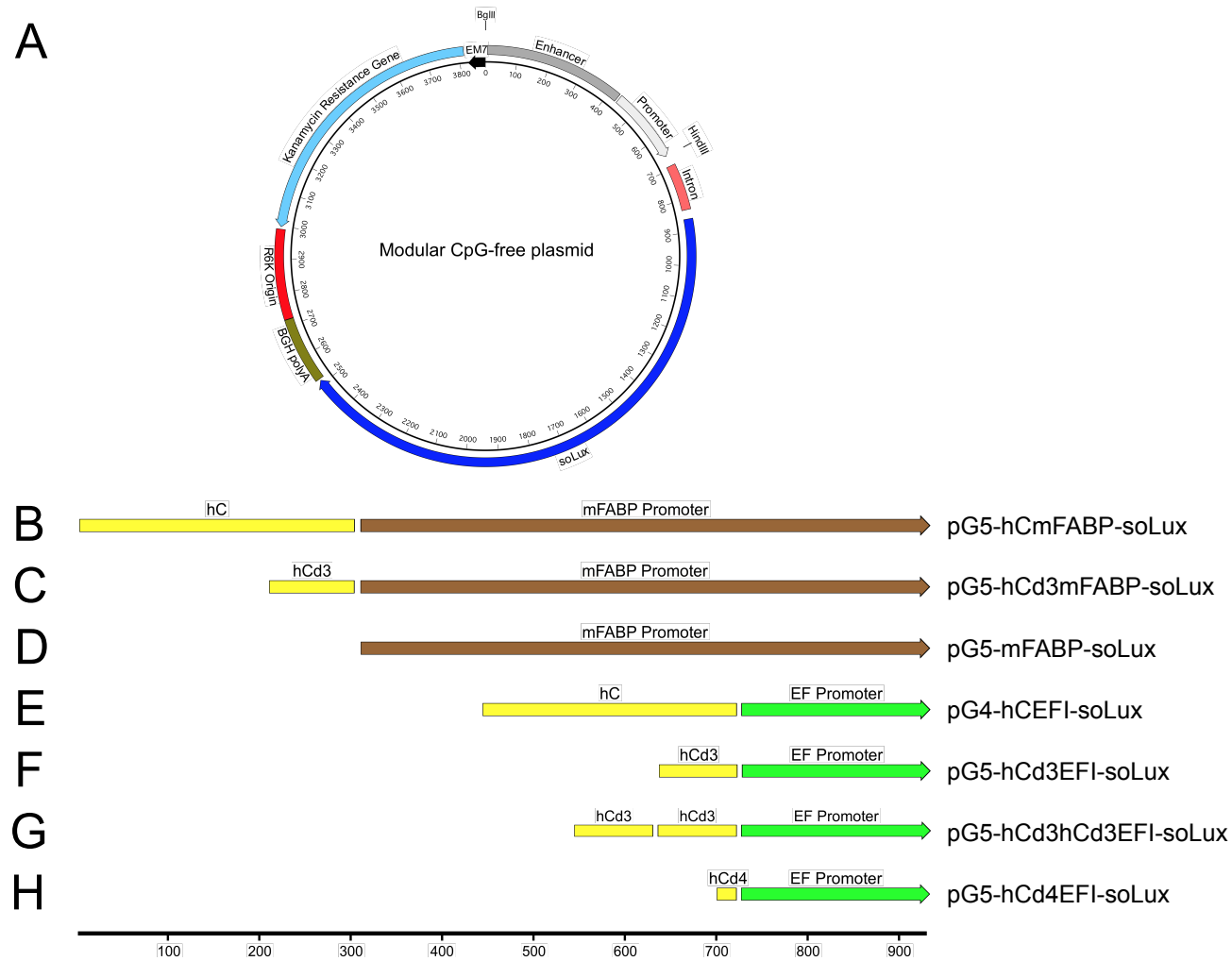
Plasmids containing new enhancer and/or promoter variants (Table 7.1 and Figure 7.1) were constructed by PCR amplification of the desired fragments (Section 2.3.1) and subsequently inserted into a pG5 plasmid fragment as described in Section 4.2.3. All plasmids used in these studies are based on a modular CpG-free design (Hyde et al., 2008) with unique restriction sites separating the enhancer/promoter components from the remainder of the plasmid (Figure 7.1A).

**Table 7.1. Overview of plasmid name, original source and enhancer length.**

Plasmid name	hC Enhancer length	Description	Figure 7.1 reference
pG5-hCmFABP-soLux	302 bp	Full-length hC enhancer placed before the mFABP promoter	B
pG5-hCd3mFABP-soLux	93 bp	Minimal enhancer placed before the mFABP promoter	C
pG5-mFABP-soLux	0 bp	No enhancer before the mFABP promoter	D
pG4-hCEFI-soLux	302 bp	Original fourth generation plasmid containing the full-length hC enhancer	E
pG5-hCd3EFI-soLux	93 bp	Minimal enhancer previously identified which mediates transgene activity similar to full-length hC (pG4-hCEFI-soLux)	F
pG5-hCd3hCd3EFI-soLux	93 bp + 93 bp	Two copies of the minimal enhancer constructed as a dual tandem enhancer	G
pG5-hCd4EFI-soLux	23 bp	Enhancer previously identified which mediates transgene activity similar to the enhancerless plasmid pG4-EFI-soLux	H

<sup>1</sup> Hyde *et al.*, 2008

The table lists the pDNA used in the enhancer studies presented in this chapter. The plasmid name, with the enhancer/promoter combination is detailed, along with the enhancer length. See Figure 7.1 for diagrammatical representations of the enhancer/promoter regions.



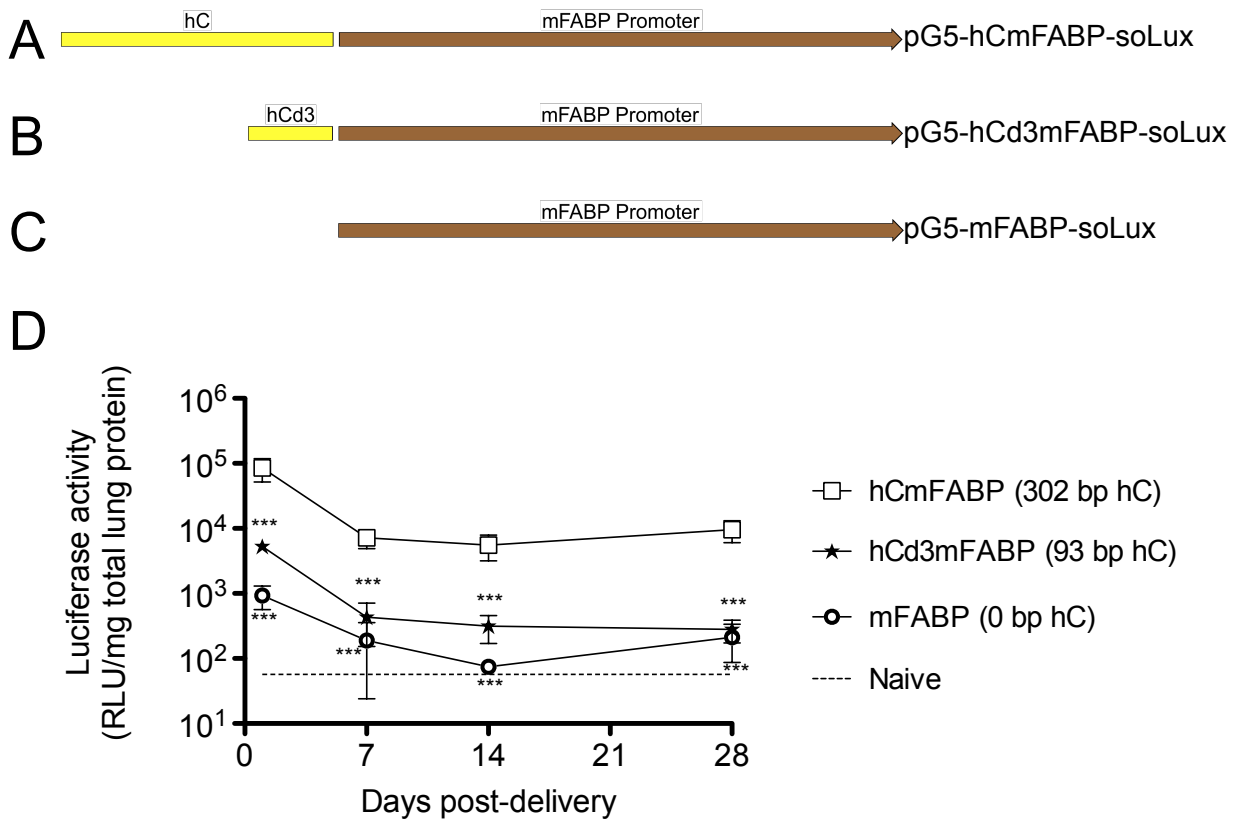
**Figure 7.1. Enhancer/promoter regions of plasmid constructs utilised in the studies in Chapter 7.**

Each CpG-free plasmid contains an intron (pink), firefly luciferase (soLux, dark blue), BGH polyA (dark green), R6K origin of replication (red) and the kanamycin resistance gene (light blue). Each plasmid varies in the enhancer and/or promoter region (B to H). A ruler is included as a visual aid to indicate the length (bp) of the enhancer/promoter region inserted between the *Bgl*III and *Hind*III restriction sites. The length of each enhancer is listed in Table 7.1.

### 7.2.2. Using the minimal hC enhancer with an alternative CpG-free promoter

To determine if the minimal 93 bp hC enhancer could mediate levels of persistent transgene activity similar to the full-length hC enhancer in conjunction with another CpG-free promoter, a new construct was made by inserting the hCd3 enhancer 5' to the mFABP promoter (Figure 7.2B). This new construct was compared for transgene activity after aerosol delivery alongside the hCmFABP enhancer/promoter (Pringle et al., 2012a) (Figure 7.2A). A plasmid containing the mFABP promoter without an enhancer (Figure 7.2C) was also administered. Figure 7.2D shows that luciferase activity mediated by the plasmid containing the hCmFABP enhancer/promoter resulted in high day 1 values, which then fell approximately 12-fold by day 7 and remained at this level until the end of the study. The hCd3mFABP enhancer/promoter directed luciferase activity at day 1 which was approximately 16-fold lower than what was observed when the hCmFABP enhancer/promoter was used ( $p < 0.001$ ). Furthermore, luciferase activity fell by approximately 19-fold between day 1 and day 28, resulting in a 34-fold difference at day 28 between hCmFABP and hCd3mFABP ( $p < 0.001$  for all time-points). The plasmid containing the mFABP promoter without an enhancer mediated lower reporter gene activity at all time-points ( $p < 0.001$ ).

This result suggests that the minimal hC enhancer does not uniformly mediate activity levels that are similar to the full-length hC enhancer when operably linked to another CpG-free promoter. This suggests that the persistent levels of luciferase activity observed from hCEFI in the mouse lung are likely to be due to a specific interaction between the minimal hC enhancer and the EF promoter, rather than a property intrinsic to the minimal enhancer.



**Figure 7.2. Comparison of luciferase activity after aerosol administration of plasmids containing the mFABP promoter and (A) hC enhancer, (B) hCd3 enhancer and (C) no enhancer.**

Female BALB/c mice ( $n = 24$  per group) were exposed to 10 mL treatment aerosols containing 2 mg of the respective plasmid construct with enhancer variants (A, B, C) complexed with PEI (Section 2.5.6).

**D:** Firefly luciferase activity was analysed at the time-points indicated using standard luminometry (Section 2.6.1,  $n = 6$  per time-point). The enhancer length is listed in brackets. Data are shown as mean  $\pm$  SD. Significant differences to values obtained from hCmFABP at specific time-points were determined using two-way ANOVA followed by Bonferroni's post-hoc tests; \*\*\*  $p < 0.001$ . For plasmid details, see Table 7.1 and Figure 7.1.

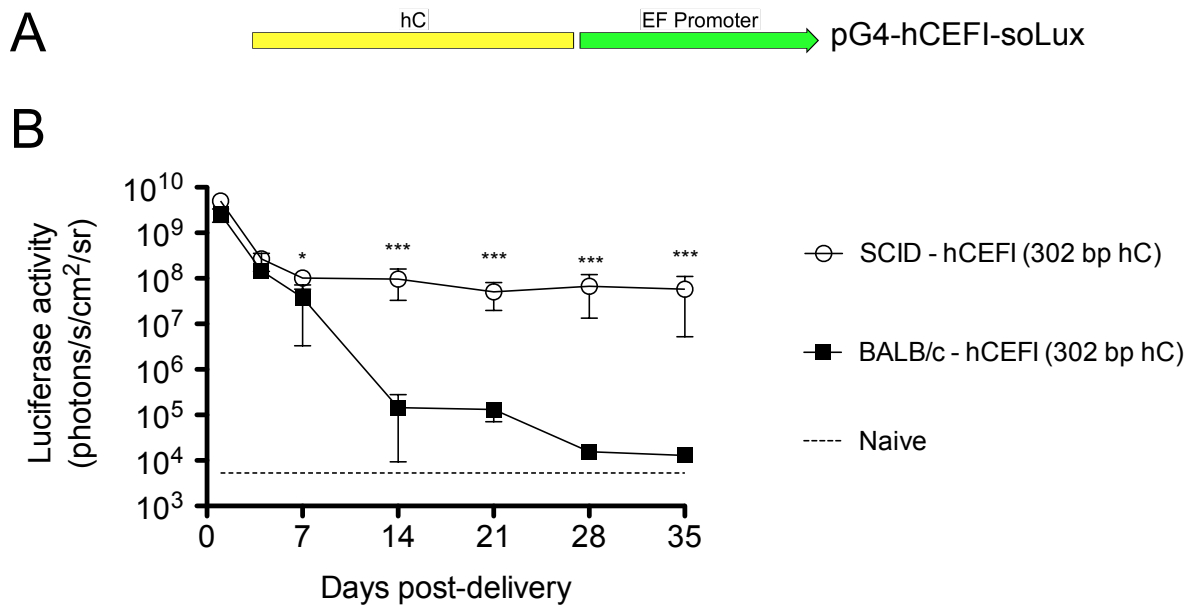
### 7.2.3. Comparison of luciferase activity after hydrodynamic tail vein injection

In order to evaluate the effect of the 93 bp minimal hC enhancer on transgene activity in an organ other than the lung, pDNA was delivered to the liver via hydrodynamic tail vein (HTV) injection. As described previously, the hCEFI enhancer/promoter did not mediate sustained activity in the liver after delivery to BALB/c mice (Pringle et al., 2010) possibly due to an anti-transgene response to the highly immunogenic luciferase protein (Podetz-Pedersen et al., 2014). Thus, mice harbouring a severe combined immunodeficiency (SCID) mutation, bred on a BALB/c background, were evaluated as a model to measure persistence of transgene activity after HTV injection of plasmids containing the hCEFI enhancer/promoter. Mice homozygous for the SCID mutation are severely deficient in functional B and T lymphocytes which affects adaptive immunity, but retain a normal innate immune response with a normal population of functional NK and macrophage cells (Bosma and Carroll, 1991). Naked DNA (Section 1.2.8) was used for HTV injections because lipid-based carriers tend to aggregate with proteins or cells after intravenous administration, leading to the disintegration of lipid vectors, DNA release and degradation (Sakurai et al., 2001). A recent publication has described the use of a PEI-based gene transfer agent to increase the level of transfected liver cells after HTV (Nakamura et al., 2013); however, given the large number of transfected cells after HTV injection of naked DNA (Liu et al., 1999; Zhang et al., 1999), other formulations were not tested.

#### 7.2.3.1. Establishing high-level transgene persistence in immuno-deficient mice

Plasmid DNA containing the hCEFI enhancer/promoter (Figure 7.3A) was delivered to the livers of BALB/c and SCID mice using HTV (Section 2.5.7). As seen in Figure 7.3B, luciferase activity (Section 2.6.3) in both mouse strains was similar at days 1 and 4 after plasmid administration. However, luciferase activity in the BALB/c strain fell abruptly after day 7 and by day 35, at the point of animal sacrifice, reporter gene activity was approximately 0.03 % of the mean day 7 values. Conversely, luciferase activity in the livers of the SCID mice remained relatively steady after the initial decrease. Indeed, the level of

transgene activity at the termination of the study (day 35) was not statistically from the level observed at day 7. However, the difference between the strains at day 35 was approximately 5000-fold and statistically significant ( $p < 0.001$ ). Together, these results suggest that the SCID mouse model may be more appropriate to compare the effect of the minimal 93 bp hC enhancer and the full-length hC enhancer after gene transfer to the mouse liver than the BALB/c strain. Furthermore, these data support the suggestion that the adaptive immune response against the transgene product may be responsible for the decrease in transgene activity between observed between day 7 and day 35 in the BALB/c mice.



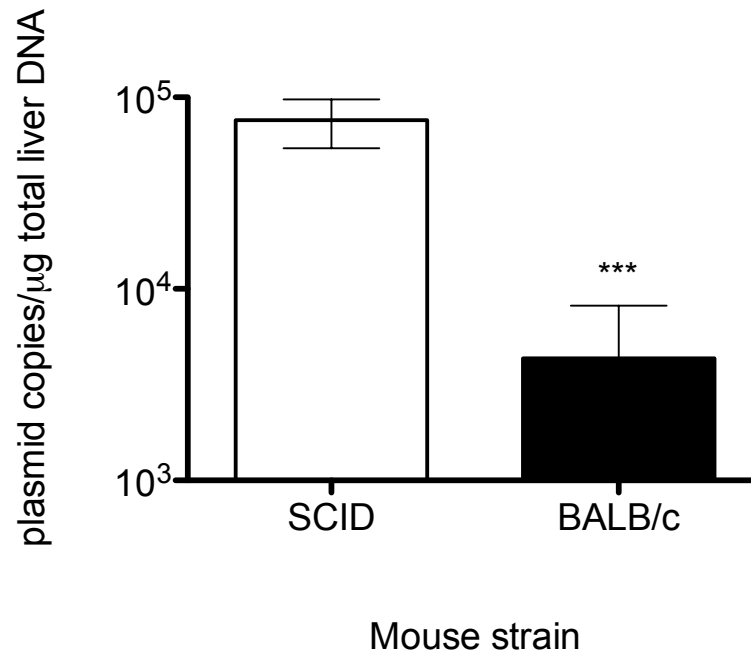
**Figure 7.3. Comparison of luciferase activity after hydrodynamic tail vein injection dosing of pG4-hCEFI-soLux in SCID and BALB/c mice.**

Female mice (n = 6-7 per strain) were dosed with pG4-hCEFI-soLux (A) at 10  $\mu\text{g}/\text{mL}$  diluted in PBS by hydrodynamic tail vein injection over 5 seconds in a volume equal to 10 % bodyweight (approximately 1.8 mL / 18  $\mu\text{g}$  per mouse) (Section 2.5.7).

**B:** Liver firefly luciferase activity was measured at the time-points indicated using BLI (Section 2.6.3). The enhancer length is listed in brackets. Data are shown as mean  $\pm$  SD. Significant differences at specific time-points were determined using two-way ANOVA followed by Bonferroni's post-hoc tests; \*  $p < 0.05$ , \*\*\*  $p < 0.001$ .

### 7.2.3.2. Quantification of plasmid DNA in the mouse liver

The result shown in Figure 7.3B suggests that the adaptive immune system may be responsible for the precipitous fall in luciferase activity after HTV injection into immuno-competent BALB/c mice. If correct, the destruction of transfected cells could be a contributing factor; therefore, the livers of the treated mice were harvested after the final imaging time-point on day 35 (Section 2.5.9) and analysed for pDNA copies by quantitative real-time PCR (Section 2.7.2). As shown in Figure 7.4, the mean number of plasmid copies in the SCID mice was approximately  $7.5 \times 10^5/\mu\text{g}$  of total liver DNA, compared with approximately  $4 \times 10^4$  copies/ $\mu\text{g}$  in the BALB/c mice ( $p < 0.001$ ). This result suggests that transfected cells expressing the luciferase protein may have been destroyed by a component of the adaptive immune system.



**Figure 7.4. Quantification of plasmid DNA in the liver of SCID and BALB/c mice.**

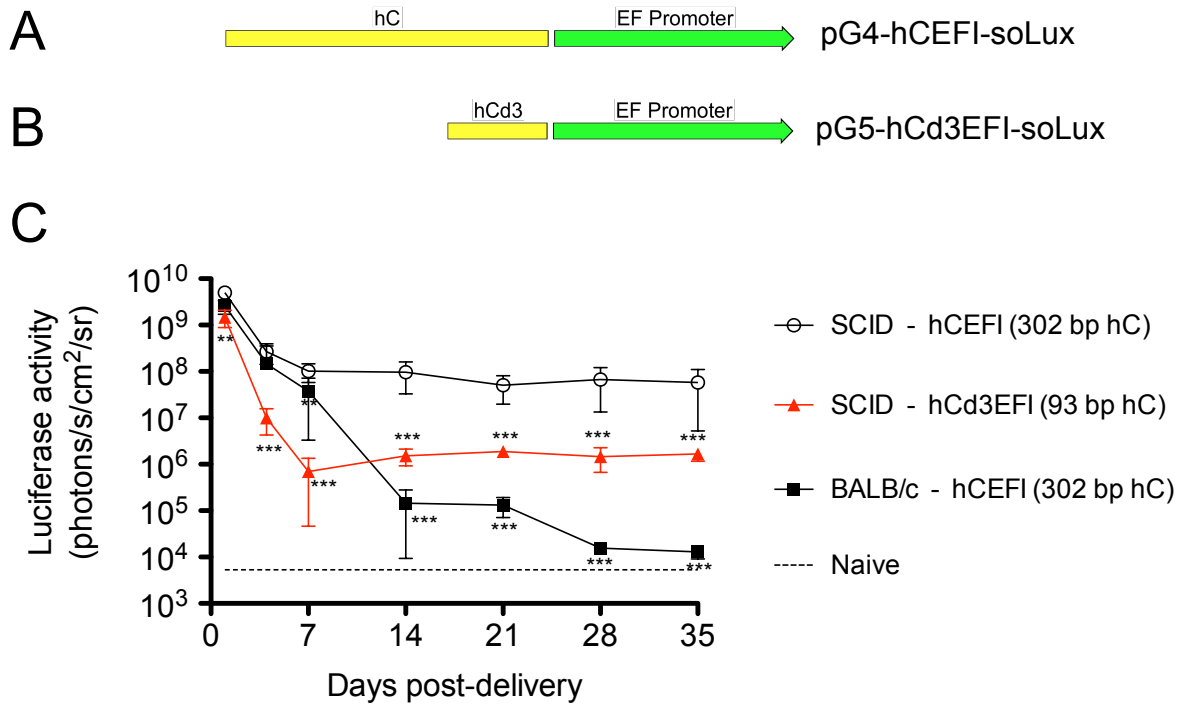
Liver tissue from the SCID and BALB/c mice was harvested (Section 2.5.9) and DNA was extracted (Section 2.7.1). Absolute levels of pDNA were determined at 35 days post-dosing by quantitative real-time PCR (Section 2.7.2) and graphically represented as plasmid copies per μg of total liver DNA. Data are shown as mean ± SD. Statistical difference was analysed by an unpaired t-test: \*\*\*  $p < 0.001$ .

### 7.2.3.3. Utility of the minimal hC enhancer in the liver

High-level activity after HTV injection of the plasmid harbouring the hCEFI enhancer/promoter was demonstrated in the SCID mouse model (Figure 7.3B). The ability of the minimal 93 bp hC enhancer (Figure 7.5B) to mediate transgene activity in the mouse liver was then compared with the full-length hC enhancer (Figure 7.5A) after HTV injection (Section 2.5.7) in SCID mice. As seen in Figure 7.5C, the minimal enhancer (red triangles) did not result in reporter gene activity similar to the full-length enhancer in SCID mice (open circles). Indeed, the fall in transgene activity was more severe in the first 7 days than observed in either the SCID or BALB/c mice after delivery of the plasmid containing the full-length hC enhancer. The day 7 values obtained from the minimal 93 bp hC enhancer yielded approximately 0.7 % of the luciferase activity mediated at the same time-point by the full-length enhancer in the SCID mice ( $p < 0.001$ ). Interestingly, transgene activity was stable after this initial fall, as observed with the hCEFI enhancer/promoter in SCID mice, and in contrast with the values observed in the BALB/c mice where luciferase activity continues to fall after day 7. However, the mean value at day 35 obtained with the minimal 93 bp hC enhancer was approximately 2 % of the transgene activity level obtained in SCID mice with the full-length hC enhancer ( $p < 0.001$ ).

These results indicate that in the absence of a confounding factor, such as the role of the adaptive immune system, it is possible to differentiate between a strong enhancer (i.e. hC) and a weak enhancer (i.e. hCd3) in the mouse liver after HTV injection. Additionally, it appears that the 5' 209 bp of the hC enhancer (absent in the hCd3 enhancer variant) may contain TFBSs required for mediating high-level transgene activity in the liver. In contrast, the removal of this sequence has been shown to be inconsequential for directing high-level transgene activity after aerosol delivery to the mouse lung (Figure 4.8). The difference in luciferase activity observed after administration to the liver and lung using the same plasmid may indicate variations in endogenous gene expression. A transcriptome analysis of the

liver and lung, based on the identification of putative TFBSs in the minimal 93 bp hC enhancer, may highlight gene expression differences in the context of understanding why the minimal enhancer mediates persistent transgene activity in the lung but not the liver.



**Figure 7.5. Comparison of luciferase activity in SCID and BALB/c mice after dosing by hydrodynamic tail vein injection of plasmids containing the (A) hC enhancer and (B) hCd3 enhancer.**

Female mice ( $n = 6-7$  per strain) were dosed with pG5-hCd3EFI-soLux (B) at  $10 \mu\text{g/mL}$  diluted in PBS by hydrodynamic tail vein injection over 5 seconds in a volume equal to 10 % bodyweight (approximately  $1.8 \text{ mL} / 18 \mu\text{g}$  per mouse) (Section 2.5.7).

**C:** The data from the hCd3 enhancer has been plotted onto the graph from Figure 7.3. Liver firefly luciferase activity was measured at the time-points indicated using BLI (Section 2.6.3). The enhancer length is listed in brackets. Data are shown as mean  $\pm$  SD. Significant differences to values obtained from SCID-hCEFI at specific time-points were determined using two-way ANOVA followed by Bonferroni's post-hoc tests; \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

#### 7.2.3.4. Gene expression in mouse lung and liver

The MatInspector analysis of TFBSs in the hCd3 enhancer (Table 5.3) predicts that there are 32 different individual matrices represented in the minimal 93 bp hC enhancer sequence. Of these 32 binding sites, a number of related sites were grouped together (namely, NF1 binding sites and estrogen-related receptor sites) for the purposes of further gene expression analysis. Furthermore, two of the predicted sites bind transcription factor (TF) proteins that consist of heterodimers. To this end, a total of 25 genes encoding the candidate TF proteins were further analysed.

The name of each relevant mouse TF protein was identified using the Universal Protein Resource ([www.uniprot.org](http://www.uniprot.org), UniProt) (Huntley et al., 2015). The gene identifier, as indicated in UniProt, was then entered into the BioGPS gene annotation database ([www.biogps.org](http://www.biogps.org)) (Wu et al., 2009), a publicly available repository of microarray data maintained by The Scripps Research Institute (La Jolla, CA, USA), which allows researchers to upload their microarray data as 'datasets'. Users of the system can then query the datasets and extract the relevant gene expression data for their own analysis. One such dataset is the Mouse MOE430 Gene Atlas, based on the Affymetrix Mouse Genome 430 2.0 arrays (Affymetrix, Santa Clara, CA, USA). This dataset is originally based on a study which analysed mRNA levels from a variety of normal male C57Bl/6 mouse organs and tissues (Lattin et al., 2008). The normalised gene expression data for the mouse lung and liver for the candidate genes were downloaded and the values entered into GraphPad Prism v5 for the calculation of mean values. The fold difference in gene expression between the lung and liver was determined by dividing the mean lung values by the mean liver values. Table 7.2 lists the Mouse Genome Informatics (MGI) gene symbol and the fold difference for each gene. A cut-off value of 2-fold change is commonly used for microarray analysis and was also used here to discriminate expression of genes that were differentially expressed between the lung and liver. Genes which reached this 2-fold cut-off are highlighted in green in Table 7.2,

whereas genes which were expressed  $< 0.5$ -fold are highlighted in red and represent genes that are relatively over-expressed in the liver. Statistical differences could not be evaluated because reported mean gene expression values were determined from only two values per gene.

From this panel of genes, eight are relatively over-expressed in the mouse lung when compared to the mouse liver, and three are over-expressed in the liver. The results presented here do not suggest that transcription factor binding occurs; neither can subsequent activation or repression of transcriptional activity be assumed. However, these results demonstrate an approach that could be employed to further understand the reasons for differences in transgene activity between different organs. This transcriptome profile analysis suggests that TF proteins that may be required for the minimal 93 bp hC enhancer to mediate persistent transgene activity are not present in the mouse liver at the necessary levels since the minimal hC enhancer did not mediate transgene activity levels similar to the full-length hC enhancer (Figure 7.5).

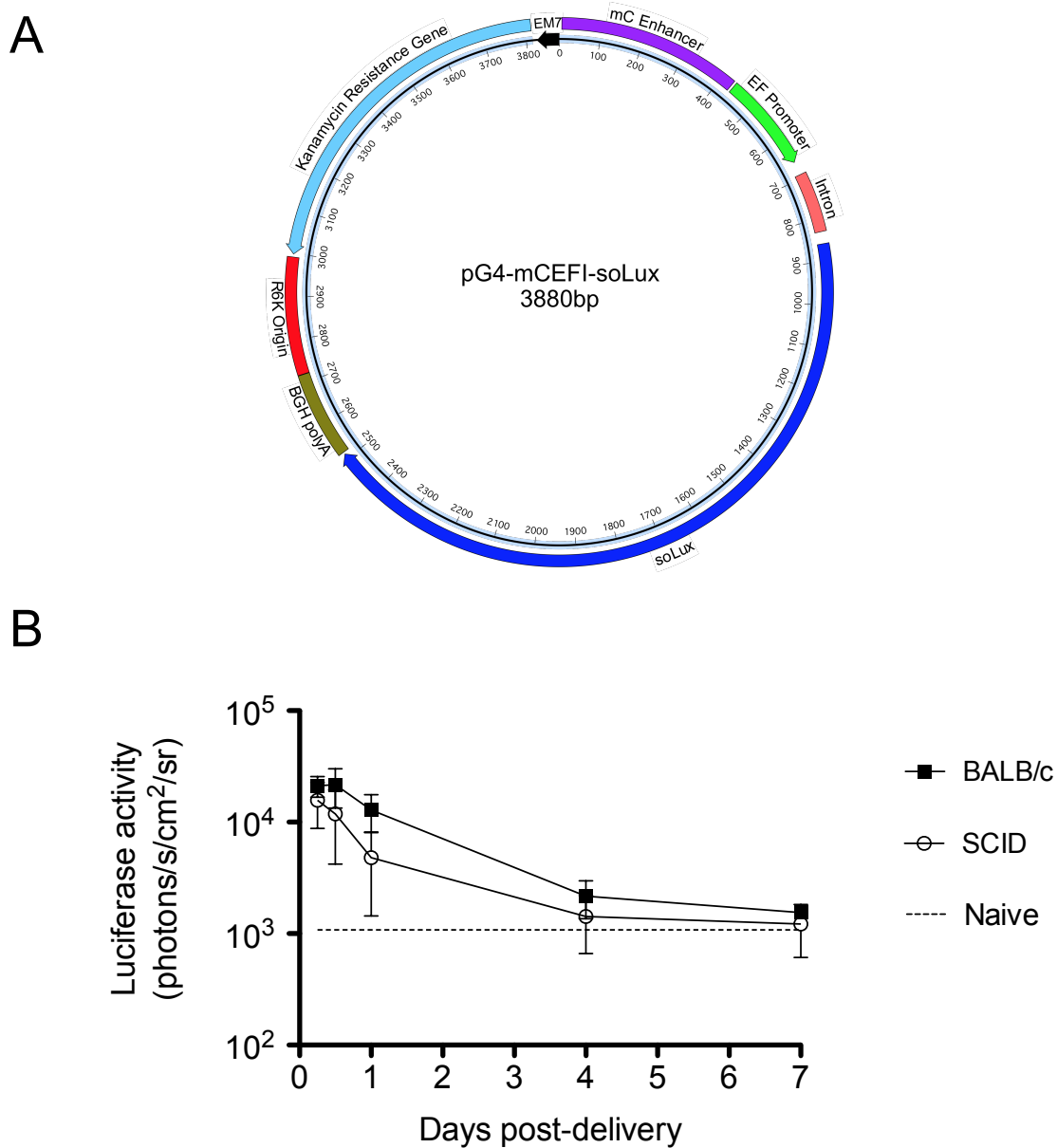
**Table 7.2. Gene expression data obtained via the BioGPS public repository.**

Gene name (MGI symbol)	Fold difference (lung versus liver)
Atf1	1.0
Bach1	1.5
Dbp	0.9
Esrra	0.9
Esrrb	0.8
Fosl1	1.0
Hivep1	3.1
Hsf1	1.7
Irx6	1.0
Jund	1.0
Nfat5	1.0
Nfia	0.3
Nfkb1	4.2
Nr1i3	0.004
Nr4a1	7.6
Phox2a	0.8
Rel	0.8
Rela	2.2
Rxrg	0.4
Satb1	18.2
Srf	4.5
Thra	6.6
Vdr	2.2
Zbtb3	1.1
Zfp423	1.0

Mean values of normalised data from mouse lung obtained from the BioGPS database were divided by the liver values to calculate the fold difference in gene expression levels. Values highlighted in green indicate over-expression in the lung (minimum cut-off of 2) whereas the values in red indicate over-expression in the liver (maximum cut-off of 0.5).

#### **7.2.4. Influence of the adaptive immune system on transient transgene activity from pG4-mCEFI-soLux**

The finding in Section 7.2.3.1 demonstrated that the adaptive immune system may play a role in transient luciferase activity after HTV injection to the liver. To investigate whether the adaptive immune system may also be playing a role in transient luciferase activity in the lung (Figure 4.2A), PEI polyplexes of pG4-mCEFI-soLux (Figure 7.6A) were aerosolised to the lungs of BALB/c and SCID mice (Section 2.5.6). As seen in Figure 7.6B, luciferase activity as measured by BLI (Section 2.6.3) in both mouse strains was transient and fell towards background levels by day 7. Peak transgene activity was observed at 12 h after aerosol dosing of pG4-mCEFI-soLux in the BALB/c mice and 6 h in the SCID mice. Levels of luciferase activity at day 7 were approximately 13 % and 30 % of day 1 values observed in BALB/c and SCID mice, respectively. Therefore, a mechanism other than an adaptive immune response is likely to be responsible for the transient luciferase activity observed after aerosol administration of pG4-mCEFI-soLux.



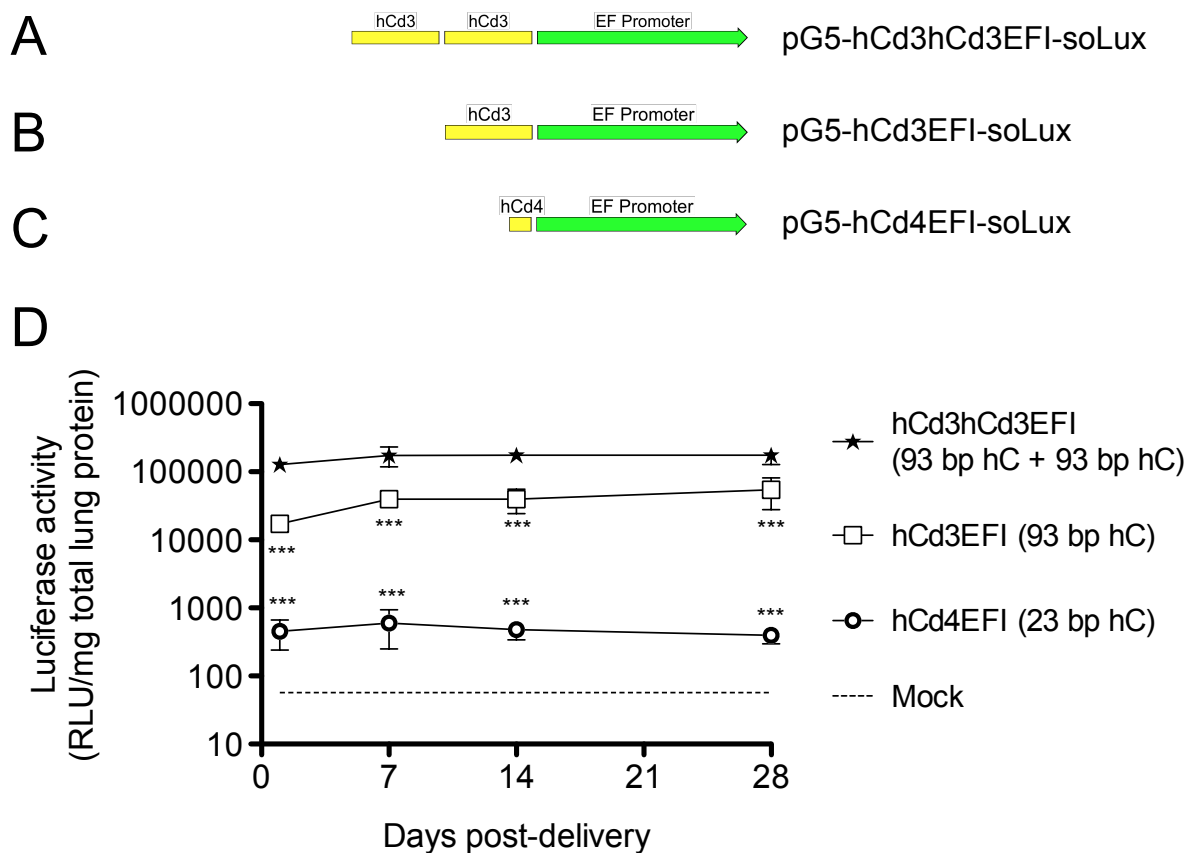
**Figure 7.6. Comparison of luciferase activity after aerosol dosing of pG4-mCEFI-soLux in BALB/c or SCID mice.**

Female mice (n = 6 per strain) were exposed to 10 mL treatment aerosols containing 2 mg of pG4-mCEFI-soLux (A) complexed with PEI (Section 2.5.6).

**B:** Lung firefly luciferase activity was measured at the time-points indicated using BLI (Section 2.6.3). Data are shown as mean  $\pm$  SD. Statistical differences between conditions were analysed by two-way ANOVA with Bonferroni's multiple comparison post-hoc tests; no statistical significance was observed.

### 7.2.5. Study utilising the minimal enhancer as a tandem dual enhancer

Given the short sequence of the minimal hC enhancer, it is conceivable that it could be used as a dual tandem enhancer to mediate transgene activity levels higher than what can be achieved by using it as a single enhancer. To evaluate this, a new plasmid was constructed which contains two copies of the minimal enhancer in tandem, operably linked with the EF promoter (Figure 7.7A). This construct was compared for transgene activity after aerosol delivery against the hCd3 enhancer (Figure 7.7B). A plasmid containing the hCd4 enhancer (Figure 7.7C) was administered as a negative control. Figure 7.7D shows that plasmid containing the dual tandem enhancer resulted in a modest (approximately seven-fold), but statistically significant ( $p < 0.001$ ), increase in luciferase activity compared with a single copy of the minimal enhancer 1 day after dosing. Furthermore, at the end of the experiment, luciferase activity mediated by the dual and single enhancer constructs was approximately 445-fold and 138-fold higher than the hCd4 enhancer, respectively ( $p < 0.001$ ). Thus, the use of two copies of the minimal enhancer in tandem can result in increased reporter activity after aerosol delivery to the mouse lung.



**Figure 7.7. Comparison of luciferase activity after aerosol administration of plasmids containing the (A) dual enhancer using the hCd3 sequence in tandem, (B) hCd3 enhancer and (C) hCd4 enhancer.**

Female BALB/c mice ( $n = 16$  per group) were exposed to 10 mL treatment aerosols containing 2 mg of the respective plasmid construct with enhancer variants (A, B, C) complexed with PEI (Section 2.5.6).

**D:** Firefly luciferase activity was analysed at the time-points indicated using standard luminometry (Section 2.6.1,  $n = 4$  per time-point). The enhancer length is listed in brackets. Data are shown as mean  $\pm$  SD. Significant differences to values obtained from hCd3hCd3EFI at specific time-points were determined using two-way ANOVA followed by Bonferroni's post-hoc tests; \*\*\*  $p < 0.001$ . For plasmid details, see Table 7.1 and Figure 7.1.

### 7.3. Discussion

Studies in previous chapters identified a minimal 93 bp derivative of the hC enhancer, which is both necessary and sufficient to mediate sustained, high-level reporter gene activity in the mouse lung after aerosol delivery when operably linked to the EF promoter. The studies in this chapter evaluated the relevance of the minimal enhancer in the context of other gene therapy applications using CpG-free plasmids.

#### 7.3.1. Use of the mFABP promoter

The minimal 93 bp hC enhancer was positioned next to another CpG-free promoter, the promoter for the murine fatty acid binding protein 4 (mFABP), identified from a screen of publicly available promoter databases (Pringle et al., 2012a). Pringle *et al.* showed that the full-length hC enhancer in conjunction with the mFABP promoter could mediate transgene expression similar to the hCEFI enhancer/promoter, after nasal instillation of GL67A lipoplexes at days 1 and 14 after delivery, and also after aerosol dosing of PEI polyplexes 28 days after delivery (Pringle et al., 2012a). Experiments described in this chapter using the minimal hC enhancer did not mimic the published results seen with the full-length hC enhancer (Figure 7.2D), possibly due to the fact that Pringle *et al.* quantified mRNA expression rather than luciferase activity. It is formally possible, though perhaps unlikely, that levels of mRNA and luciferase activity from the mFABP promoter do not correlate but this was not further pursued. Regardless, it is clear from these studies that the minimal enhancer does not reach the same levels of transcriptional activity as the full-length hC enhancer when operably linked to the mFABP promoter, further confirming the importance of promoter choice in these studies; the simple selection of a CpG-free promoter is not sufficient. This result also suggests that the specific interaction between the enhancer and promoter is crucial. Whereas the elongation factor 1 $\alpha$  gene is highly expressed in most mammalian tissues (Uetsuki et al., 1989), the mFABP gene is highly expressed in adipocytes, appendix, skin and placenta (reviewed in Chmurzynska, 2006). Therefore, the selection of a CpG-free

promoter from a 'housekeeping' gene may have been a more appropriate promoter with which to compare the effect of the minimal and full-length enhancers.

### 7.3.2. Transgene activity in the mouse liver

The liver is also an important organ in which to investigate gene transfer efficiencies for inherited diseases, such as alpha-1 antitrypsin deficiency (Alino et al., 2003) and phenylketonuria (Viecelli et al., 2014). Hydrodynamic tail vein (HTV) injection is a well-established procedure for achieving high numbers (> 20 %) of transfected liver cells (Liu et al., 1999; Zhang et al., 1999). The procedure is relatively simple to perform without the need for any special dosing equipment, is well-tolerated by mice and requires only a small amount of pDNA. Persistent transgene activity in the mouse liver with CpG-free plasmids has not previously been reported. SCID mice were used to establish a model in which to test transcriptional activity of the full-length hC enhancer compared with the minimal 93 bp hC enhancer (Figure 7.5). Quantification of the number of plasmid copies in the BALB/c and SCID mice indicated that the destruction of transfected cells might be a cause of transient luciferase activity after HTV injection (Figure 7.4).

One published study observed that transgene activity from plasmids containing the hCMV enhancer and CpG-free EF promoter persisted in immuno-deficient but not immuno-competent BALB/c mice, although the plasmids were not entirely free of CpGs (Magnusson et al., 2011). Interestingly, the plasmids used contained a CpG-free S/MAR element (Magnusson, 2010) (Section 1.3.1), indicating that the use of such elements does not automatically confer an advantage *in vivo*. The results observed by Magnusson *et al.* are also suggestive of an adaptive immune response to the luciferase protein, although other studies have demonstrated persistence of luciferase activity in the mouse liver even when CpG-rich plasmids are used (Wooddell et al., 2008). Indeed, Pringle *et al.* was able to demonstrate persistence in the liver of BALB/c mice using the UbC promoter, with stable levels of transgene activity until at least 35 days after HTV injection (Pringle et al., 2010),

although levels at day 1 after HTV injection were 70-fold lower than those observed with the hCEFI enhancer/promoter.

### **7.3.3. The role of the adaptive immune response in transient transgene activity**

The luciferase reporter gene is a sensitive tool for measuring transgene activity, but its effectiveness as an *in vivo* reporter can be affected by its immunogenicity such that recognition of the luciferase protein by the adaptive immune system may negatively impact the outcome of vector-mediated gene transfer. Indeed, murine firefly luciferase-specific CD8 T-cell epitopes have been identified in both C57BL/6 and BALB/c mouse strains (Limberis et al., 2009) the latter being the mouse strain used in the studies described in this chapter. Interestingly, the immune response to high and moderate levels of luciferase protein after HTV injection in immuno-competent mice has recently been investigated with evidence of a threshold effect for the induction of an adaptive immune response. Moderate levels of luciferase activity ( $< 10^9$  photons/sec/cm<sup>2</sup>/sr) at early time-points after HTV injection resulted in persistent transgene activity, whereas high levels of luciferase activity ( $> 10^9$  photons/sec/cm<sup>2</sup>/sr) at early time-points led to a sharp decrease soon after HTV injection (Podetz-Pedersen et al., 2014). This observation may explain the persistence of low level transgene activity after HTV injection of plasmids containing the UbC promoter, as discussed above (Pringle et al., 2010). The HTV studies described in this chapter used approximately 18 µg of pDNA per mouse; however, luciferase activity following HTV delivery was shown to peak after a 5 µg injection (Liu et al., 1999). Therefore, a study exploring the dose response after HTV injection of pDNA containing the hCEFI enhancer/promoter may identify a plasmid dose/level of luciferase activity that is tolerated by the immune system resulting in sustained transgene activity in the liver. Alternatively, the insertion of a micro-RNA-binding site such as miR142-3p, which suppresses transgene expression in cells of hematopoietic lineages such as antigen presenting cells (Brown et al., 2006), could be included in plasmid vector design. However, a study performed by Wolff *et al.* suggests that the use of the miR142-3p binding

site in non-viral vectors does not prevent the 'silencing' of luciferase activity in the liver after HTV injection (Wolff et al., 2009). Therefore, this approach was not further considered.

Luciferase activity mediated by pG4-mCEFI-soLux at 1 day after aerosol dosing or nasal instillation (Figure 4.2A) did not approach the threshold level of  $> 10^9$  photons/sec/cm<sup>2</sup>/sr, but it may be possible that a non-integrating vector as used in the studies described in this thesis could achieve higher transgene activity levels earlier than day 1. Therefore, in order to determine whether an anti-transgene response was partly responsible for the difference in transgene activity mediated by the hCEFI and mCEFI enhancer/promoter sequences in the lung (Figure 4.3B), pG4-mCEFI-soLux was delivered by aerosolisation to the lungs of BALB/c and SCID mice and luciferase activity was measured as early as 6 h after dosing (Figure 7.6B). This early time-point was used in order to determine whether very high levels of luciferase activity might have been responsible for the triggering of an immune response. However, peak bioluminescence was observed at a level of approximately  $2.2 \times 10^5$  photons/sec/cm<sup>2</sup>/sr, far below the threshold described by Podetz-Pedersen, *et al.* (Podetz-Pedersen et al., 2014). Furthermore, transient luciferase activity was observed in both immuno-competent and immuno-deficient mice, with transgene activity falling to background levels by day 7 indicating that the adaptive immune response may not be solely responsible for the precipitous fall in detectable luciferase activity. This result corroborates other studies that have demonstrated that the adaptive immune system is not solely responsible for transient transgene activity after delivery of non-viral vectors to the airways (Davies et al., 2012; Gill et al., 2001).

#### **7.3.4. Gene expression differences between mouse lung and liver**

To further understand the differences in reporter activity in the lung and the liver following delivery of plasmids containing the hC enhancer (Figure 7.5), relative gene expression differences were analysed (Table 7.2). Intriguingly, expression of *Satb1* (previously described in Section 5.3.2) in the lung is approximately 18-fold higher than in the liver. The

SATB1 TFBS is present only once in the hC enhancer and coincides with a single position within the minimal 93 bp hC enhancer (Figure 5.3). The results in Figure 7.5 suggest that the presence of this TFBS is not sufficient to mediate high-level activity in the mouse liver after HTV injection since the minimal 93 bp hC enhancer did not achieve similar levels as the full-length hC enhancer. However, the relative under-expression of this gene in the liver may indicate a further reason why plasmids containing the hCEFI enhancer/promoter did not mediate sustained transgene activity in the livers of immuno-competent mice.

### **7.3.5. Use of the minimal enhancer as a dual tandem enhancer**

The use of a dual enhancer has been evaluated, whereby the hC and mC enhancers were combined in an attempt to drive higher transgene activity, but did not confer an advantage over a single enhancer located immediately 5' to the EF promoter (Pringle et al., 2012b). In this chapter, a dual enhancer approach with the minimal 93 bp hC enhancer was explored in an attempt to exploit the shorter sequence. The inclusion of two copies of the minimal enhancer resulted in a modest seven-fold increase in transgene activity (Figure 7.7) when compared with the single enhancer. This may indicate that TF availability for the TFBSs in the dual tandem enhancer may not be a limiting factor to mediate high levels of transcriptional activity. It would be interesting to determine whether further copies of the minimal enhancer (as a triple or quadruple enhancer) could mediate even higher levels of transgene activity. Alternatively, reporter gene activity may remain similar or even decrease with more copies of the enhancer due to a number of issues, such as steric interference when using multiple identical repeats, or transcriptional interference (Section 5.3.3) due to over-represented binding sites, or limiting TF availability. This line of investigation is currently being explored as an extension to the studies performed in this thesis.

### **7.3.6. Conclusions**

These studies have demonstrated that the high-level, sustained transgene activity observed with the minimal 93 bp hC enhancer in the mouse lung is not necessarily transferrable to

other CpG-free promoters or to other organs. However, the short sequence can be exploited as a dual enhancer to mediate higher reporter gene activity than what can be achieved with a single enhancer. A potential follow-up would be to construct plasmids with multiple copies of the minimal enhancer to determine if transgene activity in the mouse lung can be further increased with possible benefit for both non-viral and viral vectors.

---

## Chapter 8: Discussion

The main cause of morbidity and mortality in patients with cystic fibrosis is due to lung destruction as a result of continuous cycles of infection and inflammation. Gene therapy-based clinical trials have focused on the amelioration of this chronic lung disease by the introduction of a functional copy of the *CFTR* gene. Over 20 clinical trials have been performed, using viral and non-viral approaches (reviewed in Griesenbach et al., 2014) both of which will probably require repeated vector administration; this has so far limited the use of viral vectors due to host immune responses against the viral proteins. Proof-of-concept for repeated dosing of non-viral vectors has been established for some time, and even in the clinic (Hyde et al., 2000), but in general only transient transgene expression was observed. Furthermore, a self-limiting flu-like response to lung delivery of the gene therapy formulation itself was also noted (Alton et al., 1999; Ruiz et al., 2001).

Efforts in extending the duration of transgene expression from non-viral vectors have focused on two main aspects: 1) reduction in the host inflammatory response to minimise the flu-like symptoms observed in patients, and 2) extending transgene expression to prolong the interval between doses. While the removal of CpGs reduced inflammation and conferred an advantage in terms of transgene activity, the development of a synthetic enhancer/promoter construct, termed hCEFI, was critical in mediating sustained and high levels of lung transgene activity after aerosol delivery of pDNA (Hyde et al., 2008). The combination of these two focused areas of research led to a pivotal Phase I/IIa clinical trial, led by the UKCFGTC, using a CpG-free plasmid containing the hCEFI enhancer/promoter, from which followed the first Phase IIb trial to assess the clinical benefit of a non-viral vector, in the lungs of CF patients (Alton et al., 2015). Given the progress in the field to date, hCEFI is the most successful enhancer/promoter to direct transgene activity levels in the lung and a broader understanding of the reasons for persistent transgene activity from this enhancer/promoter is

---

warranted. Such an understanding may contribute to the development of novel regulatory sequences for use in both non-viral and viral gene therapy applications.

### **8.1. Identification of a model for investigation of persistent transgene expression**

It was initially hoped that an airway cell culture model could be utilised to investigate the function of enhancer sequences from non-viral vectors to allow straightforward delivery and manipulation of transfected cells. However, studies in Chapter 3 showed that gene expression from plasmids containing the hC or mC enhancer sequences in cell lines (Figures 3.4 and 3.7) and in human primary airway cell cultures (Figure 3.10) did not mimic the expression profiles obtained in the mouse lung. Gene transfer studies in cell culture can benefit from high levels of transfection efficiency and gene expression, possibly due to the increased cell membrane surface area for cell uptake and increased nuclear access due to higher rates of cell division (reviewed in van Gaal et al., 2011). However, as shown in Chapter 3, none of the selected cell culture models were appropriate for studies aimed at elucidating the mechanisms that underpin extended duration of transgene expression.

Human ALI cultures can more closely recapitulate the *in vivo* environment (Figure 3.1, Table 3.1) but are still mostly used as evidence of gene transfer rather than a model for persistent *in vivo* transgene expression. For example, both the mouse nose and ALI cultures were used to determine persistence of transgene expression after intranasal administration of a SIV vector pseudotyped with fusion and hemagglutinin neuraminidase envelope proteins of the Sendai virus (F/HN-SIV) (Mitomo et al., 2010) (Section 1.2.6). Transgene activity *in vivo* (as assessed by the number of GFP-positive cells in histological sections) was stable for at least 50 days after transduction. However, luciferase activity in the ALI cultures was less stable over a similar time-frame, decreasing approximately seven-fold between days 10 and 26 after transduction. Further studies have demonstrated that a similar F/HN-SIV vector can persist in the mouse lung for at least 5 months (Pringle et al., 2014). In light of these results,

---

and the results presented in Section 3.2.6, it would be interesting to assess whether recapitulation of the respective *in vivo* results could be observed in ALI cultures of mouse as opposed to human bronchial epithelial cells. Although the same cell types are present in the human and mouse lungs, their distribution along the airways differs (Mercer et al., 1994). Therefore, it remains a possibility that the cells which are transfected/transduced in the mouse lung could differ from the cells that are present in the human ALI cultures. Indeed, studies using two different serotypes of rAAV vectors have demonstrated that there were different transduction profiles whether human or mouse ALI cultures were used, and that the human cultures could not recapitulate the *in vivo* mouse result (Liu et al., 2006). Therefore, while human ALI cultures may be useful to assess the general efficacy of gene delivery, these studies may indicate that they are not suitable as surrogates to predict mouse *in vivo* results. This may be due to the fact that despite the similarities to the human airway, ALI cultures still cannot recapitulate the complex *in vivo* airway anatomy (i.e. submucosal glands) or hormonal regulation, and they are not subject to time-varying air composition (Verkman, 2001) or may represent a more fundamental species difference.

In the absence of a suitable cell culture model, an *in vivo* mouse aerosol model was re-visited. Bioluminescence imaging (BLI) was established to monitor persistence of transgene activity in the lung. As discussed in Section 4.1, numerous studies have demonstrated the feasibility of BLI to assess *in vivo* transgene activity after viral and non-viral gene transfer to the mouse airways although the use of CpG-free plasmids had not previously been evaluated. Studies established that luciferase activity could be detected after administration of pG4-mCEFI-soLux and pG4-hCEFI-soLux with transgene activity profiles similar to those obtained with standard luminometry (Figure 4.3). The limited dynamic range of this assay meant that BLI was useful when differentiating between high- and low-expressing plasmids (Figure 4.3, Figure 4.8), but was not appropriate when comparing transgene activity between multiple low-expressing plasmids (i.e. mC enhancer variants, Section 4.2.6). Methods to increase BLI assay sensitivity in the mouse lung, such as

---

increasing the amount of D-Luciferin, may lead to an improvement in the dynamic range of the assay which would allow it to be used more readily in such non-viral airway delivery studies.

Using the mouse aerosol model, the hC enhancer region of the persistent hCEFI promoter was investigated. The results showed that different variants of the hC enhancer could substitute for the full-length enhancer present in the hCEFI promoter, and this result was similar regardless of whether BLI or standard luminometry were used (Figure 4.8). Further studies demonstrated that the 3' 93 bp of the enhancer is necessary and sufficient for high-level and persistent luciferase activity (Chapter 6); however, these effects were evaluated in the context of the enhancer encompassing the 70 bp ROI (Figure 6.1E). It is possible that different effects would have been observed if these deletions were made in the context of the minimal 93 bp hC enhancer. In the studies performed, the results suggest that the minimal 93 bp hC enhancer contains the required sequences for active transcriptional regulation. Although the 93 bp sequence represents a small portion of the full-length enhancer (~ 30 %), this is not necessarily unique amongst enhancers. For example, a 72 bp region of the SV40 viral enhancer (nominally 245 bp long) has been shown to be necessary and sufficient for mediating transgene activity after intramuscular injection of naked pDNA containing the hCMV promoter driving a SEAP transgene (Li et al., 2001). However, up until this point, the minimal required sequence of the synthetic hC enhancer had not been investigated.

## **8.2. The use of *in silico* methods identified potentially interesting TFBSs**

Following on from the identification of the minimal 93 bp hC enhancer, *in silico* methods were utilised to understand whether there were key differences in the number and type of TFBSs between the enhancers tested. Five TFBSs were identified which are present only once in the entirety of the hC enhancer, the presence of which coincided with the minimal 93 bp hC enhancer sequence (Figure 5.3); these binding sites are completely absent in the mC

enhancer. The use of emerging techniques, such as ATAC-seq (assay for transposase-accessible chromatin using sequencing), may help identify open chromatin sites (Buenrostro et al., 2013) on the pDNA and therefore, may lead to an understanding of the role of particular TFBSs on mediating transcriptional activity in the lung. However, due to the technical challenge of isolating transfected cells in the mouse lung after aerosol delivery, and also a lack of understanding of which individual plasmid molecule in each cell is contributing to transcriptional activity, the use of *in silico* methods to predict putatively important TFBSs may instead prove to be useful at this time to formulate hypotheses for the function and role of these sequences. While the use of such TFBS data analysis tools may not give direct evidence of binding or active roles in transcription, functional approaches could be later employed using the information gathered from these datasets. For example, the SATB1.01 site which was identified as the special AT-rich sequence-binding protein 1 (Figure 5.3) represents an interesting candidate TFBS (Section 5.3.2) that could be added to the mC enhancer (or variants thereof) to evaluate its effect on transcriptional activity in the mouse lung. However, each enhancer must be functionally evaluated because the context of each plasmid molecule will differ and outcome predictions cannot be made on the sequence alone.

The results presented in this thesis have largely been assessed by comparison of luciferase protein activity, mediated by the different constructs; in this context, luciferase activity has been used as a surrogate marker for transcriptional activity. The *in silico* analyses have attempted to understand the role of TFBSs in mediating or repressing transcription. The binding of TF proteins to their binding sites can modulate transcription by varying the amount of general transcription factors that are recruited to a promoter (Figure 1.5). Also, co-activators and co-repressors can bind to the bound TFs, which can further modulate activity accordingly (reviewed in Shlyueva et al., 2014). It is also possible that other interactions with TFBSs (but not specifically nuclear TF binding) could be involved. For example, TFs can modify the structure of chromatin by binding to nucleosomal DNA and subsequently recruiting additional factors to form an 'open' structure of chromatin

(euchromatin; Section 1.4.1) or by competing with nucleosomes for DNA binding, thereby displacing the nucleosomes and allowing other TF proteins to bind (Bell et al., 2011). Another explanation for the decrease in the amount of measurable luciferase protein may be a decrease in the number of plasmid molecules reaching the nucleus, possibly via the removal of TFBSs which could otherwise bind TFs acting as nuclear localisation signals (Section 1.4.4). These possibilities were not directly addressed in the studies performed here but represent an area for further investigation.

### **8.3. The hC enhancer contains very few overrepresented binding sites compared with the mC enhancer**

While the hCEFI enhancer/promoter mediates sustained, high-level transgene activity in the mouse lung, the replacement of the hC enhancer with the mC enhancer resulted in transient activity (Hyde et al., 2008). Understanding the reasons for this difference may help explain the success of the hCEFI enhancer/promoter. The use of the 'Overrepresented TFBS' program in the Genomatix software suite (Section 5.2.2.3) highlighted a key difference in the number of shared overrepresented TFBSs between the hC and mC enhancers. As shown in Table 5.11, the number of over-represented TFBSs in the hC enhancer and EF promoter is very low (one or less). However, as shown in Table 5.12, the number of shared, overrepresented TFBSs in the mC enhancer and EF promoter is much greater, ranging from nine in the full-length mC enhancer to two in mCd5. This result may provide clues as to why transgene activity appears to stabilise, and even increase, as portions of the mC enhancer are successively removed (Figure 4.9). One possibility is that transcriptional interference (TI) occurs when the mC and EF sequences are physically linked (Section 5.3.3). However, one cannot predict TI solely on the basis of over-represented binding sites. For example, the mChCd3 dual enhancer mediated high-level transgene activity (Figure 6.2) despite the fact that it contains eight shared overrepresented TFBSs with the EF promoter (data not shown), which is equivalent to the mC, mCd1 and mCd2 enhancers (Table 5.12). However, since the mC and EF sequences in mChCd3 are not linked (Figure 6.1B), the context in which to

evaluate the effect of overrepresented TFBSs on transcriptional activity may not be suitable. An alternative explanation may be that the minimal hCd3 enhancer may also play an insulator role within the context of this dual enhancer. For example, if the mC enhancer plays a role in the formation of heterochromatin, the hCd3 enhancer may block the spreading of this repressive chromatin onto the EF promoter. Insulator elements in non-viral vectors have previously been shown to partially block the effects of transgene silencing after hydrodynamic tail vein injection to the liver (Chen et al., 2008). These studies indicate that results arising from investigations into transcriptional activity must be evaluated in the context of each plasmid molecule and that the sequence alone remains insufficient to predict *in vivo* outcome.

Recent studies have made use of publicly available resources to aid design of novel regulatory sequences (Chuah et al., 2014; Nair et al., 2014; Rincon et al., 2015; Viecelli et al., 2014) using genome-wide *in silico* identification of highly expressing and highly-specific genes for specific tissues. The promoters for these genes were further screened for clusters of evolutionarily conserved TFBSs, and placed before tissue-specific promoters. This led to the identification of efficient enhancers for use in gene therapy vectors for the liver (Chuah et al., 2014; Nair et al., 2014; Viecelli et al., 2014) and the heart (Rincon et al., 2015). Interestingly, this approach, based on bioinformatics, favours the overrepresentation of given TFBSs and the co-occurrence of such sites (Chuah et al., 2014) to yield strong transcriptional activity. The identification of overrepresented binding sites may suggest transcriptional interference between enhancers and promoters, which could result in transient transgene activity; therefore, one could take the view that overrepresented binding sites should be avoided in the design of regulatory sequences (Section 5.2.2.3). However, as described above, the presence of overrepresented binding sites does not automatically lead to reduced transgene activity. Once again, these results indicate the importance of context in vector design, and that an empirical approach should be used to investigate the elements that are important in conferring sustained transgene activity.

#### 8.4. The minimal hC enhancer did not mediate high-level activity in the liver

A model for persistent transgene activity from the hCEFI enhancer/promoter in the mouse liver after HTV injection was established by using SCID mice (Figure 7.3B). In contrast to the lung, the use of the minimal 93 bp hC enhancer in the liver did not mediate similar levels to the hC enhancer (Figure 7.5C). The differences in transgene activity observed between the mouse lung and liver when using the minimal 93 bp hC enhancer further confirms that a variety of factors must be considered when designing non-viral vectors, including target organ, route of administration, strength and interaction of enhancers and promoters, and the availability of requisite transcription factor proteins. A 'one size fits all' approach to plasmid vector design is unlikely to be optimal, and the results obtained from the liver study may not imply a reason for persistence of hC in the mouse lung. Therefore, although HTV delivery to the liver results in a greater number of transfected cells (Section 7.3.2) than aerosol delivery (Section 4.3.1), these studies indicate that it may not be correct to extrapolate results from one organ or tissue to another, and that the use of a 'surrogate' organ or tissue is not always suitable.

#### 8.5. Implications

The identification of the minimal 93 bp hC enhancer has implications for viral and non-viral gene transfer applications. For example, the widespread use of rAAV in gene therapy has been hampered by the small packaging capability of the vector (approximately 5 kb, Section 1.2.5), especially when considering the large size of the *CFTR* cDNA (4.443 kb) (Flotte and Carter, 1997). However, the combination of the minimal hCd3 enhancer (Figure 4.8) and the use of a *CFTR* 'minigene' (Ostedgaard et al., 2002) may circumvent such issues. Additionally, it may be possible to truncate the EF promoter (224 bp) to fit within the space constraints of the rAAV vector. Interestingly, short, synthetic enhancers and promoters have recently been evaluated in the context of driving expression of a *CFTR* 'minigene' (Yan et al., 2015b). In this study, a synthetic 100 bp enhancer, termed F5 (Schlabach et al., 2010), was

---

linked to a synthetic 83 bp promoter (Zhang et al., 2004) and significantly improved the transcriptional activity of the CFTR transgene in the ferret lung and trachea *in vivo* (Yan et al., 2015b). The F5 enhancer was identified via a synthetic biology approach evaluating 10-mer DNA repeats assembled as 100-mer sequences. Therefore, the field is acknowledging the importance of the identification of short, strong enhancer/promoter constructs. Continued focus on such issues is likely to have a future impact on the development of new viral and non-viral vectors.

The investment in time and money required for pre-clinical safety studies, manufacture of a clinical-grade plasmid preparation, and clinical trial means that the selection of the appropriate vector is crucial. The dual 93 bp hC enhancer (Figure 7.7A) may lead to an improvement in transcriptional activity over the single minimal 93 bp hC enhancer (Figure 7.7B). However, for CF, the concept of ‘more is better’ in terms of CFTR gene expression in CF patients may not be appropriate. Indeed, activation of CFTR-specific T cells was observed in mice deficient in CFTR, but not in wild-type mice (Limberis et al., 2010) suggesting that patients who harbour the *CFTR* mutation where complete loss of protein is observed may be more likely to elicit CFTR-specific T-cell responses. Another consideration is that effects observed with the relatively small luciferase reporter gene might not extend to the larger *CFTR* cDNA. Despite these limitations, new non-viral vectors containing shortened regulatory sequences may be viewed as a favourable change by the regulatory bodies, since the use of specific elements in the vector should be justified (www.ema.europa.eu, 2001). Furthermore, the results of the multi-dose clinical trial of pG4-hCEFI-soCFTR2/GL67A suggest that more frequent dosing may be required to increase the therapeutic benefit of the non-viral formulation and the reduce the variability of overall lung function when compared with placebo was which observed in the patients (Alton et al., 2015). Therefore, the use of an improved enhancer/promoter sequence may be beneficial in terms of reducing dosing frequency, which may contribute to a less expensive treatment and improve patients’ quality

of life. Altogether, such improved enhancer/promoter sequences should be considered in the development on novel non-viral vectors.

## **8.6. Conclusion**

The field of gene therapy has only relatively recently become more aware of the requirement for choosing the appropriate enhancers and promoters to direct the desired level and persistence of expression for the target disease. The development of the hCEFI enhancer/promoter has allowed for sustained transgene activity after airway gene delivery in pre-clinical models. The persistence of transgene activity, coupled with minimal host inflammation that was achieved in those studies, enabled the work led by the UKCFGTC to deliver repeated doses of a non-viral vector to the lungs of CF patients. While such repeated doses are considered to be necessary due to the terminally differentiated nature of the lung, maximising the interval between therapeutic administrations is desirable to increase patient safety and comfort, and to decrease product costs. Therefore, an understanding of the reasons for sustained transgene expression is crucial for the clinical application of gene therapies. The studies presented here have identified sequences within enhancers that are important for directing sustained transgene expression in the mouse lung. The identification of critical binding sites within these sequences could lead to further understanding of the necessary elements and might improve future vector design for airway gene delivery.

---

## Chapter 9: References

- Aitken, M., Moss, R., Waltz, D., Dovey, M., Tonelli, M., McNamara, S., Gibson, R., Ramsey, B., Carter, B., and Reynolds, T. (2001). A phase I study of aerosolized administration of tgAAVCF to cystic fibrosis subjects with mild lung disease. *Hum Gene Ther* *12*, 1907-1916.
- Al-Dosari, M.S., and Gao, X. (2009). Nonviral gene delivery: principle, limitations, and recent progress. *AAPS J* *11*, 671-681.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). *Chromosomal DNA and its Packaging in the Chromatin Fiber* (Garland Science).
- Alino, S.F., Crespo, A., and Dasi, F. (2003). Long-term therapeutic levels of human alpha-1 antitrypsin in plasma after hydrodynamic injection of nonviral DNA. *Gene Ther* *10*, 1672-1679.
- Alton, E.W., Armstrong, D.K., Ashby, D., Bayfield, K.J., Bilton, D., Bloomfield, E.V., Boyd, A.C., Brand, J., Buchan, R., and Calcedo, R. (2015). Repeated nebulisation of non-viral CFTR gene therapy in patients with cystic fibrosis: a randomised, double-blind, placebo-controlled, phase 2b trial. *The Lancet Respiratory Medicine* -*Not available* -, *ahead of print*. doi:10.1016/S2213-2600(15)00245-3.
- Alton, E.W., Baker, A., Baker, E., Boyd, A.C., Cheng, S.H., Coles, R.L., Collie, D.D., Davidson, H., Davies, J.C., Gill, D.R., *et al.* (2013a). The safety profile of a cationic lipid-mediated cystic fibrosis gene transfer agent following repeated monthly aerosol administration to sheep. *Biomaterials* *34*, 10267-10277.
- Alton, E.W., Boyd, A.C., Cheng, S.H., Cunningham, S., Davies, J.C., Gill, D.R., Griesenbach, U., Higgins, T., Hyde, S.C., Innes, J.A., *et al.* (2013b). A randomised, double-blind, placebo-controlled phase IIB clinical trial of repeated application of gene therapy in patients with cystic fibrosis. *Thorax* *68*, 1075-1077.
- Alton, E.W., Boyd, A.C., Cheng, S.H., Davies, J.C., Davies, L.A., Dayan, A., Gill, D.R., Griesenbach, U., Higgins, T., Hyde, S.C., *et al.* (2014). Toxicology study assessing efficacy and safety of repeated administration of lipid/DNA complexes to mouse lung. *Gene Ther* *21*, 89-95.
- Alton, E.W., Stern, M., Farley, R., Jaffe, A., Chadwick, S.L., Phillips, J., Davies, J., Smith, S.N., Browning, J., Davies, M.G., *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.
- Andries, O., De Filette, M., De Smedt, S.C., Demeester, J., Van Poucke, M., Peelman, L., and Sanders, N.N. (2013). Innate immune response and programmed cell death following carrier-mediated delivery of unmodified mRNA to respiratory cells. *J Control Release* *167*, 157-166.
- Aneja, M.K., Geiger, J.P., Himmel, A., and Rudolph, C. (2009). Targeted gene delivery to the lung. *Expert Opin Drug Deliv* *6*, 567-583.
- Aneja, M.K., Imker, R., and Rudolph, C. (2007). Phage phiC31 integrase-mediated genomic integration and long-term gene expression in the lung after nonviral gene delivery. *The Journal of Gene Medicine* *9*, 967-975.

- Antequera, F., and Bird, A. (1993). CpG Islands. In DNA Methylation, J.-P. Jost, and H.-P. Saluz, eds. (Birkhauser Basel), pp. 169-185.
- Apolonia, L., Waddington, S.N., Fernandes, C., Ward, N.J., Bouma, G., Blundell, M.P., Thrasher, A.J., Collins, M.K., and Philpott, N.J. (2007). Stable gene transfer to muscle using non-integrating lentiviral vectors. *Molecular Therapy* 15, 1947-1954.
- Argyros, O., Wong, S.P., Fedonidis, C., Tolmachov, O., Waddington, S.N., Howe, S.J., Niceta, M., Coutelle, C., and Harbottle, R.P. (2011). Development of S/MAR minicircles for enhanced and persistent transgene expression in the mouse liver. *Journal of molecular medicine* 89, 515-529.
- Argyros, O., Wong, S.P., Niceta, M., Waddington, S.N., Howe, S.J., Coutelle, C., Miller, A.D., and Harbottle, R.P. (2008). Persistent episomal transgene expression in liver following delivery of a scaffold/matrix attachment region containing non-viral vector. *Gene Ther* 15, 1593-1605.
- Badding, M.A., Vaughan, E.E., and Dean, D.A. (2012). Transcription factor plasmid binding modulates microtubule interactions and intracellular trafficking during gene transfer. *Gene Ther* 19, 338-346.
- Baeuerle, P.A. (1998). I $\kappa$ B-NF- $\kappa$ B structures: at the interface of inflammation control. *Cell* 95, 729-731.
- Baguma-Nibasheka, M., Angka, H., Inanlou, M., and Kablar, B. (2007). Microarray analysis of Myf5<sup>-/-</sup>: MyoD<sup>-/-</sup> hypoplastic mouse lungs reveals a profile of genes involved in pneumocyte differentiation.
- Bao, Y., Gu, D., Feng, W., Sun, X., Wang, X., Zhang, X., Shi, Q., Cui, G., Yu, H., Tang, C., *et al.* (2014). COUP-TFII regulates metastasis of colorectal adenocarcinoma cells by modulating Snail1. *Br J Cancer* 111, 933-943.
- Bazzani, R.P. (2011). An Investigation into the Inflammatory Response to CG Dinucleotides in the Lung. D.Phil. Thesis. In Nuffield Division of Clinical Laboratory Sciences (Oxford, University of Oxford), p. 196.
- Bazzani, R.P., Cai, Y., Hebel, H.L., Hyde, S.C., and Gill, D.R. (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.
- Bell, O., Tiwari, V.K., Thoma, N.H., and Schubeler, D. (2011). Determinants and dynamics of genome accessibility. *Nat Rev Genet* 12, 554-564.
- Belur, L.R., Frandsen, J.L., Dupuy, A.J., Ingbar, D.H., Largaespada, D.A., Hackett, P.B., and Scott McIvor, R. (2003). Gene Insertion and Long-Term Expression in Lung Mediated by the Sleeping Beauty Transposon System. *Mol Ther* 8, 501-507.
- BeruBe, K., Aufderheide, M., Breheny, D., Clothier, R., Combes, R., Duffin, R., Forbes, B., Gaca, M., Gray, A., Hall, I., *et al.* (2009). In vitro models of inhalation toxicity and disease. The report of a FRAME workshop. *Altern Lab Anim* 37, 89-141.
- Beyerle, A., Kissel, T., and Stoeger, T. (2011). Toxicity of Polymeric-Based Non-Viral Vector Systems for Pulmonary siRNA Application. In *Non-Viral Gene Therapy*, P.X. Yuan, ed.

- Blackwood, E.M., and Kadonaga, J.T. (1998). Going the Distance: A Current View of Enhancer Action. *Science* 281, 60-63.
- Bode, J., Benham, C., Knopp, A., and Mielke, C. (2000). Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements). *Critical reviews in eukaryotic gene expression* 10, 73-90.
- Bode, J., Fetzer, C.P., Nehlsen, K., Scinteie, M., Hinrichs, B., Baiker, A., Piechaczek, C., Benham, C., and Lipps, H.J. (2001). The hitchhiking principle: optimizing episomal vectors for the use in gene therapy and biotechnology. *Gene Ther. Mol. Biol* 6, 33-46.
- Bosma, M.J., and Carroll, A.M. (1991). The SCID mouse mutant: definition, characterization, and potential uses. *Annual review of immunology* 9, 323-350.
- Boussif, O., Lezoualc'h, F., Zanta, M.A., Mergny, M.D., Scherman, D., Demeneix, B., and Behr, J.P. (1995). A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. *Proc Natl Acad Sci U S A* 92, 7297-7301.
- Boyle, M.P., Bell, S.C., Konstan, M.W., McColley, S.A., Rowe, S.M., Rietschel, E., Huang, X., Waltz, D., Patel, N.R., and Rodman, D. (2014). A CFTR corrector (lumacaftor) and a CFTR potentiator (ivacaftor) for treatment of patients with cystic fibrosis who have a phe508del CFTR mutation: a phase 2 randomised controlled trial. *The Lancet Respiratory Medicine* 2, 527-538.
- Brown, B.D., Venneri, M.A., Zingale, A., Sergi, L.S., and Naldini, L. (2006). Endogenous microRNA regulation suppresses transgene expression in hematopoietic lineages and enables stable gene transfer. *Nat Med* 12, 585-591.
- Brunner, S., Fürtbauer, E., Sauer, T., Kursa, M., and Wagner, E. (2002). Overcoming the nuclear barrier: cell cycle independent nonviral gene transfer with linear polyethylenimine or electroporation. *Molecular Therapy* 5, 80-86.
- Buckley, S.M.K., Howe, S.J., Rahim, A.A., Buning, H., McIntosh, J., Wong, S.P., Baker, A.H., Nathwani, A., Thrasher, A.J., Coutelle, C., *et al.* (2008). Luciferin Detection After Intranasal Vector Delivery Is Improved by Intranasal Rather Than Intraperitoneal Luciferin Administration. *Hum Gene Ther* 19, 1050-1056.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* 10, 1213-+.
- Byers, D.E., Alexander-Brett, J., Patel, A.C., Agapov, E., Dang-Vu, G., Jin, X., Wu, K., You, Y., Alevy, Y., Girard, J.-P., *et al.* (2013). Long-term IL-33-producing epithelial progenitor cells in chronic obstructive lung disease. *The Journal of Clinical Investigation* 123, 3967-3982.
- 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.
- Chadwick, S.L., Kingston, H.D., Stern, M., Cook, R.M., O'Connor, B.J., Lukasson, M., Balfour, R.P., Rosenberg, M., Cheng, S.H., Smith, A.E., *et al.* (1997). Safety of a single aerosol administration of escalating doses of the cationic lipid GL-67/DOPE/DMPE-PEG5000 formulation to the lungs of normal volunteers. *Gene Ther* 4, 937-942.

- Chalberg, T.W., Portlock, J.L., Olivares, E.C., Thyagarajan, B., Kirby, P.J., Hillman, R.T., Hoelters, J., and Calos, M.P. (2006). Integration Specificity of Phage phiC31 Integrase in the Human Genome. *Journal of Molecular Biology* 357, 28-48.
- Chan, S.Y., Speck, R.F., Ma, M.C., and Goldsmith, M.A. (2000). Distinct Mechanisms of Entry by Envelope Glycoproteins of Marburg and Ebola (Zaire) Viruses. *Journal of Virology* 74, 4933-4937.
- Chandler, R.J., Ashok, A.A., Varshney, G.K., Lafave, M.C., Wu, W., Hoffmann, V., Elkahloun, A.G., Burgess, S.M., and Venditti, C.P. (2014). Genomic Analyses Define a Common Integration Site in Hepatocellular Carcinoma Caused By Therapeutic AAV Gene Delivery: A Mouse-Specific Phenomenon or New Safety Challenge for Human Gene Therapy? *Molecular Therapy* 22, S92-S92.
- Chaudhry, A.Z., Lyons, G.E., and Gronostajski, R.M. (1997). Expression patterns of the four nuclear factor I genes during mouse embryogenesis indicate a potential role in development. *Developmental Dynamics* 208, 313-325.
- Chen, Z.Y., He, C.Y., Ehrhardt, A., and Kay, M.A. (2003). Minicircle DNA vectors devoid of bacterial DNA result in persistent and high-level transgene expression in vivo. *Mol Ther* 8, 495-500.
- Chen, Z.Y., Riu, E., He, C.Y., Xu, H., and Kay, M.A. (2008). Silencing of episomal transgene expression in liver by plasmid bacterial backbone DNA is independent of CpG methylation. *Mol Ther* 16, 548-556.
- Cheng, S.H., and Scheule, R.K. (1998). Airway delivery of cationic lipid: DNA complexes for cystic fibrosis. *Advanced drug delivery reviews* 30, 173-184.
- Chmurzynska, A. (2006). The multigene family of fatty acid-binding proteins (FABPs): Function, structure and polymorphism. *J Appl Genet* 47, 39-48.
- Christman, J.W., Sadikot, R.T., and Blackwell, T.S. (2000). The role of nuclear factor-kappa b in pulmonary diseases. *Chest* 117, 1482-1487.
- Chu, Q., Tousignant, J.D., Fang, S., Jiang, C., Chen, L.H., Cheng, S.H., Scheule, R.K., and Eastman, S.J. (1999). Binding and uptake of cationic lipid:pDNA complexes by polarized airway epithelial cells. *Hum Gene Ther* 10, 25-36.
- Chuah, M.K., Petrus, I., De Bleser, P., Le Guiner, C., Gernoux, G., Adjali, O., Nair, N., Willems, J., Evens, H., Rincon, M.Y., *et al.* (2014). Liver-specific transcriptional modules identified by genome-wide in silico analysis enable efficient gene therapy in mice and non-human primates. *Mol Ther* 22, 1605-1613.
- Clarke, L.L., Grubb, B.R., Yankaskas, J.R., Cotton, C.U., McKenzie, A., and Boucher, R.C. (1994). Relationship of a non-cystic fibrosis transmembrane conductance regulator-mediated chloride conductance to organ-level disease in Cfr(-/-) mice. *Proceedings of the National Academy of Sciences* 91, 479-483.
- Cohn, L.A., and Adler, K.B. (1991). In vitro studies of mechanisms of lung injury in the rodent. *Toxicol Pathol* 19, 419-427.
- Constant, S., Huang, S., Derouette, J.-P., and Wiszniewski, L. (2008). MucilAir: a novel in vitro human 3D airway epithelium model for assessing the potential hazard of nanoparticles and chemical compounds. *Toxicology Letters* 180, S233.

- Contag, C.H., and Bachmann, M.H. (2002). Advances in in vivo bioluminescence imaging of gene expression. *Annu Rev Biomed Eng* 4, 235-260.
- Contag, C.H., Spilman, S.D., Contag, P.R., Oshiro, M., Eames, B., Dennery, P., Stevenson, D.K., and Benaron, D.A. (1997). Visualizing gene expression in living mammals using a bioluminescent reporter. *Photochem Photobiol* 66, 523-531.
- Copreni, E., Castellani, S., Palmieri, L., Penzo, M., and Conese, M. (2008). Involvement of glycosaminoglycans in vesicular stomatitis virus G glycoprotein pseudotyped lentiviral vector-mediated gene transfer into airway epithelial cells. *The Journal of Gene Medicine* 10, 1294-1302.
- Cramer, F., Christensen, C., Poulsen, T., Badding, M., Dean, D., and Poulsen, H. (2012). Insertion of a nuclear factor kappa B DNA nuclear-targeting sequence potentiates suicide gene therapy efficacy in lung cancer cell lines. *Cancer Gene Therapy*.
- Crook, K., McLachlan, G., Stevenson, B.J., and Porteous, D.J. (1996). Plasmid DNA molecules complexed with cationic liposomes are protected from degradation by nucleases and shearing by aerosolisation. *Gene Ther* 3, 834-839.
- Curtin, J.A., Dane, A.P., Swanson, A., Alexander, I.E., and Ginn, S.L. (2008). Bidirectional promoter interference between two widely used internal heterologous promoters in a late-generation lentiviral construct. *Gene Ther* 15, 384-390.
- Dames, P., Gleich, B., Flemmer, A., Hajek, K., Seidl, N., Wiekhorst, F., Eberbeck, D., Bittmann, I., Bergemann, C., Weyh, T., *et al.* (2007). Targeted delivery of magnetic aerosol droplets to the lung. *Nat Nanotechnol* 2, 495-499.
- Darquet, A.M., Cameron, B., Wils, P., Scherman, D., and Crouzet, J. (1997). A new DNA vehicle for nonviral gene delivery: supercoiled minicircle. *Gene Ther.* 4, 1341-1349.
- Davies, J.C., Davies, G., Gill, D.R., Hyde, S.C., Boyd, A.C., Innes, A.J., Porteous, D.J., Cheng, S.H., Scheule, R.K., Higgins, T., *et al.* (2011). Safety & expression of a single dose of lipid-mediated CFTR gene therapy to the upper & lower airways of patients with CF. *Pediatric Pulmonology* 46, 281.
- Davies, J.C., Geddes, D.M., and Alton, E.W. (2001). Prospects for gene therapy in lung disease. *Curr Opin Pharmacol* 1, 272-277.
- Davies, L.A., Hannavy, K., Davies, N., Pirrie, A., Coffee, R.A., Hyde, S.C., and Gill, D.R. (2005). Electrohydrodynamic comminution: a novel technique for the aerosolisation of plasmid DNA. *Pharm Res* 22, 1294-1304.
- Davies, L.A., Hyde, S.C., Nunez-Alonso, G., Bazzani, R.P., Harding-Smith, R., Pringle, I.A., Lawton, A.E., Abdullah, S., Roberts, T.C., 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.
- Davies, L.A., McLachlan, G., Sumner-Jones, S.G., 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.
- Davies, L.A., Seguela, C., Varathalingam, A., Cheng, S.H., Hyde, S.C., and Gill, D.R. (2007). Identification of transfected cell types following non-viral gene transfer to the murine lung. *J Gene Med* 9, 184-196.

- Davis, P.B., Yasothan, U., and Kirkpatrick, P. (2012). Ivacaftor. *Nature Reviews Drug Discovery* 11, 349-350.
- Dean, D.A., Barravecchia, M., Danziger, B., and Lin, X. (2011). Use of Electroporation for Efficacious Gene Delivery to the Lungs. *Ecs Transactions* 35, 167-177.
- Dean, D.A., Machado-Aranda, D., Blair-Parks, K., Yeldandi, A.V., and Young, J.L. (2003). Electroporation as a method for high-level nonviral gene transfer to the lung. *Gene Ther.* 10, 1608-1615.
- Desai, T.J., Brownfield, D.G., and Krasnow, M.A. (2014). Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature* 507, 190-194.
- Dong, J.-Y., Fan, P.-D., and Frizzell, R.A. (1996). Quantitative analysis of the packaging capacity of recombinant adeno-associated virus. *Hum Gene Ther* 7, 2101-2112.
- Dorin, J.R., Farley, R., Webb, S., Smith, S.N., Farini, E., Delaney, S.J., Wainwright, B.J., Alton, E.W.F.W., and Porteous, D.J. (1996). A demonstration using mouse models that successful gene therapy for cystic fibrosis requires only partial gene correction. *Gene Ther.* 3, 797-801.
- Drumm, M.L., Pope, H.A., Cliff, W.H., Rommens, J.M., Marvin, S.A., Tsui, L.-C., Collins, F.S., Frizzell, R.A., and Wilson, J.M. (1990). Correction of the cystic fibrosis defect in vitro by retrovirus-mediated gene transfer. *Cell* 62, 1227-1233.
- Duan, D., Yue, Y., and Engelhardt, J.F. (2001). Expanding AAV Packaging Capacity with Trans-splicing or Overlapping Vectors: A Quantitative Comparison. *Mol Ther* 4, 383-391.
- Dvorak, A., Tilley, A.E., Shaykhiev, R., Wang, R., and Crystal, R.G. (2011). Do Airway Epithelium Air-Liquid Cultures Represent the In Vivo Airway Epithelium Transcriptome? *American journal of respiratory cell and molecular biology* 44, 465-473.
- Eastman, S.J., Lukason, M.J., Tousignant, J.D., Murray, H., Lane, M.D., St George, J.A., Akita, G.Y., Cherry, M., Cheng, S.H., and Scheule, R.K. (1997a). A concentrated and stable aerosol formulation of cationic lipid:DNA complexes giving high-level gene expression in mouse lung. *Hum Gene Ther* 8, 765-773.
- Eastman, S.J., Tousignant, J.D., Lukason, M.J., Murray, H., Siegel, C.S., Constantino, P., Harris, D.J., Cheng, S.H., and Scheule, R.K. (1997b). Optimization of formulations and conditions for the aerosol delivery of functional cationic lipid:DNA complexes. *Hum Gene Ther* 8, 313-322.
- Elouahabi, A., and Ruyschaert, J.-M. (2005). Formation and Intracellular Trafficking of Lipoplexes and Polyplexes. *Mol Ther* 11, 336-347.
- Emerson, B.M. (2002). Specificity of gene regulation. *Cell* 109, 267-270.
- Engelhardt, J., Yankaskas, J., and Wilson, J. (1992). In vivo retroviral gene transfer into human bronchial epithelia of xenografts. *Journal of Clinical Investigation* 90, 2598.
- Escoffre, J.-M., Teissié, J., and Rols, M.-P. (2010). Gene transfer: how can the biological barriers be overcome? *The Journal of membrane biology* 236, 61-74.
- Essner, J.J., Mclvor, R.S., and Hackett, P.B. (2005). Awakening gene therapy with Sleeping Beauty transposons. *Current Opinion in Pharmacology* 5, 513-519.

- Eszterhas, S.K., Bouhassira, E.E., Martin, D.I., and Fiering, S. (2002). Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. *Mol Cell Biol* 22, 469-479.
- Evans, M.J., Cabral, L.J., Stephens, R.J., and Freeman, G. (1975). Transformation of alveolar Type 2 cells to Type 1 cells following exposure to NO<sub>2</sub>. *Experimental and Molecular Pathology* 22, 142-150.
- Fasbender, A., Marshall, J., Moninger, T., Grunst, T., Cheng, S., and Welsh, M. (1997a). Effect of co-lipids in enhancing cationic lipid-mediated gene transfer in vitro and in vivo. *Gene Ther* 4, 716-725.
- Fasbender, A., Zabner, J., Zeiher, B.G., and Welsh, M.J. (1997b). A low rate of cell proliferation and reduced DNA uptake limit cationic lipid-mediated gene transfer to primary cultures of ciliated human airway epithelia. *Gene Ther* 4, 1173-1180.
- Faul, F., Erdfelder, E., Lang, A.G., and 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.
- Faust, S.M., Bell, P., Cutler, B.J., Ashley, S.N., Zhu, Y., Rabinowitz, J.E., and Wilson, J.M. (2013). CpG-depleted adeno-associated virus vectors evade immune detection. *The Journal of Clinical Investigation* 123, 2994-3001.
- Flamant, F., and Samarut, J. (2003). Thyroid hormone receptors: lessons from knockout and knock-in mutant mice. *Trends in Endocrinology & Metabolism* 14, 85-90.
- Flotte, T.R., and Carter, B.J. (1997). In vivo gene therapy with adeno-associated virus vectors for cystic fibrosis. *Advances in pharmacology* 40, 85-101.
- Flotte, T.R., Zeitlin, P.L., Reynolds, T.C., Heald, A.E., Pedersen, P., Beck, S., Conrad, C.K., Brass-Ernst, L., Humphries, M., and Sullivan, K. (2003). Phase I trial of intranasal and endobronchial administration of a recombinant adeno-associated virus serotype 2 (rAAV2)-CFTR vector in adult cystic fibrosis patients: a two-part clinical study. *Hum Gene Ther* 14, 1079-1088.
- Freimark, B.D., Blezinger, H.P., Florack, V.J., Nordstrom, J.L., Long, S.D., Deshpande, D.S., Nochumson, S., and Petrak, K.L. (1998). Cationic lipids enhance cytokine and cell influx levels in the lung following administration of plasmid: cationic lipid complexes. *J Immunol* 160, 4580-4586.
- Gally, F., Minor, M.N., Smith, S.K., Case, S.R., and Chu, H.W. (2011). Heat shock factor 1 protects against lung *Mycoplasma pneumoniae* infection in mice. *Journal of innate immunity* 4, 59.
- Gaudet, D., Methot, J., and Kastelein, J. (2012). Gene therapy for lipoprotein lipase deficiency. *Current Opinion in Lipidology* 23, 310-320.
- Gautam, A., Densmore, C.L., Xu, B., and Waldrep, J.C. (2000). Enhanced gene expression in mouse lung after PEI/DNA aerosol delivery. *Molecular Therapy* 2, 63-70.
- Gazdhar, A., Bilici, M., Pierog, J., Ayuni, E.L., Gugger, M., Wetterwald, A., Cecchini, M., and Schmid, R.A. (2006). In vivo electroporation and ubiquitin promoter - a protocol for sustained gene expression in the lung. *Journal of Gene Medicine* 8, 910-918.

- Geiger, J., Aneja, M.K., Hasenpusch, G., Yuksekdog, G., Kummerlowe, G., Luy, B., Romer, T., Rothbauer, U., and Rudolph, C. (2010). Targeting of the prostacyclin specific IP1 receptor in lungs with molecular conjugates comprising prostaglandin I2 analogues. *Biomaterials* *31*, 2903-2911.
- Gersbach, C.A., and Barbas, C.F., III (2013). Targeted Plasmid Integration into the Human Genome by Engineered Recombinases. In *Site-directed insertion of transgenes*, S. Renault, and P. Duchateau, eds. (Springer Netherlands), pp. 267-284.
- Ghosh, D. (1990). A relational database of transcription factors. *Nucleic Acids Res* *18*, 1749-1756.
- Ghosh, D. (2000). Object-oriented transcription factors database (ooTFD). *Nucleic Acids Res* *28*, 308-310.
- Ghosh, S., and Baltimore, D. (1990). Activation in vitro of NF-kappa B by phosphorylation of its inhibitor I kappa B. *Nature* *344*, 678-682.
- Giard, D.J., Aaronson, S.A., Todaro, G.J., Arnstein, P., Kersey, J.H., Dosik, H., and Parks, W.P. (1973). In vitro cultivation of human tumors: establishment of cell lines derived from a series of solid tumors. *J Natl Cancer Inst* *51*, 1417-1423.
- Gill, D.R., and Hyde, S.C. (2014). Delivery of genes into the CF airway. *Thorax* *69*, 962-964.
- Gill, D.R., Smyth, S.E., Goddard, C.A., Pringle, I.A., Higgins, C.F., Colledge, W.H., and Hyde, S.C. (2001). Increased persistence of lung gene expression using plasmids containing the ubiquitin C or elongation factor 1alpha promoter. *Gene Ther* *8*, 1539-1546.
- Goldman, M.J., Lee, P.-S., Yang, J.-S., and Wilson, J.M. (1997). Lentiviral vectors for gene therapy of cystic fibrosis. *Hum Gene Ther* *8*, 2261-2268.
- Goldman, M.J., and Wilson, J.M. (1995). Expression of alpha v beta 5 integrin is necessary for efficient adenovirus-mediated gene transfer in the human airway. *Journal of virology* *69*, 5951-5958.
- Gotzmann, J., Meissner, M., and Gerner, C. (2000). The fate of the nuclear matrix-associated-region-binding protein SATB1 during apoptosis. *Cell death and differentiation* *7*, 425-438.
- Goula, D., Benoist, C., Mantero, S., Merlo, G., Levi, G., and Demeneix, B.A. (1998). Polyethylenimine-based intravenous delivery of transgenes to mouse lung. *Gene Ther* *5*, 1291-1295.
- Griesenbach, U., and Alton, E.W.F.W. (2012). Progress in gene and cell therapy for cystic fibrosis lung disease. *Current pharmaceutical design* *18*, 642-662.
- Griesenbach, U., Featherstone, R.F., and Alton, E.W. (2014). New horizons for cystic fibrosis gene and cell therapy. *Cystic Fibrosis* *64*, 150.
- Griesenbach, U., Meng, C., Farley, R., Cheng, S.H., Scheule, R.K., Davies, M.H., Wolstenholme-Hogg, P.C., ten Hove, W., van der Hoeven, P., Sinn, P.L., *et al.* (2008). In vivo imaging of gene transfer to the respiratory tract. *Biomaterials* *29*, 1533-1540.

- Griesenbach, U., Meng, C., Farley, R., Gardner, A., Brake, M.A., Frankel, G.M., Gruenert, D.C., Cheng, S.H., Scheule, R.K., and Alton, E.W. (2009). The role of doxorubicin in non-viral gene transfer in the lung. *Biomaterials* 30, 1971-1977.
- Griesenbach, U., Vicente, C.C., Roberts, M.J., Meng, C., Soussi, S., Xenariou, S., Tennant, P., Baker, A., Baker, E., Gordon, C., *et al.* (2011). Secreted Gaussia luciferase as a sensitive reporter gene for in vivo and ex vivo studies of airway gene transfer. *Biomaterials* 32, 2614-2624.
- Grubb, B.R., and Boucher, R.C. (1999). *Pathophysiology of Gene-Targeted Mouse Models for Cystic Fibrosis*, Vol 79.
- Guo, X., and Huang, L. (2011). Recent Advances in Nonviral Vectors for Gene Delivery. *Accounts of Chemical Research* 45, 971-979.
- Hacein-Bey-Abina, S., Pai, S.-Y., Gaspar, H.B., Armant, M., Berry, C.C., Blanche, S., Bleesing, J., Blondeau, J., de Boer, H., Buckland, K.F., *et al.* (2014). A Modified gamma-Retrovirus Vector for X-Linked Severe Combined Immunodeficiency. *New England Journal of Medicine* 371, 1407-1417.
- Hacein-Bey-Abina, S., von Kalle, C., Schmidt, M., Le Deist, F.B., Wulffraat, N., McIntyre, E., Radford, I., Villeval, J.-L., Fraser, C.C., Cavazzana-Calvo, M., *et al.* (2003a). A Serious Adverse Event after Successful Gene Therapy for X-Linked Severe Combined Immunodeficiency. *New England Journal of Medicine* 348, 255-256.
- Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C., Pawliuk, R., and Morillon, E. (2003b). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 302, 415-419.
- Halbert, C.L., Allen, J.M., and Miller, A.D. (2002). Efficient mouse airway transduction following recombination between AAV vectors carrying parts of a larger gene. *Nat Biotech* 20, 697-701.
- Hara, Y., Rovescalli, A.C., Kim, Y., and Nirenberg, M. (1992). Structure and evolution of four POU domain genes expressed in mouse brain. *Proceedings of the National Academy of Sciences* 89, 3280-3284.
- Hart, A.H., Hartley, L., Ibrahim, M., and Robb, L. (2004). Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human. *Developmental Dynamics* 230, 187-198.
- Harvey, B.-G., Leopold, P.L., Hackett, N.R., Grasso, T.M., Williams, P.M., Tucker, A.L., Kaner, R.J., Ferris, B., Gonda, I., and Sweeney, T.D. (1999). Airway epithelial CFTR mRNA expression in cystic fibrosis patients after repetitive administration of a recombinant adenovirus. *The Journal of Clinical Investigation* 104, 1245-1255.
- Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., *et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108-112.
- Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T., Sato, S., Sanjo, H., Matsumoto, M., Hoshino, K., Wagner, H., Takeda, K., *et al.* (2000). A Toll-like receptor recognizes bacterial DNA. *Nature* 408, 740-745.

- Hidalgo, A., Cruz, A., and Perez-Gil, J. (2015). Barrier or carrier? Pulmonary surfactant and drug delivery. *Eur J Pharm Biopharm* <http://dx.doi.org/10.1016/j.ejpb.2015.02.014>.
- Hiemstra, P.S., McCray, P.B., and Bals, R. (2015). The innate immune function of airway epithelial cells in inflammatory lung disease. *European Respiratory Journal* *45*, 1150-1162.
- Hodges, B.L., Taylor, K.M., Joseph, M.F., Bourgeois, S.A., and Scheule, R.K. (2004). Long-term transgene expression from plasmid DNA gene therapy vectors is negatively affected by CpG dinucleotides. *Mol Ther* *10*, 269-278.
- Hong, K.U., Reynolds, S.D., Giangreco, A., Hurley, C.M., and Stripp, B.R. (2001). Clara Cell Secretory Protein-Expressing Cells of the Airway Neuroepithelial Body Microenvironment Include a Label-Retaining Subset and Are Critical for Epithelial Renewal after Progenitor Cell Depletion. *American journal of respiratory cell and molecular biology* *24*, 671-681.
- Hong, K.U., Reynolds, S.D., Watkins, S., Fuchs, E., and Stripp, B.R. (2004). Basal Cells Are a Multipotent Progenitor Capable of Renewing the Bronchial Epithelium. *The American Journal of Pathology* *164*, 577-588.
- Hou, Y., Zhou, X., Liu, J., Yuan, J., Cheng, H., and Zhou, R. (2010). Nuclear factor-Y (NF-Y) regulates transcription of mouse *Dmrt7* gene by binding to tandem CCAAT boxes in its proximal promoter. *International Journal of Biological Sciences* *6*, 655-664.
- Howe, S.J., Mansour, M.R., Schwarzwaelder, K., Bartholomae, C., Hubank, M., Kempinski, H., Brugman, M.H., Pike-Overzet, K., Chatters, S.J., de Ridder, D., *et al.* (2008). Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *The Journal of Clinical Investigation* *118*, 3143-3150.
- Hsieh, C.L. (1997). Stability of patch methylation and its impact in regions of transcriptional initiation and elongation. *Mol Cell Biol* *17*, 5897-5904.
- Huang, M., Chen, Z., Hu, S., Jia, F., Li, Z., Hoyt, G., Robbins, R.C., Kay, M.A., and Wu, J.C. (2009). Novel minicircle vector for gene therapy in murine myocardial infarction. *Circulation* *120*, S230-237.
- Huang, S., Wiszniewski, L., and Constant, S. (2011). The use of in vitro 3D cell models in drug development for respiratory diseases (INTECH Open Access Publisher).
- Hung, C., Linn, G., Chow, Y.-H., Kobayashi, A., Mittelsteadt, K., Altemeier, W.A., Gharib, S.A., Schnapp, L.M., and Duffield, J.S. (2013). Role of Lung Pericytes and Resident Fibroblasts in the Pathogenesis of Pulmonary Fibrosis. *American Journal of Respiratory and Critical Care Medicine* *188*, 820-830.
- Huntley, R.P., Sawford, T., Mutowo-Meullenet, P., Shypitsyna, A., Bonilla, C., Martin, M.J., and O'Donovan, C. (2015). The GOA database: Gene Ontology annotation updates for 2015. *Nucleic Acids Res* *43*, D1057-D1063.
- Hyde, S.C., Gill, D.R., Higgins, C.F., Trezise, A.E., MacVinish, L.J., Cuthbert, A.W., Ratcliff, R., Evans, M.J., and Colledge, W.H. (1993). Correction of the ion transport defect in cystic fibrosis transgenic mice by gene therapy. *Nature* *362*, 250-255.
- Hyde, S.C., Pringle, I.A., Abdullah, S., Lawton, A.E., Davies, L.A., Varathalingam, A., Nunez-Alonso, G., Green, A.M., Bazzani, R.P., Sumner-Jones, S.G., *et al.* (2008). CpG-free plasmids confer reduced inflammation and sustained pulmonary gene expression. *Nat Biotechnol* *26*, 549-551.

- Hyde, S.C., Southern, K.W., Gileadi, U., Fitzjohn, E.M., Mofford, K.A., Waddell, B.E., Gooi, H.C., Goddard, C.A., Hannavy, K., Smyth, S.E., *et al.* (2000). Repeat administration of DNA/liposomes to the nasal epithelium of patients with cystic fibrosis. *Gene Ther* 7, 1156-1165.
- Iborra, F.J., Pombo, A., Jackson, D.A., and Cook, P.R. (1996). Active RNA polymerases are localized within discrete transcription "factories" in human nuclei. *J Cell Sci* 109 ( Pt 6), 1427-1436.
- Ibraheem, D., Elaissari, A., and Fessi, H. (2014). Gene therapy and DNA delivery systems. *International Journal of Pharmaceutics* 459, 70-83.
- Ishii, K.J., and Akira, S. (2006). Innate immune recognition of, and regulation by, DNA. *Trends in Immunology* 27, 525-532.
- Islam, M.A., Park, T.Ä., Singh, B., Maharjan, S., Firdous, J., Cho, M.-H., Kang, S.-K., Yun, C.-H., Choi, Y.Ä., and Cho, C.-S. (2014). Major degradable polycations as carriers for DNA and siRNA. *J Control Release* 193, 74-89.
- Ivics, Z., Hackett, P.B., Plasterk, R.H., and Izsvak, Z. (1997). Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells. *Cell* 91, 501-510.
- Jans, D.A., and Hubner, S. (1996). Regulation of protein transport to the nucleus: central role of phosphorylation. *Physiological Reviews* 76, 651-685.
- Jenke, B.H.C., Fetzer, C.P., Stehle, I.M., Jönsson, F., Fackelmayer, F.O., Conradt, H., Bode, J., and Lipps, H.J. (2002). An episomally replicating vector binds to the nuclear matrix protein SAF-A in vivo. *EMBO reports* 3, 349-354.
- Jiang, C., O'Connor, S.P., Fang, S.L., Wang, K.X., Marshall, J., Williams, J.L., Wilburn, B., Echelard, Y., and Cheng, S.H. (1998). Efficiency of cationic lipid-mediated transfection of polarized and differentiated airway epithelial cells in vitro and in vivo. *Hum Gene Ther* 9, 1531-1542.
- Jiang, H.-L., Xu, C.-X., Kim, Y.-K., Arote, R., Jere, D., Lim, H.-T., Cho, M.-H., and Cho, C.-S. (2009). The suppression of lung tumorigenesis by aerosol-delivered folate-chitosan-graft-polyethylenimine/Akt1 shRNA complexes through the Akt signaling pathway. *Biomaterials* 30, 5844-5852.
- Johnson, L.G., Olsen, J.C., Sarkadi, B., Moore, K.L., Swanstrom, R., and Boucher, R.C. (1992). Efficiency of gene transfer for restoration of normal airway epithelial function in cystic fibrosis. *Nat Genet* 2, 21-25.
- Johnson, L.G., Randell, S.H., and Olsen, J.C. (2003a). Airway epithelia. In *Lentivirus Gene Engineering Protocols* (Springer), pp. 181-196.
- Johnson, L.G., Vanhook, M.K., Coyne, C.B., Haykal-Coates, N., and Gavett, S.H. (2003b). Safety and efficiency of modulating paracellular permeability to enhance airway epithelial gene transfer in vivo. *Hum Gene Ther* 14, 729-747.
- Kaepfel, C., Beattie, S.G., Fronza, R., van Logtenstein, R., Salmon, F., Schmidt, S., Wolf, S., Nowrouzi, A., Glimm, H., von Kalle, C., *et al.* (2013). A largely random AAV integration profile after LPLD gene therapy. *Nat Med* 19, 889-891.

- Kalderon, D., Roberts, B.L., Richardson, W.D., and Smith, A.E. (1984). A short amino acid sequence able to specify nuclear location. *Cell* 39, 499-509.
- Kaufman, C.D., Geiger, R.C., and Dean, D.A. (2010). Electroporation- and mechanical ventilation-mediated gene transfer to the lung. *Gene Ther* 17, 1098-1104.
- Kitson, C., Angel, B., Judd, D., Rothery, S., Severs, N.J., Dewar, A., Huang, L., Wadsworth, S.C., Cheng, S.H., Geddes, D.M., *et al.* (1999). The extra- and intracellular barriers to lipid and adenovirus-mediated pulmonary gene transfer in native sheep airway epithelium. *Gene Ther* 6, 534-546.
- Klinman, D.M., Yi, A.K., Beaucage, S.L., Conover, J., and Krieg, A.M. (1996). CpG motifs present in bacteria DNA rapidly induce lymphocytes to secrete interleukin 6, interleukin 12, and interferon gamma. *Proceedings of the National Academy of Sciences* 93, 2879-2883.
- Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 31, 89-97.
- Kobayashi, M., Iida, A., Ueda, Y., and Hasegawa, M. (2003). Pseudotyped lentivirus vectors derived from simian immunodeficiency virus SIVagm with envelope glycoproteins from paramyxovirus. *J Virol* 77, 2607-2614.
- Kobinger, G.P., Weiner, D.J., Yu, Q.-C., and Wilson, J.M. (2001). Filovirus-pseudotyped lentiviral vector can efficiently and stably transduce airway epithelia in vivo. *Nat Biotech* 19, 225-230.
- Koch, C., and Hoiby, N. (1993). Pathogenesis of cystic fibrosis. *The Lancet* 341, 1065-1069.
- Kochanek, S., Clemens, P.R., Mitani, K., Chen, H.H., Chan, S., and Caskey, C.T. (1996). A new adenoviral vector: Replacement of all viral coding sequences with 28 kb of DNA independently expressing both full-length dystrophin and beta-galactosidase. *Proceedings of the National Academy of Sciences* 93, 5731-5736.
- Konstan, M.W., Davis, P.B., Wagener, J.S., Hilliard, K.A., Stern, R.C., Milgram, L.J., Kowalczyk, T.H., Hyatt, S.L., Fink, T.L., Gedeon, C.R., *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.
- Koshkina, N.V., Agoulnik, I.Y., Melton, S.L., Densmore, C.L., and Knight, V. (2003). Biodistribution and pharmacokinetics of aerosol and intravenously administered DNA-polyethyleneimine complexes: optimization of pulmonary delivery and retention. *Mol Ther* 8, 249-254.
- Krieg, A.M., and Kline, J.N. (2000). Immune effects and therapeutic applications of CpG motifs in bacterial DNA. *Immunopharmacology* 48, 303-305.
- Krieg, A.M., Yi, A.K., Matson, S., Waldschmidt, T.J., Bishop, G.A., Teasdale, R., Koretzky, G.A., and Klinman, D.M. (1995). CpG motifs in bacterial DNA trigger direct B-cell activation. *Nature* 374, 546-549.
- Kumar, P.A., Hu, Y., Yamamoto, Y., Hoe, N.B., Wei, T.S., Mu, D., Sun, Y., Joo, L.S., Dagher, R., Zielonka, E.M., *et al.* (2011). Distal Airway Stem Cells Yield Alveoli In-vitro and during Lung Regeneration following H1N1 Influenza Infection. *Cell* 147, 525-538.

- Lalonde, M.-E., Cheng, X., and Cote, J. (2014). Histone target selection within chromatin: an exemplary case of teamwork. *Genes & development* 28, 1029-1041.
- Lam, M.T.Y., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. *Trends in Biochemical Sciences* 39, 170-182.
- Larkin, J.D., Papantonis, A., and Cook, P.R. (2013). Promoter type influences transcriptional topography by targeting genes to distinct nucleoplasmic sites. *J Cell Sci* 126, 2052-2059.
- Lattin, J.E., Schroder, K., Su, A.I., Walker, J.R., Zhang, J., Wiltshire, T., Saijo, K., Glass, C.K., Hume, D.A., and Kellie, S. (2008). Expression analysis of G Protein-Coupled Receptors in mouse macrophages. *Immunome Res* 4, 10.1186.
- Lawton, A.E., Pringle, I.A., Davies, L.A., Hyde, S.C., and Gill, D.R. (2002). Non-viral integrating vectors for the lung: Sleeping beauty transposons. *Mol Ther* 5, S323.
- Le Gall, T., Barbeau, J., Barrier, S., Berchel, M., Lemiegre, L., Jettic, J., Meriadec, C., Artzner, F., Gill, D.R., Hyde, S.C., *et al.* (2014). Effects of a Novel Archaeal Tetraether-Based Colipid on the In Vivo Gene Transfer Activity of Two Cationic Amphiphiles. *Molecular Pharmaceutics* 11, 2973-2988.
- Lechardeur, D., Verkman, A.S., and Lukacs, G.L. (2005). Intracellular routing of plasmid DNA during non-viral gene transfer. *Adv Drug Deliv Rev* 57, 755-767.
- Lee, E.R., Marshall, J., Siegel, C.S., Jiang, C., Yew, N.S., Nichols, M.R., Nietupski, J.B., Ziegler, R.J., Lane, M.B., Wang, K.X., *et al.* (1996). Detailed analysis of structures and formulations of cationic lipids for efficient gene transfer to the lung. *Hum Gene Ther* 7, 1701-1717.
- Lee, K.-H., Byun, S.S., Paik, J.-Y., Lee, S.Y., Song, S.H., Choe, Y.S., and Kim, B.-T. (2003). Cell uptake and tissue distribution of radioiodine labelled D-luciferin: implications for luciferase based gene imaging. *Nuclear Medicine Communications* 24, 1003-1009.
- Li, S., MacLaughlin, F., Fewell, J., Gondo, M., Wang, J., Nicol, F., Dean, D., and Smith, L. (2001). Muscle-specific enhancement of gene expression by incorporation of SV40 enhancer in the expression plasmid. *Gene Ther.* 8, 494-497.
- Li, S., Wu, S.P., Whitmore, M., Loeffert, E.J., Wang, L., Watkins, S.C., Pitt, B.R., and Huang, L. (1999). Effect of immune response on gene transfer to the lung via systemic administration of cationic lipidic vectors. *Am J Physiol* 276, L796-804.
- Limberis, M., Anson, D.S., Fuller, M., and Parsons, D.W. (2002). Recovery of airway cystic fibrosis transmembrane conductance regulator function in mice with cystic fibrosis after single-dose lentivirus-mediated gene transfer. *Hum Gene Ther* 13, 1961-1970.
- Limberis, M.P., Bell, C.L., Heath, J., and Wilson, J.M. (2010). Activation of transgene-specific T cells following lentivirus-mediated gene delivery to mouse lung. *Molecular Therapy* 18, 143-150.
- Limberis, M.P., Bell, C.L., and Wilson, J.M. (2009). Identification of the murine firefly luciferase-specific CD8 T-cell epitopes. *Gene Ther* 16, 441-447.
- Limberis, M.P., and Wilson, J.M. (2006). Adeno-associated virus serotype 9 vectors transduce murine alveolar and nasal epithelia and can be readministered. *Proceedings of the National Academy of Sciences* 103, 12993-12998.

- Lindee, S., and Mueller, R. (2011). Is cystic fibrosis genetic medicine's canary? *Perspect Biol Med* 54, 316-331.
- Linden, S.K., Sutton, P., Karlsson, N.G., Korolik, V., and McGuckin, M.A. (2008). Mucins in the mucosal barrier to infection. *Mucosal Immunol* 1, 183-197.
- Liu, F., Song, Y.K., and Liu, D. (1999). Hydrodynamics-based transfection in animals by systemic administration of plasmid DNA. *Gene Ther.* 6, 1258-1266.
- Liu, G., Li, D., Pasumarthy, M.K., Kowalczyk, T.H., Gedeon, C.R., Hyatt, S.L., Payne, J.M., Miller, T.J., Brunovskis, P., and Fink, T.L. (2003). Nanoparticles of compacted DNA transfect postmitotic cells. *Journal of Biological Chemistry* 278, 32578-32586.
- Liu, L., Sanz, S., Heggestad, A.D., Antharam, V., Notterpek, L., and Fletcher, B.S. (2004). Endothelial Targeting of the Sleeping Beauty Transposon within Lung. *Mol Ther* 10, 97-105.
- Liu, X., and Engelhardt, J.F. (2008). The glandular stem/progenitor cell niche in airway development and repair. *Proc Am Thorac Soc* 5, 682-688.
- Liu, X., Wu, B., Szary, J., Kofoed, E.M., and Schaufele, F. (2007). Functional Sequestration of Transcription Factor Activity by Repetitive DNA. *Journal of Biological Chemistry* 282, 20868-20876.
- Liu, X., Yan, Z., Luo, M., and Engelhardt, J.F. (2006). Species-specific differences in mouse and human airway epithelial biology of recombinant adeno-associated virus transduction. *Am J Respir Cell Mol Biol* 34, 56-64.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8Å resolution. *Nature* 389, 251-260.
- Lukacs, G.L., Haggie, P., Seksek, O., Lechardeur, D., Freedman, N., and Verkman, A. (2000). Size-dependent DNA mobility in cytoplasm and nucleus. *Journal of Biological Chemistry* 275, 1625-1629.
- Machado-Aranda, D., Adir, Y., Young, J.L., Briva, A., Budinger, G.R., Yeldandi, A.V., Sznajder, J.I., and Dean, D.A. (2005). Gene transfer of the Na<sup>+</sup>,K<sup>+</sup>-ATPase beta1 subunit using electroporation increases lung liquid clearance. *Am J Respir Crit Care Med* 171, 204-211.
- Magnusson, T. (2010). Optimised plasmids for sustained transgene expression in vivo. Ph.D. thesis. In Department of Pharmaceutical Biology-Biotechnology (Munich, Ludwig-Maximilians University), p. 80.
- Magnusson, T., Haase, R., Schleef, M., Wagner, E., and Ogris, M. (2011). Sustained, high transgene expression in liver with plasmid vectors using optimized promoter-enhancer combinations. *J Gene Med* 13, 382-391.
- Manzini, S., Vargiolu, A., Stehle, I.M., Bacci, M.L., Cerrito, M.G., Giovannoni, R., Zannoni, A., Bianco, M.R., Forni, M., Donini, P., *et al.* (2006). Genetically modified pigs produced with a nonviral episomal vector. *Proceedings of the National Academy of Sciences* 103, 17672-17677.
- Marie, C., Vandermeulen, G., Quiviger, M., Richard, M., Pr eat, V., and Scherman, D. (2010). pFARs, Plasmids free of antibiotic resistance markers, display high-level transgene expression in muscle, skin and tumour cells. *The Journal of Gene Medicine* 12, 323-332.

- Matsui, H., Grubb, B.R., Tarran, R., Randell, S.H., Gatzky, J.T., Davis, C.W., and Boucher, R.C. (1998a). Evidence for Periciliary Liquid Layer Depletion, Not Abnormal Ion Composition, in the Pathogenesis of Cystic Fibrosis Airways Disease. *Cell* 95, 1005-1015.
- Matsui, H., Johnson, L.G., Randell, S.H., and Boucher, R.C. (1997). Loss of binding and entry of liposome-DNA complexes decreases transfection efficiency in differentiated airway epithelial cells. *Journal of Biological Chemistry* 272, 1117-1126.
- Matsui, H., Randell, S.H., Peretti, S.W., Davis, C.W., and Boucher, R.C. (1998b). Coordinated clearance of periciliary liquid and mucus from airway surfaces. *The Journal of Clinical Investigation* 102, 1125-1131.
- McLachlan, G., Davidson, H., Holder, E., Davies, L.A., Pringle, I.A., Sumner-Jones, S.G., 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.
- Medina, M., Kobinger, G.P., Rux, J., Gasmi, M., Looney, D.J., Bates, P., and Wilson, J.M. (2003). Lentiviral Vectors Pseudotyped with Minimal Filovirus Envelopes Increased Gene Transfer in Murine Lung. *Mol Ther* 8, 777-789.
- Mercer, R.R., Russell, M.L., Roggli, V.L., and Crapo, J.D. (1994). Cell number and distribution in human and rat airways. *American journal of respiratory cell and molecular biology* 10, 613-624.
- Mesika, A., Grigoreva, I., Zohar, M., and Reich, Z. (2001). A regulated, NFkappaB-assisted import of plasmid DNA into mammalian cell nuclei. *Mol Ther* 3, 653-657.
- Meyer, K.B., Thompson, M.M., Levy, M.Y., Barron, L.G., and Szoka, F.C., Jr. (1995). Intratracheal gene delivery to the mouse airway: characterization of plasmid DNA expression and pharmacokinetics. *Gene Ther* 2, 450-460.
- Miller, D.G., Adam, M.A., and Miller, A.D. (1990). Gene transfer by retrovirus vectors occurs only in cells that are actively replicating at the time of infection. *Mol Cell Biol* 10, 4239-4242.
- Mitomo, K., Griesenbach, U., Inoue, M., Somerton, L., Meng, C., Akiba, E., Tabata, T., Ueda, Y., Frankel, G.M., Farley, R., *et al.* (2010). Toward gene therapy for cystic fibrosis using a lentivirus pseudotyped with Sendai virus envelopes. *Mol Ther* 18, 1173-1182.
- Miwa, A., Higuchi, T., and Kobayashi, S. (2006). Expression and polysome association of YB-1 in various tissues at different stages in the lifespan of mice. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1760, 1675-1681.
- Moss, R.B., Milla, C., Colombo, J., Accurso, F., Zeitlin, P.L., Clancy, J.P., Spencer, L.T., Pilewski, J., Waltz, D.A., and Dorkin, H.L. (2007). Repeated aerosolized AAV-CFTR for treatment of cystic fibrosis: a randomized placebo-controlled phase 2B trial. *Hum Gene Ther* 18, 726-732.
- Munkonge, F.M., Dean, D.A., Hillery, E., Griesenbach, U., and Alton, E.W.F.W. (2003). Emerging significance of plasmid DNA nuclear import in gene therapy. *Advanced drug delivery reviews* 55, 749-760.
- Murphy, S.V., and Atala, A. (2013). Cell therapy for cystic fibrosis. *J Tissue Eng Regen Med*.

- Nabel, G., and Baltimore, D. (1987). An inducible transcription factor activates expression of human immunodeficiency virus in T cells [published erratum appears in *Nature* 1990 Mar 8; 344(6262):178]. *Nature* 326, 711-713.
- Nair, N., Rincon, M.Y., Evens, H., Sarcar, S., Dastidar, S., Samara-Kuko, E., Ghandeharian, O., Man Viecelli, H., Thony, B., De Bleser, P., *et al.* (2014). Computationally designed liver-specific transcriptional modules and hyperactive factor IX improve hepatic gene therapy. *Blood* 123, 3195-3199.
- Nakamura, S., Maehara, T., Watanabe, S., Ishihara, M., and Sato, M. (2013). Improvement of hydrodynamics-based gene transfer of nonviral DNA targeted to murine hepatocytes. *BioMed Research International* 2013.
- Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F.H., Verma, I.M., and Trono, D. (1996). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* 272, 263-267.
- Naltner, A., Wert, S., Whitsett, J.A., and Yan, C. (2000). Temporal/spatial expression of nuclear receptor coactivators in the mouse lung. *American Journal of Physiology-Lung Cellular and Molecular Physiology* 279, L1066-L1074.
- Nardone, L.L., and Andrews, S.B. (1979). Cell line A549 as a model of the type II pneumocyte. Phospholipid biosynthesis from native and organometallic precursors. *Biochim Biophys Acta* 573, 276-295.
- Newell-Price, J., Clark, A.J.L., and King, P. (2000). DNA Methylation and Silencing of Gene Expression. *Trends in Endocrinology & Metabolism* 11, 142-148.
- Ochiai, H., Fujimuro, M., Yokosawa, H., Harashima, H., and Kamiya, H. (2007). Transient activation of transgene expression by hydrodynamics-based injection may cause rapid decrease in plasmid DNA expression. *Gene Ther* 14, 1152-1159.
- Ochiai, H., Harashima, H., and Kamiya, H. (2006). Intranuclear disposition of exogenous DNA in vivo: Silencing, methylation and fragmentation. *FEBS Letters* 580, 918-922.
- Oliveira, C., Silveira, I., Veiga, F., and Ribeiro, A.n.J. (2014). Recent advances in characterization of nonviral vectors for delivery of nucleic acids: impact on their biological performance. *Expert Opinion on Drug Delivery* 12, 27-39.
- Oliveira, P.H., and Mairhofer, J. (2013). Marker-free plasmids for biotechnological applications- implications and perspectives. *Trends in Biotechnology* 31, 539-547.
- Orphanides, G., and Reinberg, D. (2002). A Unified Theory of Gene Expression. *Cell* 108, 439-451.
- Ostedgaard, L.S., Zabner, J., Vermeer, D.W., Rokhlina, T., Karp, P.H., Stecenko, A.A., Randak, C., and Welsh, M.J. (2002). CFTR with a partially deleted R domain corrects the cystic fibrosis chloride transport defect in human airway epithelia in vitro and in mouse nasal mucosa in vivo. *P Natl Acad Sci USA* 99, 3093-3098.
- Palmer, D.J., and Ng, P. (2005). Helper-dependent adenoviral vectors for gene therapy. *Hum Gene Ther* 16, 1-16.
- Parks, R.J., Chen, L., Anton, M., Sankar, U., Rudnicki, M.A., and Graham, F.L. (1996). A helper-dependent adenovirus vector system: Removal of helper virus by Cre-mediated

excision of the viral packaging signal. *Proceedings of the National Academy of Sciences* 93, 13565-13570.

Perricone, M.A., Morris, J.E., Pavelka, K., Plog, M.S., O'Sullivan, B.P., Joseph, P.M., Dorkin, H., Lapey, A., Balfour, R., and Meeker, D.P. (2001). Aerosol and lobar administration of a recombinant adenovirus to individuals with cystic fibrosis. II. Transfection efficiency in airway epithelium. *Hum Gene Ther* 12, 1383-1394.

Petrick, J.S., and Klaassen, C.D. (2007). Importance of hepatic induction of constitutive androstane receptor and other transcription factors that regulate xenobiotic metabolism and transport. *Drug Metabolism and Disposition* 35, 1806-1815.

Pettit, R.S., and Fellner, C. (2014). CFTR Modulators for the Treatment of Cystic Fibrosis. *Pharmacy and Therapeutics* 39, 500.

Pickles, R.J., Fahrner, J.A., Petrella, J.M., Boucher, R.C., and Bergelson, J.M. (2000). Retargeting the coxsackievirus and adenovirus receptor to the apical surface of polarized epithelial cells reveals the glycocalyx as a barrier to adenovirus-mediated gene transfer. *J Virol* 74, 6050-6057.

Pickles, R.J., McCarty, D., Matsui, H., Hart, P.d.J., Randell, S.H., and Boucher, R.C. (1998). Limited Entry of Adenovirus Vectors into Well-Differentiated Airway Epithelium Is Responsible for Inefficient Gene Transfer. *Journal of virology* 72, 6014-6023.

Piechaczek, C., Fetzter, C., Baiker, A., Bode, J., and Lipps, H.J. (1999). A vector based on the SV40 origin of replication and chromosomal S/MARs replicates episomally in CHO cells. *Nucleic Acids Res* 27, 426-428.

Podetz-Pedersen, K.M., Vezys, V., Somia, N.V., Russell, S.J., and McIvor, R.S. (2014). Cellular immune response against firefly luciferase after sleeping beauty-mediated gene transfer in vivo. *Hum Gene Ther* 25, 955-965.

Pollard, H., Toumaniantz, G., Amos, J.-L., Avet-Loiseau, H., Guihard, G., Behr, J.-P., and Escande, D. (2001). Ca<sup>2+</sup>-sensitive cytosolic nucleases prevent efficient delivery to the nucleus of injected plasmids. *The Journal of Gene Medicine* 3, 153-164.

Pringle, I.A., Alton, E.W.F.W., Connolly, M.M., Chan, M., Davies, J.C., Davies, L.A., Gea-Sorli, S., Gill, D.R., Griesenbach, U., Hasegawa, M., *et al.* (2014). Clinical Development of an Optimal F/HN Pseudotyped SIV Vector for Cystic Fibrosis Lung Gene Therapy. *Molecular Therapy* 22, S141.

Pringle, I.A., Connolly, M.M., Lawton, A.E., Davies, L.A., Hyde, S.C., and Gill, D.R. (2010). Duration of Expression from CpG-Free Plasmids Following Hydrodynamic Delivery to the Mouse. *Mol Ther* 18, S283.

Pringle, I.A., Gill, D.R., Connolly, M.M., Lawton, A.E., Hewitt, A.M., Nunez-Alonso, G., Cheng, S.H., Scheule, R.K., Davies, L.A., and Hyde, S.C. (2012a). Rapid identification of novel functional promoters for gene therapy. *J Mol Med (Berl)*.

Pringle, I.A., Hyde, S.C., Connolly, M.M., Lawton, A.E., Xu, B., Nunez-Alonso, G., Davies, L.A., Sumner-Jones, S.G., and Gill, D.R. (2012b). CpG-free plasmid expression cassettes for cystic fibrosis gene therapy. *Biomaterials* 33, 6833-6842.

Pringle, I.A., McLachlan, G., Collie, D.D.S., Sumner-Jones, S.G., Lawton, A.E., Tennant, P., Baker, A., Gordon, C., Blundell, R., Varathalingam, A., *et al.* (2007). Electroporation

- enhances reporter gene expression following delivery of naked plasmid DNA to the lung. *Journal of Gene Medicine* 9, 369-380.
- Pringle, I.A., Raman, S., Sharp, W.W., Cheng, S.H., Hyde, S.C., and Gill, D.R. (2005). Detection of plasmid DNA vectors following gene transfer to the murine airways. *Gene Ther* 12, 1206-1214.
- Qin, L., Ding, Y., Pahud, D.R., Chang, E., Imperiale, M.J., and Bromberg, J.S. (1997). Promoter attenuation in gene therapy: interferon-gamma and tumor necrosis factor-alpha inhibit transgene expression. *Hum Gene Ther* 8, 2019-2029.
- Quiviger, M., Arfi, A., Mansard, D., Delacotte, L., Pastor, M., Scherman, D., and Marie, C. (2014). High and prolonged sulfamidase secretion by the liver of MPS-IIIa mice following hydrodynamic tail vein delivery of antibiotic-free pFAR4 plasmid vector. *Gene Ther* 21, 1001-1007.
- Ramsey, B.W., Davies, J., McElvaney, N.G., Tullis, E., Bell, S.C., Devine, P., Griese, M., McKone, E.F., Wainwright, C.E., Konstan, M.W., *et al.* (2011). A CFTR Potentiator in Patients with Cystic Fibrosis and the G551D Mutation. *New England Journal of Medicine* 365, 1663-1672.
- Ravasi, T., Suzuki, H., Cannistraci, C.V., Katayama, S., Bajic, V.B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., *et al.* (2010). An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* 140, 744-752.
- Rawlins, E.L., and Hogan, B.L. (2008). Ciliated epithelial cell lifespan in the mouse trachea and lung. *Am J Physiol Lung Cell Mol Physiol* 295, L231-234.
- Rawlins, E.L., and Hogan, B.L.M. (2006). Epithelial stem cells of the lung: privileged few or opportunities for many? *Development* 133, 2455-2465.
- Rawlins, E.L., Ostrowski, L.E., Randell, S.H., and Hogan, B.L.M. (2007). Lung development and repair: contribution of the ciliated lineage. *Proceedings of the National Academy of Sciences* 104, 410-417.
- Rettig, G.R., McAnuff, M., Liu, D., Kim, J.S., and Rice, K.G. (2006). Quantitative bioluminescence imaging of transgene expression in vivo. *Anal Biochem* 355, 90-94.
- Rincon, M.Y., Sarcar, S., Danso-Abeam, D., Keyaerts, M., Matrai, J., Samara-Kuko, E., Acosta-Sanchez, A., Athanasopoulos, T., Dickson, G., Lahoutte, T., *et al.* (2015). Genome-wide computational analysis reveals cardiomyocyte-specific transcriptional Cis-regulatory motifs that enable efficient cardiac gene therapy. *Mol Ther* 23, 43-52.
- Riordan, J.R., Rommens, J.M., Kerem, B.-s., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., and Chou, J.-L. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 245, 1066-1073.
- Riu, E., Chen, Z.Y., Xu, H., He, C.Y., and Kay, M.A. (2007). Histone modifications are associated with the persistence or silencing of vector-mediated transgene expression in vivo. *Mol Ther* 15, 1348-1355.
- Rock, J.R., Randell, S.H., and Hogan, B.L.M. (2010). Airway basal stem cells: a perspective on their roles in epithelial homeostasis and remodeling. *Dis Model Mech* 3, 545-556.

- Rogers, C.S., Stoltz, D.A., Meyerholz, D.K., Ostedgaard, L.S., Rokhlina, T., Taft, P.J., Rogan, M.P., Pezzulo, A.A., Karp, P.H., Itani, O.A., *et al.* (2008). Disruption of the CFTR Gene Produces a Model of Cystic Fibrosis in Newborn Pigs. *Science* 321, 1837-1841.
- Rose, A., Goddard, C., Colledge, W., Cheng, S., Gill, D., and Hyde, S. (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.
- Rudolph, C., Lausier, J., Naundorf, S., Müller, R.H., and Rosenecker, J. (2000). In vivo gene delivery to the lung using polyethylenimine and fractured polyamidoamine dendrimers. *The Journal of Gene Medicine* 2, 269-278.
- Rudolph, C., Ortiz, A., Schillinger, U., Jauernig, J., Plank, C., and Rosenecker, J. (2005). Methodological optimization of polyethylenimine (PEI)-based gene delivery to the lungs of mice via aerosol application. *The Journal of Gene Medicine* 7, 59-66.
- Ruiz, F.E., Clancy, J.P., Perricone, M.A., Bebok, Z., Hong, J.S., Cheng, S.H., Meeker, D.P., Young, K.R., Schoumacher, R.A., Weatherly, M.R., *et al.* (2001). A clinical inflammatory syndrome attributable to aerosolized lipid-DNA administration in cystic fibrosis. *Hum Gene Ther* 12, 751-761.
- Russell, W.M.S., and Burch, R.L. (1959). *The Principles of Humane Experimental Technique* (London: Methuen & Co, Ltd).
- Sakurai, F., Nishioka, T., Saito, H., Baba, T., Okuda, A., Matsumoto, O., Taga, T., Yamashita, F., Takakura, Y., and Hashida, M. (2001). Interaction between DNA-cationic liposome complexes and erythrocytes is an important factor in systemic gene transfer via the intravenous route in mice: the role of the neutral helper lipid. *Gene Ther.* 8, 677-686.
- Sanders, N., Rudolph, C., Braeckmans, K., De Smedt, S.C., and Demeester, J. (2009). Extracellular barriers in respiratory gene therapy. *Advanced drug delivery reviews* 61, 115-127.
- Saridey, S.K., Liu, L., Doherty, J.E., Kaja, A., Galvan, D.L., Fletcher, B.S., and Wilson, M.H. (2009). PiggyBac Transposon-based Inducible Gene Expression In Vivo After Somatic Cell Gene Transfer. *Mol Ther* 17, 2115-2120.
- Sato, A., Klaunberg, B., and Tolwani, R. (2004). In vivo bioluminescence imaging. *Comp Med* 54, 631-634.
- Schaarschmidt, D., Baltin, J., Stehle, I.M., Lipps, H.J., and Knippers, R. (2004). An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. *The EMBO Journal* 23, 191-201.
- Schakowski, F., Gorschluter, M., Buttgereit, P., Maerten, A., Lilienfeld-Toal, M.V., Junghans, C., Schroff, M., Koenig-Merediz, S.A., Ziske, C., and Strehl, J. (2007). Minimal size MIDGE vectors improve transgene expression in vivo. *in vivo* 21, 17-23.
- Schakowski, F., Gorschluter, M., Junghans, C., Schroff, M., Buttgereit, P., Ziske, C., Schottker, B., Konig-Merediz, S.A., Sauerbruch, T., Wittig, B., *et al.* (2001). A novel minimal-size vector (MIDGE) improves transgene expression in colon carcinoma cells and avoids transfection of undesired DNA. *Mol Ther* 3, 793-800.
- Scheule, R.K. (2000). The role of CpG motifs in immunostimulation and gene therapy. *Advanced drug delivery reviews* 44, 119-134.

- Scheule, R.K., StGeorge, J.A., Bagley, R.G., Marshall, J., Kaplan, J.M., Akita, G.Y., Wang, K.X., Lee, E.R., Harris, D.J., Jiang, C.W., *et al.* (1997). Basis of pulmonary toxicity associated with cationic lipid-mediated gene transfer to the mammalian lung. *Hum Gene Ther* 8, 689-707.
- 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.
- Schmid, H., Boucherot, A., Yasuda, Y., Henger, A., Brunner, B., Eichinger, F., Nitsche, A., Kiss, E., Bleich, M., Grone, H.J., *et al.* (2006). Modular activation of nuclear factor-kappaB transcriptional programs in human diabetic nephropathy. *Diabetes* 55, 2993-3003.
- Schnepp, B.C., Jensen, R.L., Chen, C.-L., Johnson, P.R., and Clark, K.R. (2005). Characterization of adeno-associated virus genomes isolated from human tissues. *Journal of virology* 79, 14793-14803.
- Sciavolino, P.J., Abrams, E.W., Yang, L., Austenberg, L.P., Shen, M.M., and Abate-Shen, C. (1997). Tissue-specific expression of murine Nkx3.1 in the male urogenital system. *Developmental Dynamics* 209, 127-138.
- Sheppard, D.N., Ostedgaard, L.S., Winter, M.C., and Welsh, M.J. (1995). Mechanism of dysfunction of two nucleotide binding domain mutations in cystic fibrosis transmembrane conductance regulator that are associated with pancreatic sufficiency. *The EMBO Journal* 14, 876-883.
- Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* 15, 272-286.
- Shu, W., Yang, H., Zhang, L., Lu, M.M., and Morrisey, E.E. (2001). Characterization of a New Subfamily of Winged-helix/Forkhead (Fox) Genes That Are Expressed in the Lung and Act as Transcriptional Repressors. *Journal of Biological Chemistry* 276, 27488-27497.
- Sinn, P.L., Anthony, R.M., and McCray, P.B. (2011). Genetic therapies for cystic fibrosis lung disease. *Human Molecular Genetics* 20, R79-R86.
- Sinn, P.L., Burnight, E.R., Hickey, M.A., Blissard, G.W., and McCray, P.B., Jr. (2005). Persistent gene expression in mouse nasal epithelia following feline immunodeficiency virus-based vector gene transfer. *J Virol* 79, 12818-12827.
- Sinn, P.L., Hickey, M.A., Staber, P.D., Dylla, D.E., Jeffers, S.A., Davidson, B.L., Sanders, D.A., and McCray, P.B., Jr. (2003). Lentivirus vectors pseudotyped with filoviral envelope glycoproteins transduce airway epithelia from the apical surface independently of folate receptor alpha. *J Virol* 77, 5902-5910.
- Stapleton, C.M., Jaradat, M., Dixon, D., Kang, H.S., Kim, S.-C., Liao, G., Carey, M.A., Cristiano, J., Moorman, M.P., and Jetten, A.M. (2005). Enhanced susceptibility of staggerer (RORalpha±sg/sg) mice to lipopolysaccharide-induced lung inflammation. *American Journal of Physiology - Lung Cellular and Molecular Physiology* 289, L144-L152.
- Stehle, I., Scinteie, M., Baiker, A., Jenke, A.W., and Lipps, H. (2003). Exploiting a minimal system to study the epigenetic control of DNA replication: the interplay between transcription and replication. *Chromosome Res* 11, 413-421.

- Stinski, M.F., and Roehr, T.J. (1985). Activation of the major immediate early gene of human cytomegalovirus by cis-acting elements in the promoter-regulatory sequence and by virus-specific trans-acting components. *J Virol* **55**, 431-441.
- Stonebraker, J.R., Wagner, D., Lefensty, R.W., Burns, K., Gendler, S.J., Bergelson, J.M., Boucher, R.C., O'Neal, W.K., and Pickles, R.J. (2004). Glycocalyx restricts adenoviral vector access to apical receptors expressed on respiratory epithelium in vitro and in vivo: role for tethered mucins as barriers to luminal infection. *Journal of virology* **78**, 13755-13768.
- Stoner, G.D., Hallman, M., and Troxell, M.C. (1978). Lecithin biosynthesis in a clonal line of lung adenoma cells with type II alveolar cell properties. *Exp Mol Pathol* **29**, 102-114.
- Stoner, G.D., Kikkawa, Y., Kniazeff, A.J., Miyai, K., and Wagner, R.M. (1975). Clonal isolation of epithelial cells from mouse lung adenoma. *Cancer Res* **35**, 2177-2185.
- Stormo, G.D. (2000). DNA binding sites: representation and discovery. *Bioinformatics* **16**, 16-23.
- Sumner-Jones, S.G., and Gill, D.R. (2014). Cystic Fibrosis: Gene Therapy. In eLS (John Wiley & Sons, Ltd).
- Sun, X., Sui, H., Fisher, J.T., Yan, Z., Liu, X., Cho, H.-J., Joo, N.S., Zhang, Y., Zhou, W., Yi, Y., *et al.* (2010). Disease phenotype of a ferret CFTR-knockout model of cystic fibrosis. *The Journal of Clinical Investigation* **120**, 3149-3160.
- Tannous, B.A. (2009). Gaussia luciferase reporter assay for monitoring biological processes in culture and in vivo. *Nat. Protocols* **4**, 582-591.
- Tessarz, P., and Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol* **15**, 703-708.
- Thornhill, S.I., Schambach, A., Howe, S.J., Ulaganathan, M., Grassman, E., Williams, D., Schiedlmeier, B., Sebire, N.J., Gaspar, H.B., Kinnon, C., *et al.* (2008). Self-inactivating Gammaretroviral Vectors for Gene Therapy of X-linked Severe Combined Immunodeficiency. *Mol Ther* **16**, 590-598.
- Thyagarajan, B., Olivares, E.C., Hollis, R.P., Ginsburg, D.S., and Calos, M.P. (2001). Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol Cell Biol* **21**, 3926-3934.
- Trojer, P., and Reinberg, D. (2007). Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Molecular Cell* **28**, 1-13.
- Tros de Ilarduya, C., Sun, Y., and Duzgunes, N. (2010). Gene delivery by lipoplexes and polyplexes. *European Journal of Pharmaceutical Sciences* **40**, 159-170.
- Uetsuki, T., Naito, A., Nagata, S., and Kaziro, Y. (1989). Isolation and characterization of the human chromosomal gene for polypeptide chain elongation factor-1 alpha. *Journal of Biological Chemistry* **264**, 5791-5798.
- Ura, T., Okuda, K., and Shimada, M. (2014). Developments in Viral Vector-Based Vaccines. *Vaccines* **2**, 624-641.

- van Gaal, E.V.B., van Eijk, R., Oosting, R.S., Kok, R.J., Hennink, W.E., Crommelin, D.J.A., and Mastrobattista, E. (2011). How to screen non-viral gene delivery systems in vitro? *J Control Release* 154, 218-232.
- Van Goor, F., Hadida, S., Grootenhuys, P., Stack, J., Burton, B., Olson, E., Wine, J., Frizzell, R., Ashlock, M., and Negulescu, P. (2010). Rescue of the protein folding defect in cystic fibrosis in vitro by the investigational small molecule, VX-809. *Journal of Cystic Fibrosis*.
- Vandermeulen, G., Marie, C., Scherman, D., and Preat, V. (2011). New Generation of Plasmid Backbones Devoid of Antibiotic Resistance Marker for Gene Therapy Trials. *Mol Ther* 19, 1942-1949.
- Vargas, J.J., Gusella, G., Najfeld, V., Klotman, M., and Cara, A. (2004). Novel integrase-defective lentiviral episomal vectors for gene transfer. *Hum Gene Ther* 15, 361-372.
- Vasu, V.T., Cross, C.E., and Gohil, K. (2009). Nr1d1, an Important Circadian Pathway Regulatory Gene, Is Suppressed By Cigarette Smoke in Murine Lungs. *Integrative Cancer Therapies* 8, 321-328.
- Vaysse, L., Gregory, L.G., Harbottle, R.P., Perouzel, E., Tolmachov, O., and Coutelle, C. (2006). Nuclear-targeted minicircle to enhance gene transfer with non-viral vectors in vitro and in vivo. *The Journal of Gene Medicine* 8, 754-763.
- Verkman, A.S. (2001). Lung disease in cystic fibrosis: is airway surface liquid composition abnormal? *American Journal of Physiology - Lung Cellular and Molecular Physiology* 281, L306-L308.
- Verkman, A.S., Song, Y., and Thiagarajah, J.R. (2003). Role of airway surface liquid and submucosal glands in cystic fibrosis lung disease. *Am J Physiol Cell Physiol* 284, C2-15.
- Viecelli, H.M., Harbottle, R.P., Wong, S.P., Schlegel, A., Chuah, M.K., VandenDriessche, T., Harding, C.O., and Thony, B. (2014). Treatment of phenylketonuria using minicircle-based naked-DNA gene transfer to murine liver. *Hepatology* 60, 1035-1043.
- Virella-Lowell, I., Poirier, A., Chesnut, K., Brantly, M., and Flotte, T. (2000). Inhibition of recombinant adeno-associated virus (rAAV) transduction by bronchial secretions from cystic fibrosis patients. *Gene Ther.* 7, 1783-1789.
- Wagner, J.A., Messner, A.H., Moran, M.L., Daifuku, R., Kouyama, K., Desch, J.K., Manley, S., Norbash, A.M., Conrad, C.K., and Friborg, S. (1999). Safety and Biological Efficacy of an Adeno-Associated Virus Vector,-Cystic Fibrosis Transmembrane Regulator (AAV-CFTR) in the Cystic Fibrosis Maxillary Sinus. *The Laryngoscope* 109, 266-274.
- Wagner, J.A., Reynolds, T., Moran, M.L., Moss, R.B., Wine, J.J., Flotte, T.R., and Gardner, P. (1998). Efficient and persistent gene transfer of AAV-CFTR in maxillary sinus. *The Lancet* 351, 1702-1703.
- Wan, H., Dingle, S., Xu, Y., Besnard, V.r., Kaestner, K.H., Ang, S.-L., Wert, S., Stahlman, M.T., and Whitsett, J.A. (2005). Compensatory Roles of Foxa1 and Foxa2 during Lung Morphogenesis. *Journal of Biological Chemistry* 280, 13809-13816.
- Wang, G., Zabner, J., Deering, C., Launspach, J., Shao, J., Bodner, M., Jolly, D.J., Davidson, B.L., and McCray Jr, P.B. (2000). Increasing epithelial junction permeability enhances gene transfer to airway epithelia in vivo. *American journal of respiratory cell and molecular biology* 22, 129-138.

- Warming, S., Suzuki, T., Yamaguchi, T.P., Jenkins, N.A., and Copeland, N.G. (2004). Early B-cell factor-associated zinc-finger gene is a frequent target of retroviral integration in murine B-cell lymphomas. *Oncogene* 23, 2727-2731.
- Whitcutt, M.J., Adler, K.B., and Wu, R. (1988). A biphasic chamber system for maintaining polarity of differentiation of cultured respiratory tract epithelial cells. *In Vitro Cell Dev Biol* 24, 420-428.
- Whitfield, T.W., Wang, J., Collins, P.J., Partridge, E.C., Aldred, S.F., Trinklein, N.D., Myers, R.M., and Weng, Z. (2012). Functional analysis of transcription factor binding sites in human promoters. *Genome Biol* 13, R50.
- Wilke, M., Buijs-Offerman, R.M., Aarbiou, J., Colledge, W.H., Sheppard, D.N., Touqui, L., Bot, A., Jorna, H., de Jonge, H.R., and Scholte, B.J. (2011). Mouse models of cystic fibrosis: phenotypic analysis and research applications. *J Cyst Fibros* 10 Suppl 2, S152-171.
- Wilson, M.H., Coates, C.J., and George, A.L. (2007). PiggyBac Transposon-mediated Gene Transfer in Human Cells. *Mol Ther* 15, 139-145.
- Wiseman, J.W., Goddard, C.A., McLelland, D., and Colledge, W.H. (2003). A comparison of linear and branched polyethylenimine (PEI) with DCChol/DOPE liposomes for gene delivery to epithelial cells in vitro and in vivo. *Gene Ther* 10, 1654-1662.
- Wolff, L.J., Wolff, J.A., and Sebestyen, M.G. (2009). Effect of tissue-specific promoters and microRNA recognition elements on stability of transgene expression after hydrodynamic naked plasmid DNA delivery. *Hum Gene Ther* 20, 374-388.
- Wongrakpanich, A., Adamcakova-Dodd, A., Xie, W., Joshi, V.B., Mapuskar, K.A., Geary, S.M., Spitz, D.R., Thorne, P.S., and Salem, A.K. (2014). The Absence of CpG in Plasmid DNA-Chitosan Polyplexes Enhances Transfection Efficiencies and Reduces Inflammatory Responses in Murine Lungs. *Mol Pharm*.
- Wooddell, C.I., Reppen, T., Wolff, J.A., and Herweijer, H. (2008). Sustained liver-specific transgene expression from the albumin promoter in mice following hydrodynamic plasmid DNA delivery. *J Gene Med* 10, 551-563.
- Wooddell, C.I., Subbotin, V.M., Sebestyen, M.G., Griffin, J.B., Zhang, G., Schleef, M., Braun, S., Huss, T., and Wolff, J.A. (2010). Muscle damage after delivery of naked plasmid DNA into skeletal muscles is batch dependent. *Hum Gene Ther* 22, 225-235.
- Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C.L., Haase, J., Janes, J., Huss, J.W., 3rd, *et al.* (2009). BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* 10, R130.
- Wu, F., Levchenko, I., and Filutowicz, M. (1995). A DNA segment conferring stable maintenance on R6K gamma-origin core replicons. *J Bacteriol* 177, 6338-6345.
- <http://www.cysticfibrosis.org.uk> UK CF Registry, Annual data report 2010. Last modified: 2011 December 14 [cited: 2014 December 11] Available from: URL [https://www.cysticfibrosis.org.uk/media/108230/CR\\_Annual\\_Data\\_Report\\_2010\\_Dec\\_11.pdf](https://www.cysticfibrosis.org.uk/media/108230/CR_Annual_Data_Report_2010_Dec_11.pdf).
- <http://www.ema.europa.eu> (2001). The European Agency for the Evaluation of Medicinal Products. Note for Guidance on the Quality, Preclinical and Clinical Aspects of Gene Transfer Medicinal Products. Last modified: 2001 April 24 [cited: 2015 April 03] Available from: URL

- [http://www.ema.europa.eu/ema/pages/includes/document/open\\_document.jsp?webContentId=WC500003987](http://www.ema.europa.eu/ema/pages/includes/document/open_document.jsp?webContentId=WC500003987).
- Xenariou, S., Griesenbach, U., Ferrari, S., Dean, P., Scheule, R.K., Cheng, S.H., Geddes, D.M., Plank, C., and Alton, E.W.F.W. (2006). Using magnetic forces to enhance non-viral gene transfer to airway epithelium in vivo. *Gene Ther.* *13*, 1545-1552.
- Xenariou, S., Griesenbach, U., Liang, H.D., Zhu, J., Farley, R., Somerton, L., Singh, C., Jeffery, P.K., Ferrari, S., Scheule, R.K., *et al.* (2007). Use of ultrasound to enhance nonviral lung gene transfer in vivo. *Gene Ther.* *14*, 768-774.
- Xenariou, S., Liang, H.D., Griesenbach, U., Zhu, J., Farley, R., Somerton, L., Singh, C., Jeffery, P.K., Scheule, R.K., Cheng, S.H., *et al.* (2010). Low-frequency ultrasound increases non-viral gene transfer to the mouse lung. *Acta Bioch Bioph Sin* *42*, 45-51.
- Xu, M., and Cook, P.R. (2008). The role of specialized transcription factories in chromosome pairing. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* *1783*, 2155-2160.
- Yamaya, M., Finkbeiner, W.E., Chun, S.Y., and Widdicombe, J.H. (1992). Differentiated structure and function of cultures from human tracheal epithelium. *Am J Physiol* *262*, L713-724.
- Yan, Z., Stewart, Z.A., Sinn, P.L., Olsen, J.C., Hu, J., McCray, P.B., Jr., and Engelhardt, J.F. (2015a). Ferret and pig models of cystic fibrosis: prospects and promise for gene therapy. *Hum Gene Ther Clin Dev* *26*, 38-49.
- Yan, Z., Sun, X., Feng, Z., Li, G., Fisher, J., Stewart, Z., and Engelhardt, J. (2015b). Optimization of rAAV-mediated Expression for Large Transgenes Using a Synthetic Promoter and Tandem Array Enhancers. *Hum Gene Ther.*
- Yang, Y., Li, Q., Ertl, H., and Wilson, J.M. (1995). Cellular and humoral immune responses to viral antigens create barriers to lung-directed gene therapy with recombinant adenoviruses. *Journal of virology* *69*, 2004-2015.
- Yean, D., and Gralla, J. (1997). Transcription reinitiation rate: a special role for the TATA box. *Mol Cell Biol* *17*, 3809-3816.
- Yew, N.S., Marshall, J., Przybylska, M., Wysokenski, D.M., Ziegler, R.J., Rafter, P.W., Li, C., Armentano, D., and Cheng, S.H. (1999a). Increased duration of transgene expression in the lung with plasmid DNA vectors harboring adenovirus E4 open reading frame 3. *Hum Gene Ther* *10*, 1833-1843.
- Yew, N.S., Przybylska, M., Ziegler, R.J., Liu, D.P., and Cheng, S.H. (2001). High and sustained transgene expression in vivo from plasmid vectors containing a hybrid ubiquitin promoter. *Molecular Therapy* *4*, 75-82.
- Yew, N.S., Wang, K.X., Przybylska, M., Bagley, R.G., Stedman, M., Marshall, J., Scheule, R.K., and Cheng, S.H. (1999b). Contribution of plasmid DNA to inflammation in the lung after administration of cationic lipid:pDNA complexes. *Hum Gene Ther* *10*, 223-234.
- Yew, N.S., Zhao, H., Przybylska, M., Wu, I.H., Tousignant, J.D., Scheule, R.K., and Cheng, S.H. (2002). CpG-depleted plasmid DNA vectors with enhanced safety and long-term gene expression in vivo. *Mol Ther* *5*, 731-738.

- Yew, N.S., Zhao, H., Wu, I.H., Song, A., Tousignant, J.D., Przybylska, M., and Cheng, S.H. (2000). Reduced inflammatory response to plasmid DNA vectors by elimination and inhibition of immunostimulatory CpG motifs. *Mol Ther* 1, 255-262.
- Yu, H., Burton, B., Huang, C.-J., Worley, J., Cao, D., Johnson Jr, J.P., Urrutia, A., Joubran, J., Seepersaud, S., Sussky, K., *et al.* (2012). Ivacaftor potentiation of multiple CFTR channels with gating mutations. *Journal of Cystic Fibrosis* 11, 237-245.
- Zabner, J., Cheng, S.H., Meeker, D., Launspach, J., Balfour, R., Perricone, M.A., Morris, J.E., Marshall, J., Fasbender, A., Smith, A.E., *et al.* (1997). Comparison of DNA-lipid complexes and DNA alone for gene transfer to cystic fibrosis airway epithelia in vivo. *Journal of Clinical Investigation* 100, 1529-1537.
- Zabner, J., Couture, L.A., Gregory, R.J., Graham, S.M., Smith, A.E., and Welsh, M.J. (1993). Adenovirus-mediated gene transfer transiently corrects the chloride transport defect in nasal epithelia of patients with cystic fibrosis. *Cell* 75, 207-216.
- Zabner, J., Karp, P., Seiler, M., Phillips, S.L., Mitchell, C.J., Saavedra, M., Welsh, M., and Klingelutz, A.J. (2003). Development of cystic fibrosis and noncystic fibrosis airway cell lines. *American Journal of Physiology - Lung Cellular and Molecular Physiology* 284, L844-L854.
- Zhang, G., Budker, V., and Wolff, J.A. (1999). High levels of foreign gene expression in hepatocytes after tail vein injections of naked plasmid DNA. *Hum Gene Ther* 10, 1735-1737.
- Zhang, L., Wang, D., Fischer, H., Fan, P.-d., Widdicombe, J.H., Kan, Y.W., and Dong, J.-y. (1998). Efficient expression of CFTR function with adeno-associated virus vectors that carry shortened CFTR genes. *Proceedings of the National Academy of Sciences* 95, 10158-10163.
- Zhang, L.N., Karp, P., Gerard, C.J., Pastor, E., Laux, D., Munson, K., Yan, Z., Liu, X., Godwin, S., and Thomas, C.P. (2004). Dual therapeutic utility of proteasome modulating agents for pharmaco-gene therapy of the cystic fibrosis airway. *Molecular Therapy* 10, 990-1002.
- Zhang, Y., Stabernack, C.R., Dutton, R., Sonner, J., Trudell, J.R., Mihic, S.J., Yamakura, T., Harris, R.A., Gong, D., and Eger, E.I. (2001). Luciferase as a model for the site of inhaled anesthetic action. *Anesth Analg* 93, 1246-1252.
- Zhao, F.-Q., Zheng, Y., Dong, B., and Oka, T. (2004a). Cloning, genomic organization, expression, and effect on beta-casein promoter activity of a novel isoform of the mouse Oct-1 transcription factor. *Gene* 326, 175-187.
- Zhao, H., Doyle, T.C., Coquoz, O., Kalish, F., Rice, B.W., and Contag, C.H. (2005). Emission spectra of bioluminescent reporters and interaction with mammalian tissue determine the sensitivity of detection in vivo. *J Biomed Opt* 10, 41210.
- Zhao, H., Hemmi, H., Akira, S., Cheng, S.H., Scheule, R.K., and Yew, N.S. (2004b). Contribution of Toll-like Receptor 9 Signaling to the Acute Inflammatory Response to Nonviral Vectors. *Mol Ther* 9, 241-248.
- Zhou, L., Dey, C.R., Wert, S.E., Duvall, M.D., Frizzell, R.A., and Whitsett, J.A. (1994). Correction of lethal intestinal defect in a mouse model of cystic fibrosis by human CFTR. *Science* 266, 1705-1708.

Ziady, A.-G., Gedeon, C.R., Miller, T., Quan, W., Payne, J.M., Hyatt, S.L., Fink, T.L., Muhammad, O., Oette, S., and Kowalczyk, T. (2003a). Transfection of airway epithelium by stable PEGylated poly-L-lysine DNA nanoparticles in vivo. *Molecular Therapy* 8, 936-947.

Ziady, A.-G., Gedeon, C.R., Muhammad, O., Stillwell, V., Oette, S.M., Fink, T.L., Quan, W., Kowalczyk, T.H., Hyatt, S.L., and Payne, J. (2003b). Minimal toxicity of stabilized compacted DNA nanoparticles in the murine lung. *Molecular Therapy* 8, 948-956.