Gene Therapy for Surfactant Protein B Deficiency using Recombinant AAV Vectors

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



Helena Christiana Maria Meyer-Berg

Gene Medicine Research Group
Radcliffe Department of Medicine

and

St Cross College
University of Oxford

Supervisors: Prof. Deborah R. Gill and Prof. Stephen C. Hyde

Trinity Term 2021

Gene Therapy for Surfactant Protein B Deficiency using Recombinant AAV Vectors

Helena Christiana Maria Meyer-Berg

Radcliffe Department of Medicine and St Cross College, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy, Trinity Term 2021

Abstract

Gene therapies are being developed for rare genetic diseases and offer hope for conditions with few treatment options. Surfactant protein B (SP-B) deficiency is a rare genetic disease of the lung causing severe respiratory distress in neonates and is fatal within the first months of life. Recombinant adeno-associated virus (rAAV) is being investigated for delivery of functional SP-B to the lungs of SP-B deficient neonates, to produce surfactant for secretion onto the lung surface to increase chances of survival.

To screen AAV serotypes for efficient targeting of alveolar type II (ATII) cells of the lung parenchyma, rAAV5, rAAV6.2 and rAAV9 expressing reporter transgenes were delivered to murine lungs via intranasal instillation. Serotype 5 mediated the highest transduction levels in the target ATII cells (19.4 %, $p < 0.05$ versus naive). The *in situ* hybridisation method RNAscope was adapted to include a DNase digestion step to allow for detection of transgene mRNA expression after rAAV delivery and showed that the hCEFI promoter mediated the highest expression levels in ATII cells (11.1 transgene mRNA copies per cell,

$p < 0.05$ versus rAAV5.SP-B MG and $p < 0.0001$ versus all other groups). Therefore, rAAV5.hCEFI.SP-B was the rAAV vector selected for use in the murine lung. To validate the vector for human lung delivery, eight AAV capsids were screened in human models of the lung parenchyma: precision cut lung slices (PCLS) generated from lung resections and lung bud organoids (LBO) generated from human embryonic stem cells. The rAAV5 vector did not transduce the PCLS model, or the LBO after microinjection to the lumen; but vectors rAAV6, rAAV6.2 and rAAV6.2FF showed distinct levels of transduction in both models ($p < 0.005$ versus naive), indicating their suitability for gene delivery to the human lung parenchyma and highlighting the limitations of the murine model for AAV serotype selection in lung. Reporter gene expression using the CMV promoter was more robust than the hCEFI promoter in both the PCLS (hCEFI expression not detected) and LBO model (about 5-fold), but hCEFI mediated distinct levels of expression long-term. To test the capability of rAAV6.2 to mediate therapeutic protein expression in a human model of the SP-B deficient lung, embryonic stem cells were gene edited to incorporate the most common disease-causing 121ins2 mutation in SP-B deficiency. Subsequently, LBOs were generated from this edited cell line by sequential differentiation and delivery of rAAV6.2.CMV.SP-B resulted in expression of mature SP-B protein, highlighting the translational potential of the selected rAAV vector.

Graphic abstract

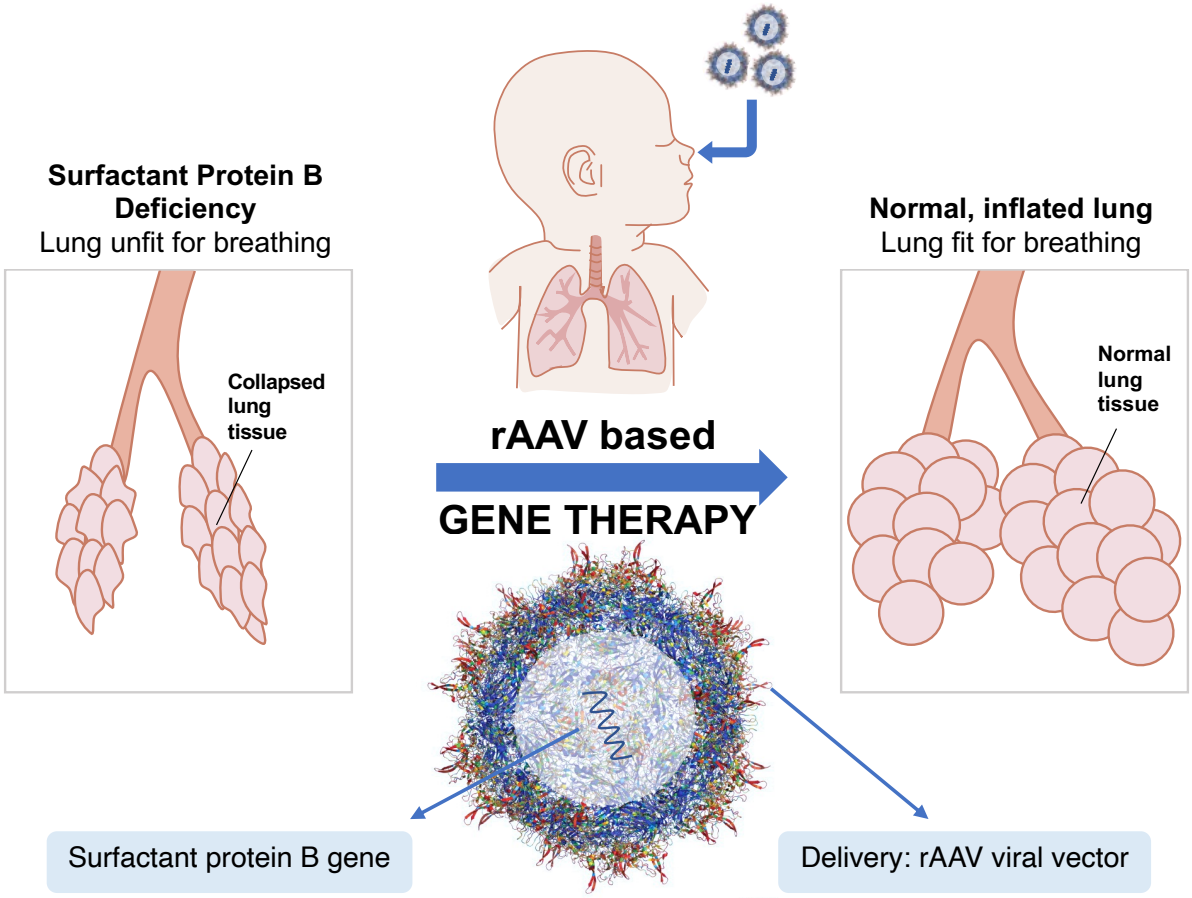


Figure partly modified from http://www2.hawaii.edu/~yzuo/research1-surfactant.html and Eric Zinn, Luk Vandenberghe: Adeno-associated virus: fit to serve, Current Opinion in Virology, 2014 (AAV capsid).

Acknowledgements

First and foremost, I wish to thank my DPhil supervisors, Prof. Deborah Gill and Prof. Steve Hyde for their continuous support, wisdom, and the freedom to explore more and less fruitful ideas. These past years were truly formative, taught me a lot about science, an exciting field, and myself – and that is in no small part because of you.

Thank you for the support, enthusiasm and discussions, but also the giggles, and all the shared cake and doughnuts to all the past and present members of the gene medicine group: Dr Ian Pringle, Dr Jack Tan, Dr Agata Antepowicz, Dr Joost van Haasteren, Dr Altar Munis, Dr Toby Gamlen, Dr Kamran Miah, Dr Yue Du, Dr Rosie Munday, Omar Habib, Eoin Mac Réamoinn, Aimee Ruffle, Wian Satyapertiwi, Mariana Viegas, Rebecca Dean, Catriona Conway, Rachel Ashworth, and last but definitely not least, Tina Garland. I further would like to thank Dr Alberto Zambrano, Dr Maria Pilar de Lucas and Lucia Zhou Yang, for welcoming me at ISCIII, Madrid as a collaborator, support of the lung organoid studies, as well as training me in their generation. Special thanks to Dr Esme Magro-Lopez for joining us in Oxford to support the generation of gene editing protocols in embryonic stem cells. Further, I wish to thank my thesis committee Dr Philip Hublitz and Prof. Mona Bafadhel for productive discussions.

Thank you to all the moat house inhabitants for providing me with a home and family in Oxford – Marianne, Danielle, Pernille, Brooke, Sam, Moritz, Federico. Thank you, Shamir, for sharing your science-, trivia- and food-appreciation. Thank you, Selena, for your friendship and shaking tough days off in the gym with me.

I am deeply grateful to my family, for their love and encouragement, particularly to my parents for always believing in me. Alexander, thank you for being my partner through all the adventures in life.

Funding

This thesis was supported by a Scatcherd European Scholarship, German National Academic Foundation (*Studienstiftung des deutschen Volkes*) PhD Scholarship as well as a Medical Research Council (MRC) DPhil Scholarship. Training in organoid generation was supported by MRC supplementary funding.

Declaration

This thesis has been composed by myself and has not been used in any previous application for a degree. The results presented were obtained by myself and contributions made by others have been acknowledged within the figure legends or text as appropriate. All sources of published or unpublished information have been specifically acknowledged by means of a reference or personal communication notation, respectively.

List of abbreviations

AAP	assembly-activating protein
AAT	alpha-1 antitrypsin
AAV	adeno-associated virus
AAVR	AAV receptor
ABCA3	ATP binding cassette subfamily A member 3
AdV	adenovirus
ATII cell	alveolar type II cell
ATP	adenosine triphosphate
BALF	bronchoalveolar lavage fluid
BBB	blood brain barrier
BSA	bovine serum albumin
ca.	circa
CAG	CMV enhancer element, chicken beta-actin promoter, splice acceptor of rabbit beta-globin
Cas9	CRISPR-associated protein 9
CASI	CMV enhancer, chicken β-actin promoter, UBC enhancer with splice donor and acceptor sequences
CBA	chicken beta-actin
cDNA	complimentary DNA
CFTR	cystic fibrosis transmembrane conductance regulator
CMV	cytomegalovirus
CMVie	CMV immediate early promoter
CNS	central nervous system
CpG	cytosine-guanosine dinucleotide
CRISPR	clustered regularly interspaced short palindromic repeat
DMD	Duchenne muscular dystrophy
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulfoxide
DNA	desoxyribonucleic acid
DPPC	dipalmitoylphosphatidylcholine
DSB	double strand break
dsDNA	double stranded DNA
EDTA	ethylenediaminetetraacetic acid
EF1a	elongation factor 1 a
EGFP	enhanced green fluorescent protein
EGFR	epidermal growth factor receptor
ELISA	enzyme-linked immunosorbent assay
EMA	European Medicines Agency
ER	endoplasmic reticulum
ESC	embryonic stem cells
FBS	fetal bovine serum
FDA	Food and Drug Administration

FGFR1	fibroblast growth factor receptor 1
GC	genome copies
GFP	green fluorescent protein
HAE	human airway epithelial culture
hCEF	CMV enhancer element with EF1a promoter
hCEFI	CMV enhancer element with EF1a promoter and CpG free Intron
HGFR	hepatocyte growth factor receptor
HITI	homology-independent targeted integration
HIV	human immunodeficiency virus
HPRE	hepatitis B virus post-transcriptional regulatory element
HS	heparan sulphate
ICC	immunocytochemistry
ICE	interference of CRISPR edits
IHC	immunohistochemistry
IM	intramuscular
IMS	industrial methylated spirit
IPF	idiopathic pulmonary fibrosis
iPSC	induced pluripotent stem cells
ISH	*in situ* hybridisation
ITR	inverted terminal repeat
IV	intravenous
kDa	kilodalton
LB	lamellar body
LB broth	Luria Bertani broth
LBO	lung bud organoid
LD stain	live/dead stain
LTR	long terminal repeat
MAA I	*Maackia amurensis* lectin I
MAA II	*Maackia amurensis* lectin II
mAAP	membrane associated accessory protein
MES	2-ethanesulfonic acid
miRNA	microRNA
mRNA	messenger RNA
MVB	multivesicular body
NHEJ	non-homologous end joining
NHP	non-human primate
OCT	optimal cutting temperature
PAP	pulmonary alveolar proteinosis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDGFR	platelet-derived growth factor receptor
PEI	polyethylenimine
Pen/Strep	penicillin/streptomycin
PFA	paraformaldehyde
PRE	post-transcriptional regulatory element

PSC	pluripotent stem cells
qPCR	quantitative PCR
rAAV	recombinant adeno-associated virus
rAdV	recombinant adenovirus
RDS	respiratory distress syndrome
rLV	recombinant lentivirus
RNA	ribonucleic acid
RNAi	RNA interference
RNP	ribonucleoprotein
ROCK	Rho-associated, coiled-coil containing protein kinase
rpm	rounds per minute
RT	room temperature
rtTA	reverse tetracycline transactivator
SA	sialic acid
SCID-X1	X-linked severe combined immunodeficiency
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
sgRNA	single guide RNA
SIN	self-inactivating
SIV	simian human immunodeficiency virus
SMA	spinal muscular atrophy
SNA	*Sambucus nigra* agglutinin
SNP	single nucleotide polymorphism
SP-A	surfactant protein A
SP-B	surfactant protein B
SP-C	surfactant protein C
SP-D	surfactant protein D
ssDNA	single strand DNA
T-X-100	Triton-X-100
TALEN	transcription activator-like effector nucleases
TBS	Tris buffered saline
TBS-T	1x TBS with 0.05 % Tween
TFF	tangential flow filtration
TIDE	tracking indels by decomposition
Tris	Tris(hydroxymethyl)aminomethane
TU	transducing units
UBC	ubiquitin C
WPRE	woodchuck hepatitis virus post-transcriptional regulatory element
ZNF	zinc finger nucleases

Table of contents

Abstract ..... *I*

Graphic abstract ..... *III*

Acknowledgements ..... *IV*

Funding ..... *V*

Declaration ..... *VI*

List of abbreviations ..... *VII*

Table of contents ..... *X*

List of figures ..... *XV*

List of tables ..... *XVIII*

1 Chapter 1 – Introduction 1

1.1 Gene therapy for rare diseases of the lung 1

1.2 Surfactant Protein B deficiency 3

1.2.1 The lung and its parenchyma 3

1.2.2 Pulmonary surfactant 5

1.2.2.1 Surfactant composition 5

1.2.2.2 Surfactant homeostasis 6

1.2.3 Surfactant Protein B biosynthesis 10

1.2.4 Genetic disorders of surfactant protein B 14

1.3 Gene therapy – principles of design 16

1.3.1 Gene addition approach 17

1.3.1.1 Transgene cassette design 17

1.3.1.2 Transgene optimisation 20

1.3.2 Gene editing approach 22

1.3.3 Viral and non-viral delivery strategies for gene therapy products 24

1.4 The lung as target for gene therapy 26

1.4.1 Delivery of lung gene therapies 26

1.4.2 Obstacles for lung gene therapy 26

1.4.3 Clinical experience of lung gene therapy 27

1.5 Adeno-associated viral vector-based gene therapy 29

1.5.1 Adeno-associated virus biology 29

1.5.2 Recombinant adeno-associated viral vectors 31

1.5.3 Limitations of recombinant AAV-based gene therapy 35

1.5.4 AAV serotypes and recombinant AAV vector tropisms 37

1.5.5 AAV capsid engineering 46
1.6 Previous studies on gene therapy for SP-B deficiency 48
1.6.1 Studies *in vivo* 48
1.6.2 Studies in organoids 50
1.7 Aims 51
2 Chapter 2 – Materials and Methods 52
2.1 Common reagents and solutions 52
2.2 Plasmids and molecular cloning 52
2.2.1 Polymerase chain reaction 53
2.2.2 Restriction enzyme digest 53
2.2.3 Gel electrophoresis and extraction 53
2.2.4 Ligation 54
2.2.5 Plasmid amplification 54
2.2.6 Plasmid analytics 55
2.2.7 Plasmids in this study 55
2.3 Sanger sequencing 57
2.4 Mammalian Cell Culture 57
2.4.1 Lines and maintenance 57
2.4.2 Transfection 59
2.4.3 Creation of a CRISPR/Cas9-edited stem cell line 60
2.4.3.1 Electroporation of stem cells for Cas9 RNP delivery 60
2.4.3.2 Analysis of editing efficiency 61
2.4.3.3 Isolation of a gene edited stem cell clone 62
2.5 Viral vector production 63
2.5.1 Recombinant adeno-associated viral vector production 63
2.5.2 Recombinant lentiviral vector production 66
2.6 Animal studies 67
2.6.1 Maintenance 67
2.6.2 Intranasal dosing 68
2.6.3 Lung harvest 68
2.6.4 Bronchoalveolar lavage 69
2.7 Precision cut lung slices 69
2.7.1 Generation and culture of precision cut lung slices 69
2.7.2 Vector dosing to precision cut lung slices 70
2.8 Lung bud organoids 71
2.8.1 Generation and culture of lung bud organoids 71
2.8.2 Microinjection 71

2.8.3 Cas9 microinjection for gene editing in lung bud organoids 72
2.9 Preparation of fixed-frozen sections 72
2.10 Immuno- and lectin histochemistry 73
2.11 Immunocytochemistry 76
2.12 Western blot analysis 77
2.13 Flow cytometry 79
2.14 RNAscope 80
2.14.1 RNAscope pretreatment and assay 80
2.14.2 Nuclease pretreatment to uncouple RNA and DNA signal 81
2.15 Microscopy and image analysis 84
2.16 Statistical analysis 87
3 Chapter 3 – Selection of a candidate rAAV vector in the murine lung *in vivo* 88
3.1 Introduction 88
3.2 Results 90
3.2.1 Investigation of AAV serotypes suitable for transduction of the murine lung parenchyma *in vivo* 90
3.2.2 Design and validation of rAAV expression cassettes for SP-B expression and viral vector production 97
3.2.3 Development of RNAscope to quantify therapeutic transgene expression after rAAV delivery 104
3.2.4 Identification of promoters driving high SP-B mRNA expression levels in murine ATII cells 112
3.2.5 SP-B protein expression levels in murine BALF after rAAV delivery *in vivo* 124
3.3 Discussion 128
3.3.1 Selection of rAAV5 as a candidate serotype for transduction of murine ATII cells *in vivo* 128
3.3.2 Validation of rAAV SP-B expression cassettes for rAAV production 130
3.3.3 Identification of a promoter with high expression levels in murine ATII cells after rAAV transduction using a newly established *in situ* hybridisation method 132
4 Chapter 4 – Examination of rAAV gene therapy vectors in human lung models 136
4.1 Introduction 136
4.1.1 Selection of a lung organoid model to mimic rAAV dosing to the human lung 137
4.1.2 Precision cut lung slices as a human lung *ex vivo* model to study viral infection 142

4.2 Results 144
4.2.1 Precision cut lung slices for rAAV vector screening 144
4.2.1.1 Use of murine PCLS as an *ex vivo* model for rAAV transduction 144
4.2.1.2 Screening of AAV serotypes in a human PCLS 148
4.2.2 Human lung bud organoids as a model for rAAV vector screening 159
4.2.2.1 LBO model generation 159
4.2.2.2 Establishment of a method suitable for dosing to the LBO lumen 163
4.2.2.3 Identification of AAV serotypes suitable for transduction of LBOs 163
4.2.2.4 Long-term rAAV expression in the LBO model 168
4.2.2.5 Verification of the LBO's distal lung identity and characterisation of AAV entry receptors 171
4.2.3 Identification of LBOs as a model of SARS-CoV-2 entry 176
4.3 Discussion 178
4.3.1 Vector screening in a model of the human adult lung 178
4.3.2 Human LBOs as a model for rAAV screening in context of entry receptor expression 180
4.3.3 Complementarity of LBO and PCLS model 184
5 Chapter 5 – Gene therapy in a human disease model of SP-B deficiency 187
5.1 Introduction 187
5.2 Results 189
5.2.1 Establishment of a method for generation of a human ESC line edited in *SFTPB* 189
5.2.2 Generation of a 121ins2 gene edited human ESC line 196
5.2.3 Characterisation of 121ins2 edited human ESC clones 200
5.2.4 Generation of LBOs from the 121ins2 edited stem cell line and its parental line 204
5.2.5 Gene therapy to restore SP-B expression in SP-B deficient LBOs 210
5.3 Discussion 215
5.3.1 SP-B deficient LBOs as model for SP-B gene transfer 215
5.3.2 Gene editing in SP-B deficient LBOs carrying the 121ins2 mutation 217
6 Chapter 6 – Discussion 219
6.1 Translational potential of rAAV vectors for gene therapy for SP-B deficiency 219
6.1.1 AAV serotype 6.2 220
6.1.2 Promoters for therapeutic transgene expression in ATII cells 230
6.1.3 Investigation of rAAV gene therapy in the mouse model of SP-B deficiency 234

6.2 Challenges for translation of rAAV gene therapy for SP-B deficiency 235
6.2.1 Immune system interaction....235
6.2.2 Longevity of rAAV vector expression in the lung 239
6.2.2.1 Integrating lentivirus as an alternative vector choice 242
6.2.2.2 Gene editing approaches....242
6.3 Acceleration of gene therapy development....247
6.4 Conclusion....250
References....251

List of figures

Figure 1.1 Schematic of the human lung and its parenchyma 4

Figure 1.2 Model of SP-B function 9

Figure 1.3 Biosynthesis of mature SP-B 12

Figure 1.4 Schematic of the AAV genome 30

Figure 1.5 Schematic of rAAV production 34

Figure 1.6 Phylogenetic tree of AAV serotypes 1-13 39

Figure 1.7 Alignment of the VP1 sequences of AAV1-9 41

Figure 2.1 RNAscope detection of nucleic acid targets 83

Figure 3.1 Transduction in murine lungs after intranasal delivery of AAV serotypes 5, 9 and 6.2....91

Figure 3.2 Representative images of lung sections from naive mice and mice transduced with AAV serotypes 5, 9, and 6.2 95

Figure 3.3 Quantification of transduction rate in ATII cells after intranasal dosing of mice with rAAV5, rAAV9 and rAAV6.2 96

Figure 3.4 Schematic of rAAV expression cassettes for SP-B transgene delivery 100

Figure 3.5 Western Blot of A549 cells transfected with plasmids expressing SP-B and soSP-B transgenes....101

Figure 3.6 Production yields of rAAV5 vectors containing SP-B and soSP-B expression cassettes 103

Figure 3.7 RNAscope in murine lung sections of naive mice and mice dosed with rAAV5.CMV.SP-B....107

Figure 3.8 RNAscope in lung sections of rAAV-dosed mice using RNase and DNase digestion to understand the origin of RNAscope 111

Figure 3.9 Uncoupling of the WPRE mRNA expression signal from DNA in sections of mice dosed with rAAV vectors expressing SP-B under various promoters....115

Figure 3.10 Quantification of SP-B transgene mRNA expression in transduced ATII cells after transduction of mice with rAAV vectors expressing SP-B under various promoters 117

Figure 3.11 Quantification of SP-B transgene mRNA expression in transduced *Sftbp* positive cells after rAAV gene transfer 122

Figure 3.12 Percentage of transduced ATII cells in DNase treated and untreated murine lung sections....123

Figure 3.13 Western blot analysis of SP-B expression in murine BALF 126

Figure 4.1 Stages of human lung development 141

Figure 4.2 Transduction of murine PCLS with rAAV5 and rAAV9.....146

Figure 4.3 Viability of human PCLS from d3 to d14 after generation.....149

Figure 4.4 Images of human PCLS imaged *en face* for native EGFP fluorescence on day 4 after transduction with AAV serotypes.....151

Figure 4.5 Images of human PCLS transduced with various AAV serotypes or mock transduced.....153

Figure 4.6 Representative images of sections from mock transduced human PCLS stained for the ATII cell marker pro-SP-C 158

Figure 4.7 Key stages of LBO generation 161

Figure 4.8 Native EGFP fluorescence in LBOs on day 5 after viral vector transduction165

Figure 4.9 Quantification of EGFP brightness intensity in *en face* images of rAAV transduced or mock-injected LBOs 167

Figure 4.10 Long-term EGFP expression facilitated by CMV or hCEFI promoter after rAAV6.2 delivery to LBOs.....169

Figure 4.11 Immunohistochemistry for markers of alveolar type II cells in LBOs 172

Figure 4.12 Immunohistochemistry and lectin stain for AAV entry receptors.....175

Figure 4.13 Immunohistochemistry for entry receptors of SARS-CoV-2.....177

Figure 5.1 Schematic of editing strategies to create a human ESC line for generation of SP-B deficient LBOs.....191

Figure 5.2 Identification of sgRNAs for efficient *SFTPB* targeting in HEK293T cells.....192

Figure 5.3 Comparison of 4D nucleofector electroporation programmes for efficient RNP delivery in RUES2 cells 195

Figure 5.4 Clone screening via restriction fragment length analysis and subsequent confirmation of gene editing by Sanger sequencing.....199

Figure 5.5 Karyostat microarray analysis to confirm normal karyotype 201

Figure 5.6 Immunocytochemistry for stem cell markers of the RUES2 cell line and 121in2 edited RUES2 clone 10 203

Figure 5.7 Brightfield images of LBO generation key stages from RUES2 cells and 121ins2 edited RUES2.....205

Figure 5.8 Examination of definitive endoderm induction efficiency by flow cytometry for c-KIT and CXCR4 206

Figure 5.9 Verification of anterior-foregut endoderm induction by immunocytochemistry for FOXA2 207

Figure 5.10 Western blot analysis for SP-B expression in wildtype and gene edited LBOs as well as transfected control cells 209

Figure 5.11 Western blot analysis for SP-B expression in 121ins2 edited LBO after rAAV6.2.CMV.SP-B microinjection 212

Figure 5.12 ICE analysis for therapeutic gene editing in 121ins2 edited LBOs 214

Figure 6.1 Schematic of CRISPR/Cas9 editing and DSB outcomes 245

List of tables

Table 1.1 Naturally occurring AAV serotypes, their receptors and tropisms....44

Table 2.1 Plasmids....56

Table 2.2 Cell lines....59

Table 2.3 Single guide RNAs for editing in *SFTPB*....61

Table 2.4 Primers for analysis of gene editing in *SFTPB*....63

Table 2.5 Iodixanol gradient preparation for one virus purification or 3 gradients....65

Table 2.6 Viral vectors used in this study....66

Table 2.7 Sequences of CRISPR/Cas9 gene editing tools to correct 121ins2 mutation in *SFTPB* exon 4....72

Table 2.8 Lectin histochemistry....74

Table 2.9 Antibodies for immunohistochemistry....75

Table 2.10 Antibodies for immunocytochemistry....77

Table 2.11 Antibodies for immunoblotting....79

Table 2.12 Antibodies for flow cytometry....80

Table 2.13 RNAscope probes....82

Table 2.14 ImageJ macro for channel subtraction....85

Table 2.15 ImageJ macro to count nuclei....85

Table 2.16 ImageJ macro to count transduced cells....86

Table 2.17 ImageJ macro to count RNAscope signal dots in selected image area....86

Table 4.1 Transduction level of viral vectors in human PCLS....157

Table 4.2 Comparison of AAV serotype dependent transduction level in human PCLS and LBOs....186

Table 6.1 Comparison of lung models for rAAV gene therapy investigation....229

Chapter 1 – Introduction

1.1 Gene therapy for rare diseases of the lung

Genetic mutations are the cause of about 72 % of the estimated 6000-10 000 rare diseases known today [1, 2]. A high proportion of rare diseases are classified as orphan diseases which means there are no adequate treatment options. Although every single rare disease is by definition rare (defined as affecting less than 1 in every 2000 people in the EU), cumulatively they are thought to affect more than 30 million individuals in the EU [3]. This corresponds to ca. 7 % of the EU's population, creating a large unmet need for novel treatments.

The lung is the organ with the irreplaceable function of gas exchange. Malfunction of the lung critically affects individuals' life quality with potentially fatal outcome. At an estimate, 1-2 million people in the EU are affected by a rare disease with primary or secondary lung involvement [4]. Within rare lung diseases, there are variable levels of understanding of pathogenesis, severity, prevalence and treatment options. For example, acquired pulmonary alveolar proteinosis (PAP) is a rare lung autoimmune disease causing severe respiratory distress, the pathogenesis of which is not fully understood (OMIM #300770). Idiopathic pulmonary fibrosis (IPF) is a chronic, disease of the lung, likewise with unclear cause (OMIM #178500). Lung fibrosis progressively hampers the gas exchange capacity of the lung. In contrast, the pathogenesis of most monogenic diseases is well understood, such as hereditary PAP caused by mutations in *CSF2RA* and *CSF2RB* (OMIM #300770 and #614370, respectively). Other well-characterised monogenic diseases are cystic fibrosis (OMIM #219700) and genetic diseases of pulmonary surfactant including mutations in *SFTPB*, *SFTPC* and *ABCA3* (OMIM #265120, #610913 and #610921, respectively).

Gene therapy is the introduction of genetic information into cells for therapeutic benefit. Gene therapy could be helpful for treating rare diseases of the lung with a known genetic target, for instance by the addition of genetic information for loss-of-function mutations. In contrast to symptomatic treatment, gene therapy has the potential for being curative because the underlying genetic condition is treated. Further, gene therapy has the potential for continuous production of therapeutic proteins, which could be advantageous in the lung, where disruption of homeostasis causes injury and fibrosis.

Despite several licenced gene therapy products [5], gene therapy is still a non-standard treatment due to the high costs associated and its novelty compared with other treatments. Therefore, preferred targets for gene therapy are diseases with clearly defined genetic targets, ideally involving only one gene as defined for monogenic diseases, that are without treatment options and severe enough to justify the costs and associated risks. For instance, cystic fibrosis is a genetic disease of the lung caused by mutations in the *CFTR* gene leading to thickened mucus which obstructs breathing and causes recurring lung infections. The predicted survival is approximately 30-40 years and the prevalence in the UK is 1:2500 [6]. Specific treatment is available for some, but not all known *CFTR* mutations. Therefore, cystic fibrosis may be considered as an attractive target for gene therapy, but the hurdles for safety and efficacy are quite high due to emerging availability of new drugs and increasing expected life span; however preparations for a gene therapy clinical trial with a product suitable for all mutations are underway [7]. For genetic diseases of pulmonary surfactant, however, there are no specific treatments available and life expectancy is shorter. Amongst them, Surfactant Protein B (SP-B) deficiency has been extensively researched. While ABCA3 and SP-C deficiencies have more variable outcomes, SP-B deficiency is fatal early in life creating urgency which makes even highly experimental treatments more acceptable after risk/benefit analyses.

SP-B deficiency is considered an ultrarare monogenic disease affecting 1 in 1 million live births with well-understood pathophysiology caused by mutations in *SFTPB* [8, 9]. The disease is fatal within the first months of life and traditional treatments such as surfactant replacement therapy were ineffective [10, 11]. Therefore, SP-B deficiency is an excellent target for gene therapy as well as a strategic starting point for developing gene therapies for the human lung, specifically for genetic diseases of pulmonary surfactant.

1.2 Surfactant Protein B deficiency

1.2.1 The lung and its parenchyma

The lung is the organ of gas exchange which can be functionally divided into airways and the gas-exchanging lung parenchyma (Figure 1.1). The lung faces a unique challenge: providing a large gas transmittable surface for gas exchange with the environmental gases, while maintaining the fluid balance of the gas exchanging cells and providing host defence against particles and pathogens. To maximise surface for gas exchange, the lung is organised into alveoli which form a mesh with the surface area of ca. 70 m^2 [12]. To keep the alveolar cells hydrated and provide a medium for immune cells, the alveolar space is lined by liquid. The resulting bioengineering problem is the surface tension at the air-liquid interface which generates collapsing forces at the end of exhalation and prevents respiration. This problem is solved by pulmonary surfactant which reduces surface tension [13, 14], such that the absence of functional pulmonary surfactant prevents respiration. For its function, pulmonary surfactant needs to fulfil three essential requirements: fast adsorption to the air-liquid interface to seal the liquid surface from the air, compressibility during expiration, and quick lateral redistribution during inhalation [15].

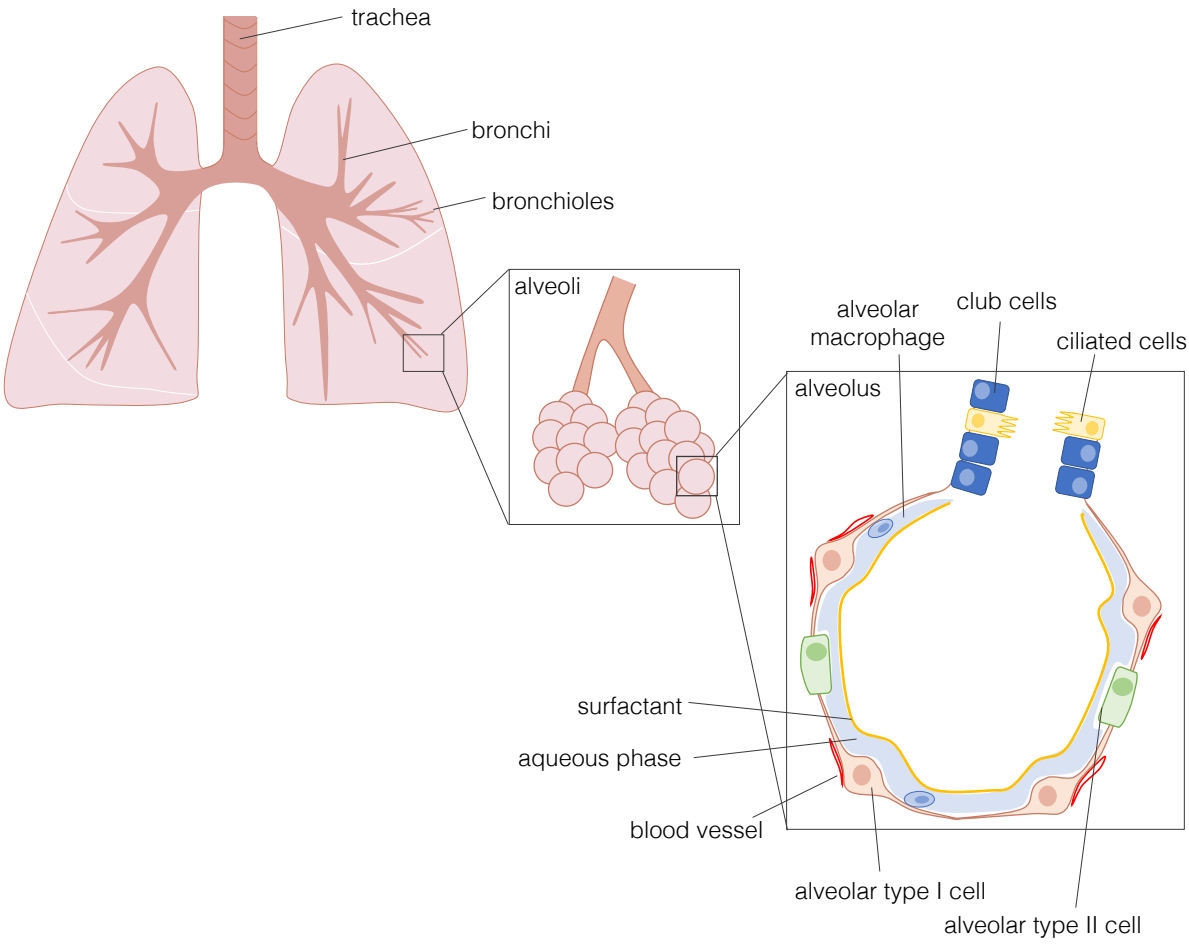


Figure 1.1 Schematic of the human lung and its parenchyma

The human lung is organised into five lung lobes. Air moves from the nasal cavity through progressively subdividing and narrowing airways, starting from the trachea, via the bronchi and bronchioles to the lung parenchyma. Air-sacs called alveoli are the structural units of the lung parenchyma and harbour the cuboidal surfactant-producing alveolar type II cells and alveolar type I cells that mediate gas-exchange of the blood. The aqueous phase of the alveoli provides a medium for immune cells such as alveolar macrophages. Pulmonary surfactant covers the aqueous phase to lower surface tension and prevent atelectasis. The alveoli are connected to the airways via the bronchioalveolar duct junction that contains airway cells such as club cells and ciliated cells.

1.2.2 Pulmonary surfactant

1.2.2.1 Surfactant composition

Pulmonary surfactant is a thin film consisting of protein and lipids lining the alveolar space. Surfactant is secreted by alveolar type II (ATII) cells in the distal lung and is mainly composed of lipids as well as proteins (ca. 90 % and 10 % of weight, respectively [16, 17]). The lipid component is predominantly composed of phospholipids that form bilayers, amongst other structures, in surfactant. The most common lipid is dipalmitoylphosphatidylcholine (DPPC) contributing ca. 40 % of the surfactant mass [16]. Because of its saturated acyl chains, DPPC can be tightly packed laterally in the lipid bilayer, which is crucial in surfactant compression during the exhalation-caused decrease of the alveolar surface [18]. Further, surfactant contains other saturated and unsaturated zwitterionic phosphatidylcholines, anionic phosphatidylglycerols and a small proportion of phosphatidylethanolamine and sphingomyelin [16, 18, 19]. Surfactant evolved to a ca. 50:50 mixture of saturated and unsaturated phospholipids [18] which is an atypical composition for cell membranes, but is relatively conserved among terrestrial animals [16, 20, 21]. The saturated versus unsaturated lipid ratio is crucial for surfactant function and is highly regulated. The saturated lipids provide stability especially during compression and unsaturated lipids provide the necessary fluidity of the surfactant [18]. Neutral lipids such as cholesterol further contribute to fluidity [16].

Aside from lipids, pulmonary surfactant contains four surfactant proteins (SP): the hydrophobic SP-B and SP-C and the hydrophilic SP-A and SP-D. The hydrophilic SPs are involved in innate immunity and belong to the collectin protein family. They bind and opsonise pathogens on the alveolar surface and modulate the inflammatory response by presenting their ‘prey’ to alveolar macrophages [22]. The hydrophobic surfactant

proteins SP-B and SP-C are associated with pulmonary surfactant function and SP-B is crucial in formation of the surface film and surfactant remodelling during the dynamics of breathing [13].

1.2.2.2 Surfactant homeostasis

The highly hydrophobic SP-B and SP-C are synthesised as large precursors together with surfactant lipids at/in the endoplasmic reticulum (ER). The SP-B and SP-C proteins travel via vesicular transport from the Golgi apparatus via multivesicular bodies (MVB) to the lamellar bodies (LB) [13]. The LB is the pulmonary surfactant storage organelle in ATII cells. The SP-B and SP-C proteins are processed during their journey through the cell organelles to their mature forms (SP-B discussed in detail in 1.2.3). Surfactant lipids are transported across the LB limiting membrane via ABCA3, an ATP-dependent transmembrane transporter essential for LB formation and its absence leads to abnormal or absent organelles [23, 24]. As LBs mature, they reach a densely packed and dehydrated state which is organised in a characteristic fashion [25]. SP-B is necessary for achieving this tightly packed state by connecting adjacent bilayers [26] and absence of SP-B leads to disorganised, 'bubbly' LBs [27]. Therefore, it is thought that the highly packed and ordered state of LBs is necessary for fast and efficient adsorption to the surfactant film and the increase of entropy during unwinding of the LB is a promoting factor [25].

Surfactant is secreted into the alveolar aqueous phase by exocytosis of lamellar bodies. The primary physiological stimulus of LB exocytosis is the mechanical stretching of the alveoli during inhalation as sensed by ATI cells [28, 29]. Multiple other stimuli that also lead to LB exocytosis are integrated by signalling pathways that finally result in elevation of intracellular Ca^{2+} concentration. For instance, extracellular SP-B itself can trigger efficient exocytosis of LBs, potentially to provide a feedback-loop for surfactant lipid

secretion caused by free, surfactant membrane-dissociated SP-B [30]. Elevated Ca^{2+} concentration triggers fusion of LBs with the cell membrane, an ATP-dependent process involving convergence of the membranes by SNARE proteins.

After exocytosis, LBs undergo structural rearrangement that ultimately leads to adsorption to the surface film. Firstly, LBs can be directly adsorbed to the surface film. This process is thought to be facilitated by ring-shaped SP-B oligomers built from SP-B dimers ([31], Figure 1.2). Structural modelling suggests that SP-B oligomers can bridge two membranes by causing perturbation of lipids at the ring interface and channelling the lipids through the central pore ([32], Figure 1.2). This process is also thought to facilitate surfactant expansion and compression during the dynamics of breathing in the lung where the surfactant film needs to expand and shrink in time with every breath.

Alternatively, LBs can unpack in the aqueous sub-phase after exocytosis and form the tubular myelin intermediate where surfactant membranes are arranged in a lattice structure [17, 33]. SP-A supports the structure by connecting the corners of the square membrane lattices [34]. In SP-A deficient mice, tubular myelin is absent, but lung function remains normal [35]. These findings suggest that tubular myelin is not an essential intermediate for formation of a functional surface film.

Surfactant needs regular replacement, because it becomes inactivated in the airspace, e.g. by the highly oxidative environment. It is estimated that 10 % of the surfactant pool is recycled per hour [36]. The majority of surfactant is recycled in ATII cells which is supported by alveolar macrophages and is key for lung function as disruption leads to PAP and respiratory distress [8, 37]. In the alveolar space, there are multiple surfactant vesicles, some containing ready-to-use surfactant, other smaller vesicles containing consumed surfactant ready for recycling. The SP-C and D proteins are thought to be

involved in surfactant recycling by regulating the vesicle size of surfactant [38, 39]. Consistent with this role, absence of SP-D leads to accumulation of phospholipids in the alveolar space [40]. Surfactant is absorbed by ATII cells by both an SP-A-dependent, clathrin-mediated endocytosis or a non-clathrin pathway that is independent of SP-A [41]. In summary, pulmonary surfactant solves a complex bioengineering problem that requires lipids and proteins to act in multifunctional roles to maintain the delicate balance of surfactant homeostasis during the dynamics of breathing.

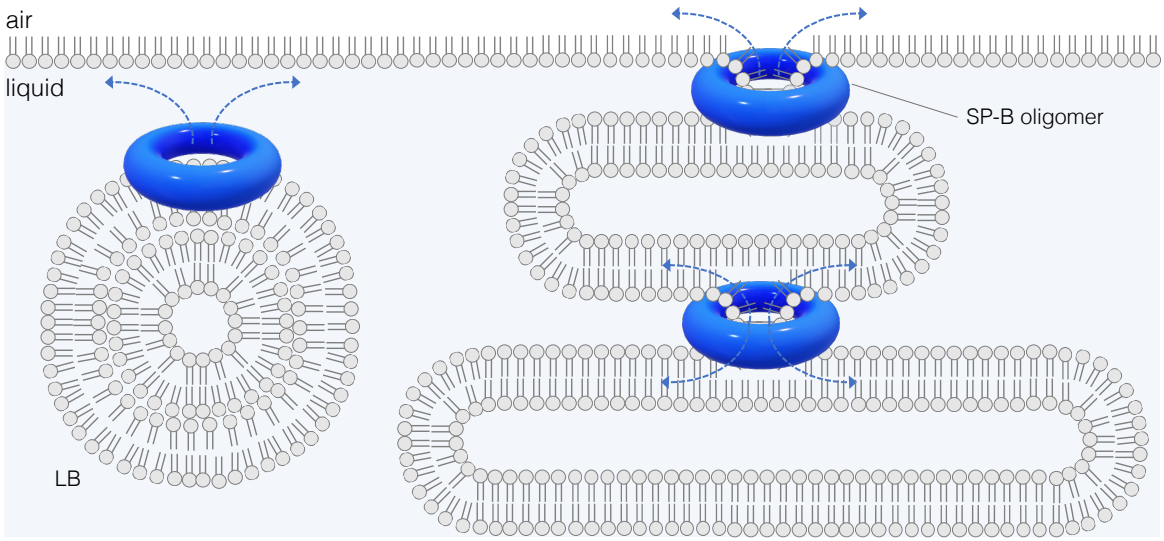


Figure 1.2 Model of SP-B function

Liekkinen *et al.* [32] proposed a model of SP-B function. The ring-shaped SP-B oligomer (shown in blue) causes perturbation of lipids at the ring interface and channels lipids through its hydrophobic, central pore, which shields the hydrophobic lipid chains. Accordingly, this allows for quick absorption of LBs to the air-liquid interface and quick deposition and lateral distribution from the surface film to its associated, multi-layered membrane reservoirs and *vice versa* during the dynamics of breathing.

1.2.3 Surfactant Protein B biosynthesis

SP-B is encoded in the *SFTPB* gene on human chromosome 2 (2p12-2p11.2). The ca. 10 kb *SFTPB* gene encodes 10 introns and 11 exons (Figure 1.3). *SFTPB* is expressed in both ATII and club cells, but only ATII cells efficiently produce the mature SP-B protein by complex proteolytical processing and can package it in LBs for surfactant deposition [42, 43]. The function of SP-B expression in club cells is not fully understood, however, the SP-B processing intermediate, in particular its N-terminal sequence, was found in club cells and bronchoalveolar lavage fluid (BALF) where it was linked to an antimicrobial function [44].

SFTPB encodes a 381 amino acid preproprotein containing large flanking sequences (Figure 1.3) which are thought to have a chaperone function for the highly hydrophobic mature SP-B. Deletion of the N-terminal propeptide sequence, but not the C-terminal sequence, disrupts intracellular trafficking [45]. SP-B is proteolytically processed in at least three steps during the passage from the Golgi apparatus through the multivesicular bodies to generate the mature 79 amino acid SP-B monomer protein in the LBs [43, 46]. Formation of the final homodimer form proceeds only from the monomer, therefore dimerization likely happens after proteolytic processing is complete [47].

The 79 amino acid SP-B monomer contains a high proportion of hydrophobic residues and high α-helical content [48, 49]. It belongs to the saposin-like protein family, that share a common fold and cysteine at six highly conserved positions, which form three intramolecular disulphide-bridges [48, 50]. While there is no crystal structure of SP-B, its structure has been modelled in accordance with saposin B [50]. Importantly, the SP-B monomer contains a seventh cysteine essential for homodimerization. Although optimal SP-B function seems to require homodimerization, SP-B with the cysteine

residue mutated can still rescue a mouse model of SP-B deficiency [51]. SP-B has a positive net charge of seven thought to be essential in interacting with negatively charged phospholipid head groups [48].

The SP-C protein is also synthesised as a large precursor protein, with a highly hydrophobic 35 amino acid protein consisting of a single membrane spanning helix, but its pulmonary surfactant function is less well understood. In patients, SP-C deficiency is characterised by the accumulation of misprocessed SP-C and diffuse alveolar damage with variable severity caused by altered folding and trafficking [52]. SP-C$^{(-/-)}$ mice survive and do not show respiratory distress syndrome (RDS) indicative of lack of functional surfactant as do SP-B deficient mice, but are more susceptible to infection, inflammation, intracellular lipid accumulation and fibrosis, indicating a role in innate immunity and/or surfactant recycling [53, 54]. Absence of SP-B causes incomplete processing of pro-SP-C [55]. One proposed mechanism for the impact of SP-B on SP-C processing is that the proteins directly interact and SP-B acts to expose the SP-C proteolytic cleavage sites [56, 57]. SP-C is thought to be involved in surfactant function alongside, and in complex with, SP-B for optimal function and adsorption to the surface film [57, 58]. The synthesis of other surfactant associated proteins was also found to be closely interconnected, e.g. ABCA3 knock-out animal models showed no LBs and aberrant processing of SP-B indicating that ABCA3 is necessary for LB formation and correct SP-B processing requires LBs [23, 59].

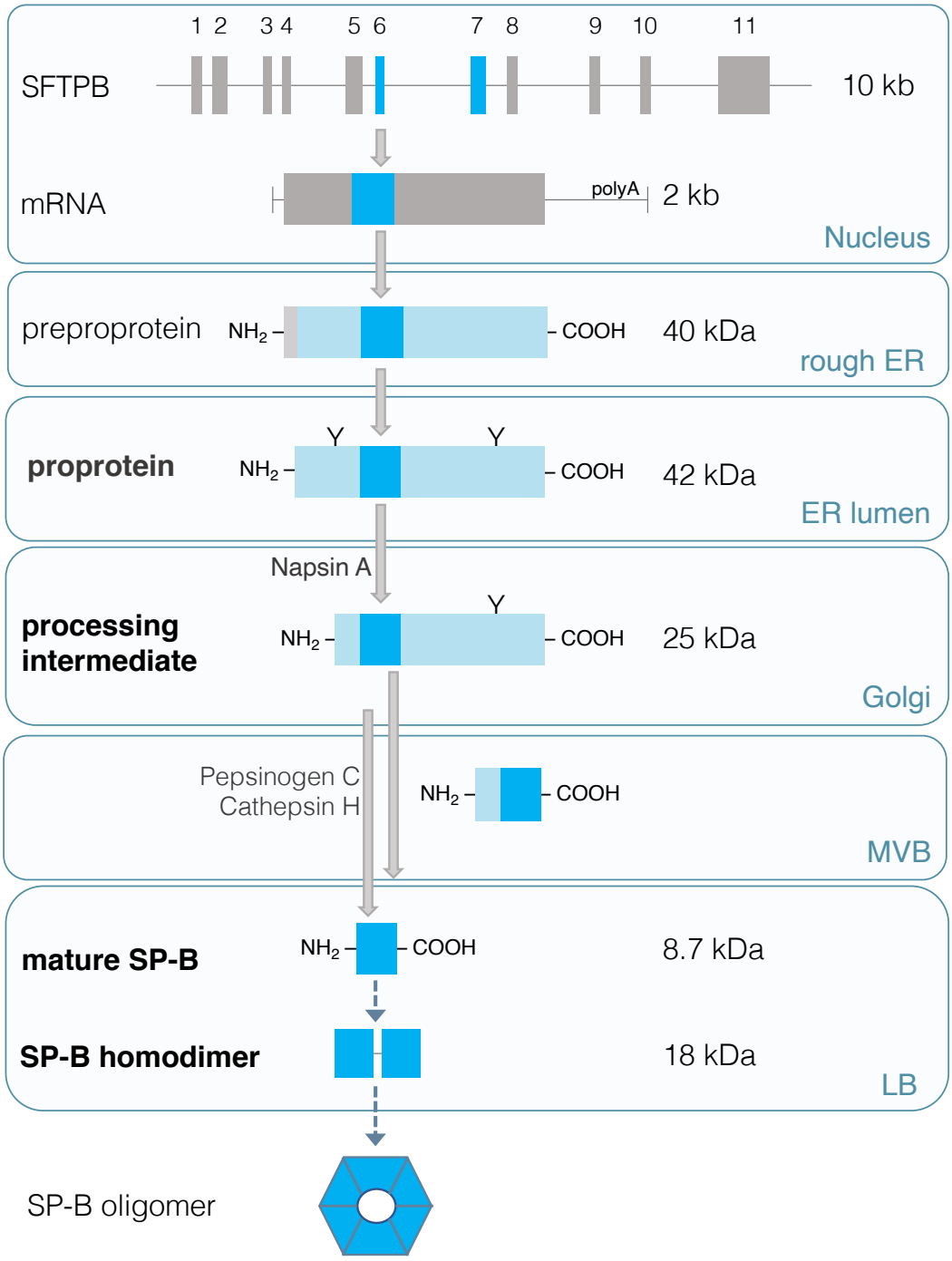


Figure 1.3 Biosynthesis of mature SP-B

The SP-B protein is encoded in exons 1-10 of the *SFTPB* gene that transcribes in ca. 2 kb mRNA. The last exon 11 is untranslated. The mature SP-B protein is encoded in exon 6 and 7. The preproprotein is translated at the rough ER and the 23 amino acid signal sequence is cleaved cotranslationally during ER import [60, 61]. The resulting 42 kDa proprotein [9, 62] contains the amino acid sequence of the mature protein flanked by N- and C-terminal sequences and is glycosylated in the ER lumen. The

Golgi apparatus is the place of glycosylation modification and cleavage of the N-terminal sequence by Napsin A resulting in the 25 kDa processing intermediate [63, 64]. In at least two proteolytical processing steps involving Pepsinogen C and Cathepsin H, SP-B is processed to its mature form during the passage through the multivesicular bodies (MVB) to lamellar bodies (LB) [43, 46, 63, 65]. The 8.7 kDa human mature SP-B [62, 66] is also observed as a homodimer (ca. 18 kDa [9]) and structural modelling suggests that SP-B dimers form non-covalent, ring-shaped oligomers, potentially hexamers [31]. The longer-lived SP-B intermediates that are commonly observed (in e.g. Western Blots) are shown in bold.

1.2.4 Genetic disorders of surfactant protein B

SP-B deficiency (OMIM 265120) is a very rare autosomal recessive genetic disease (1 in 1 million live births affected [67, 68]) that causes alveolar collapse and severe RDS in neonates. Clinical findings in these full-term neonates are similar to RDS in preterm neonates due to insufficient surfactant production [10]. However, treatment with corticosteroids or exogenous surfactant is ineffective in these patients [10]. Complete absence of SP-B is invariably fatal and the only treatment options are lung transplantation or palliative care. The timeframe for possible lung transplantation is very short, as patients usually die within the first 3 months of life [11, 69, 70]. That, and the low availability of donor organs, contribute to the poor prognosis of SP-B deficient neonates. Notably, SP-B deficiency is an autosomal recessive disease such that heterozygous parents and siblings are usually free from symptoms [71].

The most common mutation in *SFTPB* associated with SP-B deficiency is 121ins2 (C > GAA) in exon 4, with an estimated gene frequency of 1 in 1000 individuals, accounting for ca. 66 % of cases [9, 72]. The 121ins2 mutation causes a frameshift and a premature stop codon in exon 6, resulting in an unstable mRNA excluding the coding region of mature SP-B [73]. The mRNA is degraded by nonsense-mediated decay which results in the virtually complete absence of *SFTPB* mRNA and SP-B protein as well as fatal RDS in individuals homozygous for 121ins2 [10, 73]. The 121ins2 mutation can also be lethal in combination with a mutation associated with reduced mature SP-B expression, as reported for ca. 6 % of normal levels [74, 75]. Other rarer mutations in *SFTPB* associated with SP-B deficiency have been described, including nonsense, missense, splice-site and other frameshift mutations [9, 11, 76, 77]. Further, large deletions including exons 7 and 8 as well as in exons 8 and 9 (Figure 1.3) caused fatal SP-B deficiency [78-80]. *SFTPB* is associated with a high frequency of SNPs and high recombination rate [81]. However, in total, ca. 85 mutations are currently reported for

'*SFTPB*' on ClinVar (NCBI database for mutations associated with human disease, www.simple-clinvar.broadinstitute.org) of which 36 are classified as benign or likely benign.

Histological markers of SP-B deficiency include alveolar wall thickening, epithelial-cell dysplasia and pulmonary fibrosis. The most characteristic feature is the accumulation of granular lipoprotein aggregates in the alveolar spaces that often contain desquamated ATII cells and foamy alveolar macrophages [9, 82]. Using electron microscopy, enlarged disorganised LBs with a 'bubbly' appearance are visible and tubular myelin is absent [83]. Because of the interconnected roles in surfactant homeostasis, the absence of SP-B also leads to accumulation of extracellular proteins such as misprocessed SP-C precursor protein and altered phospholipid composition [10, 84].

Treatment using artificial surfactant or mechanical ventilator support as undertaken in premature babies are not effective therapies for SP-B deficiency [10, 11]. The ineffectiveness of conventional surfactant replacement therapy highlights the complex biophysics of surfactant components which artificial surfactant cannot yet replace. Similarly, SP-B protein replacement therapy could prove difficult to develop because of the complex synthesis of the highly hydrophobic protein, which will likely require packaging with lipids in a LB-like structure for protein stability. The tightly packed structure of LBs seems also relevant for surfactant function, which will be not easy to imitate synthetically. Lastly, any kind of replacement therapy will require highly frequent dosing to the lung parenchyma because of the high rate of surfactant recycling [36].

Gene therapy has the potential to permanently treat SP-B deficiency. As the aim of SP-B gene therapy is the restoration of SP-B expression in surfactant-producing cells, physiological systems of surfactant packaging and secretion can be utilised, making the

biodistribution and functionality of surfactant less of a concern. While ca. 6 % of normal adult SP-B levels as measured by Western blot still resulted in a fatal outcome [74], 12-16 % of mature SP-B expression, resulting from a homozygous splice site mutation, enabled an individual to breathe without intubation for ca. 3 months. However, with supplemental oxygen and declining lung function ultimately requiring lung transplantation [77]. This shows that there is likely a critical SP-B level necessary for lung function, but that even low SP-B levels can positively influence lung function. As heterozygotes for 121ins2 are without symptoms, this means that somewhere between 6 % and 50 % SP-B protein expression is likely beneficial for those affected by SP-B deficiency. As a first step, gene therapy could be expected to expand the window for transplant beyond the age of ca. 2 months (76 days mean reported [85]) that currently leads to a poor survival rate (31 % 5 year survival rate after listing for transplantation [85]), therefore ca. half of the parents chose compassionate care in a study conducted in the US [85] and transplantation is not usually attempted in the UK [11].

1.3 Gene therapy – principles of design

Gene therapy was first postulated in the 1970s as a way to treat patients with monogenic loss-of-functions disease [86], but efficacy was first shown only in 2000, treating X-linked severe combined immunodeficiency (SCID-X1, [87]). Early successes were overshadowed by unforeseen insertional mutagenesis and immune-toxicities, leading to leukaemia and fatalities [88] which highlighted the underlying complex biology of gene therapy and that understanding (viral) vector biology is key. Since then, many challenges have been overcome and gene therapy has expanded to many applications. The first gene therapy drugs were approved in China in 2003 (Gendicine, oncolytic virus), followed by the European Medicines Agency (EMA) in 2012 (Glybera™

for lipoprotein lipase deficiency) and by the Food and Drug Administration (FDA) in the USA in 2017 (Kymriah, chimeric antigen receptor (CAR) T cell therapy) and several gene therapies have been successfully developed since [88-90]. In haemophilia B gene therapy, a single gene therapy dose of recombinant adeno-associated viral (rAAV) vector resulted in long-term expression (> 5 years) and remarkable clinical benefit with no late-associated toxicities [91, 92]. For beta-thalassaemia, a potentially curative gene therapy treatment was EMA approved [93]. A particular successful form of gene addition therapy using an exogenous gene is CAR T cell therapy, where T cells are equipped with cancer-targeting antibodies by *ex vivo* gene transfer to target cancer cells. This approach has been enormously successful in treating various forms of leukaemia in the clinic, especially CD19-positive acute lymphocytic leukaemia with a 54 % complete response rate in meta-analysis [94]. These examples show that gene therapy can be effective with huge patient benefit.

1.3.1 Gene addition approach

To design a gene therapy, there are two major components to consider: what genetic information to deliver, and how to deliver that information. One common approach is gene addition, where the therapeutic gene is added alongside the mutated gene. While this approach can be utilised for suppressing toxic gene expression via RNA interference (RNAi), the most prevalent application is therapeutic protein expression where the mere lack of a certain protein is disease-causing as well as there being no toxic effect of any protein expressed from the mutant gene. This is likely to be the case in monogenic, recessive, loss-of-function mutations, such as SP-B deficiency.

1.3.1.1 Transgene cassette design

Gene therapy with the aim of gene addition requires the design of a therapeutic transgene cassette to enable expression of the therapeutic protein. There are three

major components to design: the promoter sequence, the transgene and the termination sequence. A schematic of a typical transgene cassette is shown in Figure 1.5 in Section 1.5.2.

The promoter sequence is tasked with driving transgene expression in the target tissue at the appropriate level, and commonly consists of an enhancer element (containing transcription enhancing, transcription-factor binding sites) and the promoter (point of gene transcription initiation). Common, constitutive (usually active) highly expressing promoters, include the immediate-early CMV (cytomegalovirus) promoter/enhancer, EF1a (elongation factor 1 a) promoter, UBC (ubiquitin C) promoter and CBA (chicken beta-actin) [95]. Further combinations of these promoter/enhancer elements have been designed, such as the CAG promoter (CMV enhancer element, chicken beta-actin promoter, splice acceptor of rabbit beta-globin) [96] and the hCEFI (Cytosine-guanosine dinucleotide-free form of the **h**uman **C**MV enhancer element, **EF**1a promoter and a synthetic **I**ntron) [97]. Amongst the ubiquitous promoters mentioned, CMV, CAG and CBA are considered strongest, but strength of expression can differ amongst specific cell types [95, 98, 99]. This highlights that adequate promoters for gene therapy must be chosen experimentally depending on the cellular target. If the size of the expression cassette is constrained, the promoter length can be adjusted by designing minimal promoters that often retain, or even enhance, transcription strength [95, 100]. A disadvantage of viral promoters such as CMV is that they are prone to epigenetic silencing, likely as part of a cellular defence mechanism [98, 101].

In contrast with ubiquitous promoters, transgene expression can be restricted to a given tissue or cell type using a specific promoter. This improves safety of the gene therapy, as off-target tissue expression can be toxic, but their use can be restricted by size or inadequate transcription strength. Possible solutions are design of minimal promoters

and/or pairing with suitable enhancer element as proved effective for pairing the CMV enhancer element with the *SFTPC* promoter [102]. Indeed, since 2015 a trend towards tissue specific promotors was observed amongst clinical trials using rAAV vectors ([103], rAAV discussed in detail in Section 1.5).

Another tool to restrict transgene expression to the target tissue is the addition of target sites for microRNAs (miRNAs), typically in the 3' untranslated region of the transgene cassette [95, 104]. If the miRNA target sequence contained within the transgene mRNA is complementary to a miRNA expressed by a cell, this leads to knock-down of transgene expression by RNA interference.

There are several expression cassette elements that can have positive effect on transgene expression, such as promoter proximal introns [105], which possibly enhance mRNA stability and transport out of the nucleus to the ribosome [106, 107]. For example, deletion of the UBC promoter proximal intron in cell lines strongly reduced promoter activity [108]. A study in lung carcinoma cells showed a positive effect of a promoter proximal SV40 intron on expression [109]. Depending on the version of CMV promoter used, the intron doubled the expression or had only modest effects highlighting the importance of the combination of expression cassette elements for expression outcome. Most studies report only minor effects of introns (up to 2-fold improvement, [109-113]); however, there are some studies reporting strong effects [114, 115], such as a study using introns engineered from bacterial elements that reported 40-100-fold increase over conventional rAAV expression cassettes [114].

Further, post-transcriptional regulatory elements (PRE) from viruses, such as Woodchuck hepatitis virus PRE (WPRE) and hepatitis B virus PRE (HPRE) have been shown to increase transgene expression [95, 116]. They act by enhancing nuclear export of viral mRNA [117, 118]. This effect seems to be particularly important in intron-

less viral mRNAs, as no positive effect of PREs on expression in intron-containing mRNAs was reported [119-121]. Further, cell line and transgene dependencies have been reported [122, 123].

Lastly, the polyadenylation of transcript strongly influences mRNA stability, nuclear export and translation. The choice of the polyadenylation signal can boost expression, as shown for the late SV40 polyA [111, 113, 115]. Overall, expression increase of the individual polyadenylation signals was less cell type dependent [95].

In summary, the elements of the transgene cassette design must be carefully selected in the light of its application and delivery strategy to ensure therapeutic expression levels in the target cells with limited and safe off-target expression. Although transgene cassettes are built from modules, the cross-influence of these, and differences observed *in vivo* versus *in vitro*, must be carefully considered. Optimal combinations for high and specific expression in target cells still need to be determined empirically.

1.3.1.2 Transgene optimisation

Similar to the components of the expression cassette described above, the transgene cDNA (complementary DNA) can also be optimised for higher expression by codon optimisation using silent codon mutagenesis. Transgene codons can be adjusted to favour codons with high tRNA concentration in the target species to accelerate translation and thereby elevate expression levels. Further, pairing of high and low frequency codons may influence translation speed [124].

Cytosine-guanosine dinucleotide (CpG) motifs have also been shown to influence expression. Unmethylated CpG motifs are recognised as foreign microbial nucleotides and can elicit an inflammatory immune response [125, 126]. *De novo* methylations of CpG motifs induce epigenetic silencing, causing loss of transgene expression [127, 128]. The silent mutagenesis of CpG motifs can reduce immunogenicity and prolong

transgene expression in *in vitro* and *in vivo* studies using plasmid DNA [97, 127, 128]. Similar positive effects were observed when using rAAV for transgene delivery. Compared with the conventional lacZ transgene encoding bacterial β-galactosidase, removal of CpG motifs resulted in robust and stable transgene expression and lower immunogenicity as characterised by less transgene-reactive $CD8^+$ T cells and lower IFNγ secretion [129]. However, there are also *in vitro* and *in vivo* studies using plasmid DNA where removal of CpG motifs did not increase transgene expression [130-132]. This highlights that changes in the transgene sequence can affect expression both positively and negatively via transcription and translation, complicating the postulation of rigid rules for transgene design.

The cDNA version of a gene, without introns, is commonly used in expression cassettes to circumvent vector size constraints. However, the removal of genomic splice sites and SNPs can unintentionally create intronic consensus and near-consensus splice donor and acceptor sequences within the cDNA, which can result in a variable degree of truncated or destabilised mRNAs [124, 133]. Potential outcomes are loss-of-function mutation and loss of expression. Unintended splicing can be circumvented by careful review of cDNA sequences and, where appropriate, introduction of silent mutations to eliminate cryptic splice signals and thus boost transgene expression [134].

In summary, there are virtually endless combinations of codons coding for the same amino acid sequence, but codon choice influences posttranscriptional modifications and translation and therefore protein expression levels. Bioinformatics tools are available to facilitate the application of known design rules; however, such rules remain imperfect and ultimately designs must still be validated empirically.

1.3.2 Gene editing approach

Gene editing is used to repair a mutated gene that causes disease, to replace it with a functional copy and/or prevent its expression, which is beneficial in genetic diseases with toxic gene expression. During gene editing, a nuclease targets a position in the genome to induce a double strand break (DSB). The triggered DNA damage response of the cell subsequently repairs the DSB, and the cell can use different repair pathways which can be exploited for gene editing; the results can include large knock-ins, knock-outs as well as small, precise edits.

Amongst available gene editing tools, the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein 9 (Cas9) system has revolutionised the field, because it can be easily adapted to new genetic targets compared with previously used zinc finger nucleases (ZFN) and transcription activator-like effector nucleases (TALEN), while still remaining highly specific. The CRISPR system originates from the bacterial adaptive immune system which inserts foreign nucleic acids in the CRISPR region [135-137]. The CRISPR/Cas system was reprogrammed as a gene editing tool in 2012 by Jinek *et al.* [138]. The two RNA components of CRISPR/Cas were fused in a chimeric single guide RNA (sgRNA) resulting in a convenient two-component editing system: sgRNA and effector Cas nuclease. The sgRNA can be altered without the need to alter the Cas9, abolishing the requirement for time-consuming ZFN or TALEN protein engineering.

After the adaptation of CRISPR/Cas9 as a gene editing system, it was rapidly fitted to edit mammalian cells [139-141] and facilitated the acceleration of research in gene function and the creation of disease models in cell lines and animals, such as mice and monkeys [142, 143]. One of the first studies demonstrating the therapeutic potential of CRISPR/Cas9 was to correct the *Dmd* mutation in Duchenne muscular dystrophy

(DMD); first in zygotes [144], and later in post-natal mice by rAAV CRISPR delivery [145]. The progression from the zygote to animal has been essential in the development of therapeutic CRISPR editing because gene correction in post-natal animals requires a strategy suitable for *in vivo* delivery that models the treatment of patients much more closely. The therapeutic effect of gene editing has been demonstrated in models of many genetic diseases, such as cystic fibrosis [146] and alpha-1 antitrypsin (AAT) deficiency [147]. Clinical trials are underway using therapeutic gene editing with ZNFs for, e.g. SCID-X1, Wiskott-Aldrich syndrome and sickle cell anaemia [148]. Genetic blood disorders are regarded as a safe entry point for gene editing due to the possibility of *ex vivo* treatment and subsequent quality control of edited cells. Cas9 gene editing has been used for T cell therapy to knock out genes associated with T cell fatigue. The cell products showed improved transferred T cell persistence compared to previous trials and were safe at early time points [149]. The gene editing field has moved at a fast pace and a first Cas9 *in vivo* gene editing trial for congenital blindness was initiated in September 2019 (NCT03872479, phase I/II).

CRISPR-induced edits are permanent resulting in permanent correction; however, off-target edits do occur, with unforeseeable and possibly negative consequences. Gene editing as a gene therapy application introduces yet another variable, and potential advantages over gene addition approaches need to be carefully considered for the specific target disease. In the case of SP-B deficiency, where it is the lack of expression, rather than toxic (mutant) gene expression, which is disease-causing, then the comparatively more straight-forward gene addition is the first choice, especially when considering timely translation to the clinic.

1.3.3 Viral and non-viral delivery strategies for gene therapy products

Efficient delivery to the target tissue is crucial for the efficacy of virtually all drugs but is a specific challenge in macromolecular therapies. Broadly, gene therapy delivery strategies require a viral or non-viral (synthetic) vector. Non-viral vectors include nucleic acids, such as plasmid DNA, (chemically modified) mRNA and in the case of gene editing, ribonucleoprotein (RNP) complexes. Therapeutic nucleic acids can be delivered 'naked' via physical methods or, more commonly, complexed with lipid/polymeric materials into nanoparticles to condense and protect the nucleic acid thereby improving transfection and subsequent expression efficiency. In clinical trials, expression was safe, but detected at low, often sub-therapeutic levels and with limited duration [150]. However, the transient nature of expression may be seen as an additional safety feature. Compared with viral vectors, the interaction between vector and host-cell is thought to be simpler to understand and control. Using a viral vector for transgene delivery adds another layer of complexity, although they have mediated safe and stable expression at therapeutic levels in the clinic and the vast majority of approved gene therapy products today are based on viral vectors.

Common differentiation criteria for viral vectors are whether they are integrating (inserted into cellular genome) or non-integrating (exist as episomes alongside cellular genome). Retroviral vectors integrate into the host chromatin allowing for persistent expression. Gamma-retroviral vectors were extensively used in the early days of gene therapy, but were shown to cause oncogenic genotoxicity [151, 152]. The field has largely moved on to recombinant lentiviral (rLV) vectors which appear to have no integration-preference for oncogenic loci [7]. In particular, third generation rLV vectors contain a self-inactivating (SIN) 3'LTR to abolish its promoter function and are produced in a four-plasmid packaging system to avoid generation of replication-competent lentivirus [153-155]. Further desirable features of rLV vectors include the ability to

transduce both dividing and non-dividing cells, persistent expression, potential for repeated administration and large packaging capacity [7]. Lentiviral vectors are commonly based on human or simian immunodeficiency virus (HIV, SIV) and have excelled in gene transfer to haematopoietic cells such as in CAR T cell therapy. To broaden their tropism, rLV vectors can be pseudotyped, as done with the F ('fusion') and HN ('hemagglutinin-neuraminidase') envelope proteins of the respiratory Sendai virus to improve lung gene transfer [156]. The resulting rSIV.F/HN vector demonstrated high level expression levels in human airway epithelial cultures (HAE) and long-term expression in the murine lung, including parenchyma and airways [7].

Adenovirus (AdV) and AAV are non-integrating viruses. AdV is a double stranded DNA virus with many naturally occurring serotypes and a capsid diameter of 90-100 nm. Recombinant AdV (rAdV) vectors have a large packaging capacity (up to ca. 36 kb for 'gutless' vectors) and are commonly used in vaccination and anticancer immunotherapies as they are strongly immunogenic [157], a characteristic that makes them less suitable for long-term gene expression, such as required for the treatment of monogenic diseases. A specialised application of rAdV is using replication-competent rAdV as an oncolytic virus for cancer therapy which takes advantage of the cytolytic activity of AdV at the end of its life cycle [157, 158].

One of the most extensively studied viral gene therapy vectors is rAAV, based on a small DNA virus (ca. 25 nm capsid diameter) with limited packaging capacity (ca. 4.7 kb), relatively low immunogenicity and being non-pathogenic in humans. Several naturally occurring capsids known as serotypes have been identified. Recombinant AAV has been investigated in numerous clinical trials, with some rAAV-based gene therapy products already reaching approval and many more being in the pipeline. This vector is discussed in detail in Section 1.5.

1.4 The lung as target for gene therapy

1.4.1 Delivery of lung gene therapies

The systemic delivery route, administered via intravenous injection, has been shown to result in high levels of transgene expression in the liver as well as other organs including lung [159-162]. Intravenous injection delivers vector to the whole body, so the vector is more diluted compared with targeted delivery to the lung. Interestingly, intravenous injection for delivery to the lung tends to target endothelial rather than epithelial cells [163, 164]. However, airway and respiratory epithelia can be targeted more specifically, by delivery directly to the lung lumen via inhalation (intranasal administration) or via the trachea (intratracheal administration), or directly to the airways by dosing single lung lobes using a nebulisation/spray catheter [165]. The most popular lung delivery method is aerosolization, facilitating the maximum spread throughout the organ with an extensive surface.

1.4.2 Obstacles for lung gene therapy

Despite naturally good accessibility, or more likely exactly because of it, there are several barriers to lung gene therapy delivery by inhalation. Since the lung has evolved to defend itself against inhaled pathogens, the question emerges whether targeting the lung via inhalation is as advantageous as it intuitively seems.

In the conducting airways, the barriers of the airway surface are the mucus gel layer and periciliary layer. Airway mucus is a heterogeneous mixture, primarily made from large glycoproteins forming mucin fibres [166]. Mucus forms a porous, gel-like structure that can trap inhaled particles, pathogens and gene therapy vectors alike [167]. In the case of pre-existing immunity, which is common for viral vectors, antibodies secreted into the airway surface liquid can add another hurdle by neutralising gene therapy vectors [167]. The periciliary layer is located below the mucus and comprises a finer

mesh structure than mucus [168]. The pores in the periciliary layer are theoretically big enough to be permissive for small vectors such as rAAV [168] and there is limited evidence that AAV, and not AdV, moves through the periciliary layer [169].

In the lung parenchyma, the barriers include pulmonary surfactant and alveolar macrophages. Pulmonary surfactant was shown to negatively impact gene delivery via cationic lipid-based vectors, where components of pulmonary surfactant, such as phospholipids, caused destabilisation of the vectors, potentially by interaction with the vector lipids, followed by aggregation of the vector [170, 171]. Non-viral vectors based on cationic polymers were more stable in surfactants compared with lipid-based vectors [172], suggesting they would be the better choice for delivery to the lung parenchyma. Alveolar macrophages are critical for host defence and lung homeostasis and as part of the innate immune system, they engulf inhaled particles such as pathogens. Surface charge of particles influences phagocytosis activity of alveolar macrophages. Surfactant proteins A and D and, in the case of pre-existing immunity, pulmonary antibodies can boost particle phagocytosis by modulating surface charge allowing for convergence of the macrophages [167, 173].

Pulmonary drug delivery results in a high local concentration of vector, which reduces pressure on vector production and helps restrict unwanted off-target expression. Therefore, pulmonary delivery is the preferred administration route and barriers of the lung, which might also be changed by the underlying genetic condition, is an important consideration for the design of pulmonary gene therapy.

1.4.3 Clinical experience of lung gene therapy

The natural accessibility of the lung was a factor leading to the undertaking of some of the earliest gene therapy trials in the lung, administering vector using nebulisers. A phase IIb clinical trial for cystic fibrosis using aerosolisation of a non-viral plasmid

DNA/cationic lipid (GL67A, cationic lipid) formulation showed modest effects on lung function. However, the study still demonstrated the safety of the vector used and the potential of gene therapy to improve lung function for cystic fibrosis. The authors concluded that improvements in efficacy were needed before the therapy is suitable for clinical care and suggested the use of more efficient gene transfer vectors [174].

The first clinical study using rAAV in the lung was also performed to treat cystic fibrosis, using AAV serotype 2 [175, 176]. In a series of trials published between 1998 and 2007, rAAV proved to be safe but ultimately ineffective [176-179]. One possible reason for the lack of efficacy, is the absence or low abundance of AAV2 viral entry receptors on the apical surface [180, 181]. When the gene therapy vector for this trial was designed, considerably fewer serotypes were available and the relationship between capsids, their distinct entry receptors and how that translated into cell and tissue tropisms were less well understood. In the lung, viral entry receptors are differently distributed on the apical versus basolateral surface of the lung and for pulmonary gene transfer the receptor must be accessible from the lumen-facing, apical lung surface. Moreover, receptor distribution varies between human and animal models which complicates screening and selection of suitable viral vectors [180, 182, 183]. At the time point of the trials it was further not known that some AAV serotypes, including AAV2, bind adhesively to mucus causing steric obstruction [184, 185], which would influence the choice of a vector candidate for airway delivery, especially in designing a gene therapy for cystic fibrosis.

With the newly acquired knowledge about rAAV vectors and the potential pitfalls in their selection, there is now an entirely new opportunity to select rAAV vectors for efficient gene transfer to the lung. The previously demonstrated safety of rAAV in the lung is an excellent starting point to develop a pulmonary gene therapy. The safety data could further accelerate the authorisation of clinical trials for urgently required treatment of SP-B deficiency.

1.5 Adeno-associated viral vector-based gene therapy

1.5.1 Adeno-associated virus biology

Adeno-associated virus (AAV) is a DNA virus from the family *Parvoviridae,* which was first discovered as a contamination in adenovirus preparations [186]. Because AAV replication is dependent on coinfection with a helper virus, such as AdV, it is further classified in the genus *Dependoparvovirus* and has been found in multiple species including human and non-human primates (NHP, [187]). The general agreement is that AAV does not cause disease in humans (discussed in detail in Section 1.5.3).

The AAV capsid is assembled from 60 subunits of the three capsid proteins VP1, VP2 and VP3 (1:1:10 ratio). The resulting icosahedral capsid of ca. 25 nm in diameter harbours the ca. 4.7 kb genome which can be either sense or anti-sense ssDNA and is flanked by inverted terminal repeats (ITR, [188]). The AAV *rep* gene encodes four proteins (Figure 1.4), which are (in order of their molecular masses): Rep78, Rep68, Rep52 and Rep40 and are required for viral replication and genome integration. It is thought that the large Rep proteins are required for AAV DNA replication and serve as packaging signals, while the helicase function of Rep52 and Rep40 moves the viral DNA through a pore into the pre-assembled capsid [189]. The AAV *cap* gene encodes the three capsid subunits VP1, VP2 and VP3 through different start codons and alternative splicing, with the result that all three capsid proteins share the same C-terminal sequence [190]. The *cap* gene further encodes the assembly-activating protein (AAP) which supports capsid assembly [191, 192] and is indispensable for the assembly of some serotypes [193]. The AAV life cycle is not fully understood and other proteins relevant to viral replication encoded in the *cap* gene have been recently identified (mAAP) or predicted by computational studies [194, 195].

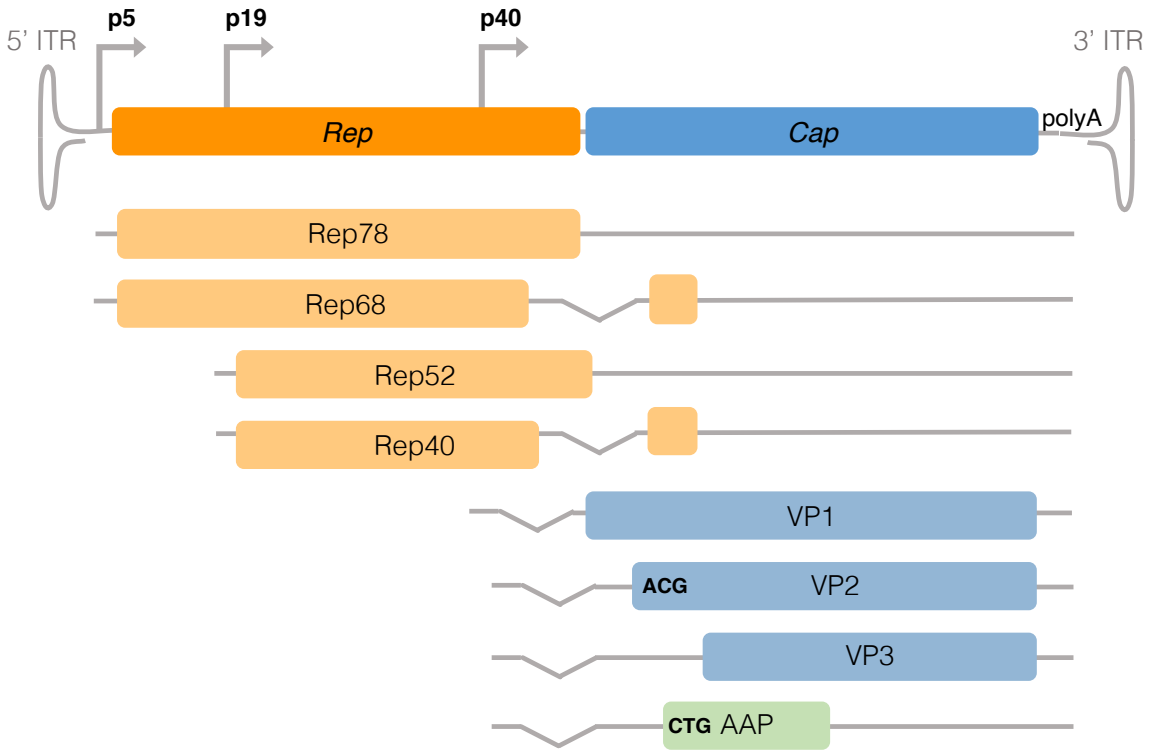


Figure 1.4 Schematic of the AAV genome

The 4.7 kb AAV genome is sandwiched between two 145 bp palindromic sequences known as Inverted Terminal Repeats (ITR), which form T-shaped hairpins and serve as packaging signal and viral origins of replication [196]. The AAV genome consists of two genes: *rep* in ORF1 and *cap* in ORF2. Rep proteins Rep78 and Rep68 are transcribed using the p5 promoter; Rep52 and Rep40 using the p19 promoter. Rep40 and Rep68 are the results of splicing. The *cap* gene is transcribed using the p40 promoter and by alternative splicing there are two mRNA versions. One translates into VP1, the other in VP2, VP3 and AAP [197]. The unusual ACG and CTG initiating codons of VP2 and AAP are indicated, for all other AAV proteins the commonly used ATG initiating codon is utilised. The VP2 start codon and splicing promote VP1:VP2:VP3 expression in 1:1:10 ratio [197].

1.5.2 Recombinant adeno-associated viral vectors

Wildtype AAV can be converted into a transgene delivery tool by excising the AAV genome between the ITRs and replacing it with a transgene expression cassette. The resulting rAAV vector is typically produced using a triple plasmid transfection approach ([198], Figure 1.5). The ITRs are preserved because they are essential for efficient viral genome packaging [199] and are the only sequences of AAV origin remaining within rAAV particles [200]. Because rAAV is 'gutless', genes essential for virus production such as *rep* and *cap* are provided *in trans*. Using a triple transfection system increases safety by decreasing the likelihood of replication-competent or integration-competent virus being generated from the supplied viral elements [198]. Further, it provides flexibility for the production of serotype/transgene combinations by changing just one plasmid during transfection.

There are several naturally occurring AAV capsids called serotypes (13 identified to date) that exhibit various cell and tissue tropisms [159]. The AAV serotype 2 (AAV2) is the most extensively studied serotype and the ITRs and *rep* gene of virtually all rAAV vectors are based on AAV2. AAV serotypes are able to target multiple organs including liver, lung, muscle and heart, as well as the central nervous system (CNS), which is easier to target due to the remarkable ability of AAV to cross the blood-brain barrier (BBB, [201]). Because rAAV can transduce dividing as well as non-dividing cells, it has become the dominant choice for CNS delivery, especially the eye.

The AAV capsid enters its target cell by interaction with distinct cell receptors for the particular serotype, usually glycan receptor(s) and protein co-receptor(s). Binding to the cell surface via the entry receptors leads to internalisation by endocytosis (clathrin-dependent and independent pathways, [202]). The low pH in the endosome and late endosome triggers conformational changes necessary for transduction such as

exposure of the phospholipase A2 unit of VP1 [203]. Phospholipase A2 is thought to be involved in endosomal escape and/or nuclear entry [190]. Endosomal escape results in the release of rAAV into the cytoplasm and subsequent nuclear entry for successful transduction. Proteasomal or endosomal degradation of AAV capsids can prevent transduction [204]. Only after arrival in the nucleus, does rAAV uncoat its DNA [203], after which, the single strand rAAV genome is converted to double-strand either by multi-vector single-strand annealing or by *de novo* second strand synthesis; ultimately this double strand conversion step seems to be rate-limiting [205]. Double-stranded rAAV organises in its episomal form by ITR-driven recombination into circular monomers, dimers or concatemers [206].

The rAAV genome does not usually integrate and remains episomal as it lacks the *rep* gene that encodes the Rep protein necessary for integration of wildtype AAV at the preferred integration locus *AAVS1*. For rAAV, only low rates of integration (ca. 0.1 %) have been reported which are likely caused by DSBs and non-homologous end joining (NHEJ) integration [207, 208]. In line with these findings, high rates of rAAV genome integration at Cas9 cut sites, but not elsewhere, have been reported [209]. Episomal rAAV is highly stable and provides long-term expression, especially in post-mitotic tissues [206], but in highly-proliferating tissues, rAAV expression is likely to be lost over time. Episomal DNA is not replicated along with the genomic DNA during cell division and therefore expression is eventually lost by progressive dilution.

In a clinical setting, it is advantageous that rAAV is stable across wide temperature and pH changes. For example, rAAV1 has been shown to be stable over 12 weeks at 4 °C and across four freeze-thaw cycles [210], mainly due to the resistance of rAAV to physical stress such as shearing forces, which means that it can be delivered by vibrating mesh or jet nebulisation to the lung with little loss of infectivity [211, 212].

The very first rAAV clinical trials used rAAV2 to treat cystic fibrosis where it was well-tolerated, but ineffective [177]. Other early rAAV gene therapy clinical trials aimed to treat haemophilia A and B. Here, rAAV gene therapy also proved to be safe in pre-clinical and clinical studies, and important lessons about the interference of rAAV gene therapy with the immune system were learnt, such as the effect of anti-capsid pre-existing immunity on rAAV gene transfer and the possibility of an anti-transgene adaptive immune response [213]. Another important observation was the activation of cytotoxic $CD8^+$ T cells specific to transduced cells presenting AAV capsid epitopes leading to elimination of transduced cells and diminished transgene expression which prompted the development of early intervention or prophylactic immune suppression protocols using corticosteroids for rAAV gene therapy clinical trials [92, 213]. Since then, rAAV continued to prove generally safe in the clinic and showed efficacy in multiple trials (reviewed in [214, 215]). The evidence led to the approval of Glybera to treat congenital lipoprotein lipase deficiency following muscle administration (EMA approval 2012); Luxturna to treat the retinal disease Leber congenital amaurosis (EMA approval 2018) and Zolgensma to treat the neuromuscular disorder spinal muscular atrophy (FDA approval 2019, EMA approval 2020, [214]). With several ongoing clinical trials, many more therapies are on the horizon and the FDA expects 10-20 approvals of gene therapy products before 2025 [216].

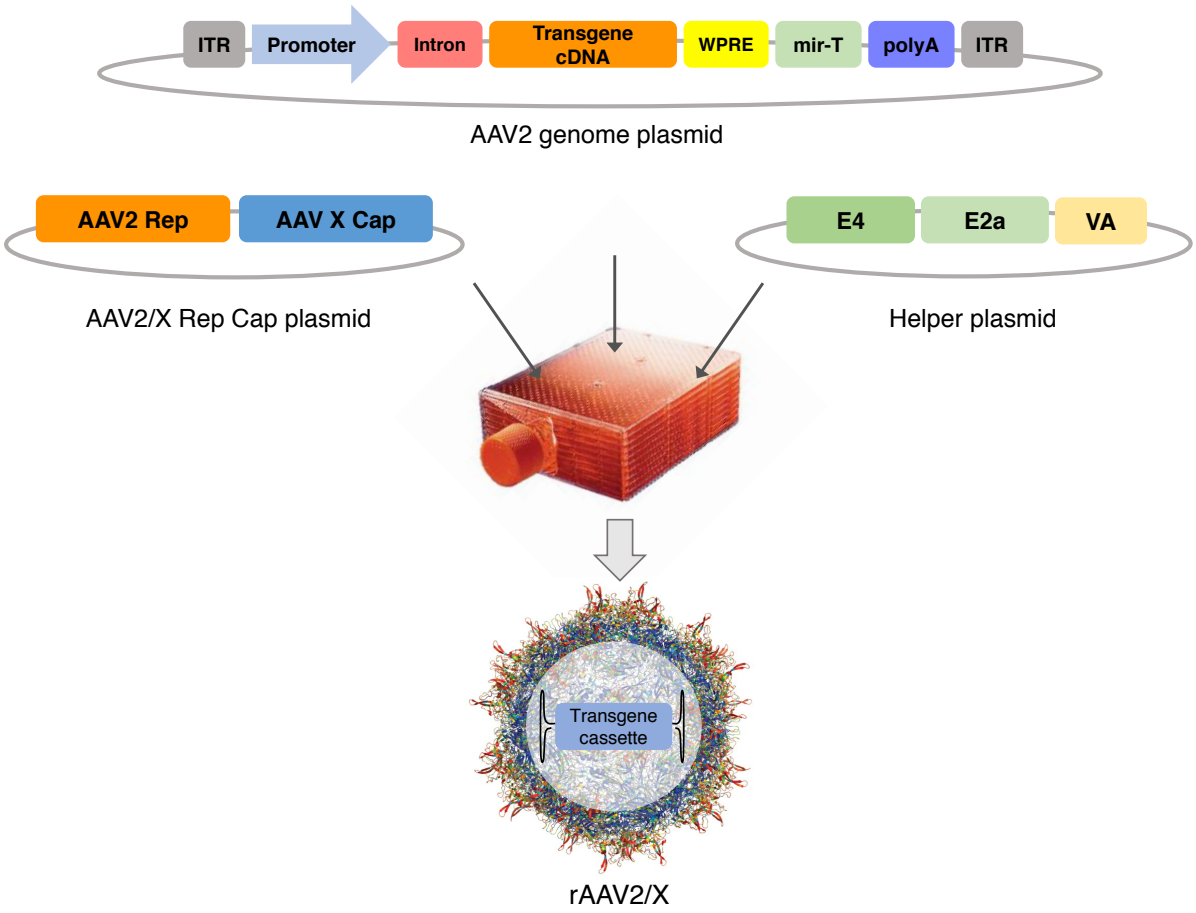


Figure 1.5 Schematic of rAAV production

Recombinant AAV is typically produced by triple transfection (Section 2.5.1). The producer cells are transfected with the genome plasmid containing the transgene expression cassette between the ITRs, the *rep/cap* plasmid and helper plasmid. The ITRs and *rep* gene are commonly based on AAV2. The helper plasmid contains genes for the adenoviral proteins E4, E2a and VA RNA necessary for replication. The necessary adenoviral E1 gene is typically provided by the producer cell line [190]. The *cap* gene encodes the capsid and *rep* viral AAV proteins necessary for DNA packaging into the virion. The producer cells can be adherent cells, which are commonly grown in multi-layered HYPERflasks to increase production scale or suspension cells. After transfection, rAAV is harvested, purified and concentrated. AAV capsid image modified from [217].

1.5.3 Limitations of recombinant AAV-based gene therapy

One major drawback of rAAV is its declining expression over time, because of its episomal nature. Although expression is eventually lost, long-term expression in dividing cells has been observed ([218], at least 52 weeks in murine lung), but it has also been argued that the non-integrating nature of rAAV is an advantage, as it minimises the risk of genotoxicity. An important disadvantage of rAAV is the relatively small packaging size of ca. 4.7 kb (5.2 kb at the very maximum [219]), which limits the choice of transgene cassette elements, such as large promoters. Also, some transgenes are too big for rAAV expression cassettes, requiring circumvention strategies such as alternative vector choice or the design of minimal/truncated genes, e.g. mini-dystrophin for treatment of DMD [220].

Seroepidemiological studies imply that at least serotypes AAV1-3 and AAV5-9 are endemic to humans [221, 222]; thus there is a high prevalence of pre-existing immunity and neutralising antibodies rendering parts of the population potentially ineligible for rAAV gene therapy based on these AAV serotypes [223]. An effort has been made to isolate AAV variants from other species to avoid pre-existing immunity [224]. However, not all isolates are sufficiently different to evade pre-existing immunity. Another rAAV challenge is the adaptive immune response towards AAV capsids which complicates efficient readministration of the vector. This is problematic in combination with the episomal nature of rAAV and in the lung, which has evolved to maintain strong host defence systems (as discussed in Section 1.4.2) including secretion of antibodies to the lung lumen. However, strategies to either make rAAV expression persist, or to allow for repeat administration, could solve this problem. A recent study using enzymatic digestion of IgG *in vivo* showed promising effects on the repeat administration of rAAV in NHPs [225].

Although rAAV vectors have proven safe in numerous clinical trials, severe toxicities were first reported for rAAV9 variants at high intravenous (IV) doses in 2018 [226, 227]. High-dose rAAV9 variant (2E14 genome copies (GC)/kg, IV) administered to juvenile NHPs and piglets led to severe toxicity; the NHPs exhibited severe and acute hepatic damage requiring euthanasia of one animal and degeneration of dorsal root ganglia sensory neurons [227]. No hepatic damage was observed in piglets, but neurotoxicity was observed by ataxia and histopathology and required euthanasia of all animals. A meta-analysis of rAAV studies in NHP found dorsal root ganglion disease in 32 % of animals dosed IV which was mild to moderate in most cases and was affected by delivery route, age at injection and dose, but not by serotype or vector purification method [228]. There are also recurring discussions about a potential link between AAV and hepatocellular carcinoma, particularly in the rare event of genome integration, however, this link if real at all [229, 230], is deemed very small [231].

In late-phase clinical studies based on rAAV9 to treat spinal muscular atrophy (SMA), that eventually led to the approval of Zolgensma, only manageable levels of thrombocytopenia and liver transaminase elevations, a measure of liver damage, were observed after dosing individuals with 2E14 GC/kg. In the course of the studies two deaths were reported. According to cumulative safety data reported by the sponsoring company Novartis, after approval, both deaths were unrelated to rAAV gene therapy based on the autopsy findings [232]. In a gene therapy trial using a very high dose of rAAV8 conducted in 2020, the highest dose ever used in rAAV trials (3E14 GC/kg, IV, the average adult body has ca. 3E13 cells [233]), three patients died and sepsis was identified as the immediate cause of death [234, 235]. From the clinical data, questions remain as to whether high IV doses of rAAV are generally a risk, if the observed toxicities are serotype-dependent and if different age at dosing between the trials (with younger children being less vulnerable) play a role [236], while NHP data points as well

to age-dependent effects on sensory neuropathies, but serotype-independency [228]. These findings show that (perhaps unsurprisingly in retrospect) limits on rAAV doses exist, and research is underway to understand the circumstances when these limits are more flexible.

A current challenge for the clinical translation of gene therapies is the associated high costs. The first approved rAAV gene therapy Glybera cost over 1 million EUR and was withdrawn from market largely due to economic reasons. Since then, pharmaceutical companies are exploring outcome-based payment models to convince payers, such as annual rates to allow for a comparison of alternative, symptomatic treatments versus the cost gene therapy. One of the reasons for the overall high cost of rAAV gene therapy is the cost of manufacturing which is dramatic in comparison with small molecule drugs. Possible strategies to reduce cost are more cost-efficient manufacturing (at large-scale) as well as more efficient rAAV vectors reducing the required dose and potentially avoiding the toxicities described above.

1.5.4 AAV serotypes and recombinant AAV vector tropisms

The AAV family of viruses is diverse with several serotypes and hundreds of naturally occurring variants. Primarily, AAV serotypes have been isolated from humans and NHPs [221, 237] and share a high degree of sequence homology, between ca. 60-99 %, with AAV4 and AAV5 being the most distinct from each other and the remaining serotypes (Figure 1.6). Depending on sequence homology and antigenic reactivity, serotypes and variants can be ordered in a phylogenetic tree and categorised in genetic groups called clades [221]. The clades are often less meaningful for vector tropism, as even a single amino acid change (‘singleton change’) can influence vector tropism [238]. However, broad tropisms can be attributed for some clades, such as liver tropism for clade E.

The AAV capsid is built from the three capsid subunits VP1, VP2, and VP3 with overlapping amino acid sequences (Figure 1.4, above). The largest subunit VP1 comprises a unique region (138 aa) which contains phospholipase 2 activity and the VP1/VP2 shared sequence harbours the nuclear localisation sequence ([239], Figure 1.7). Both of these sequences are located inside the capsid [240]. The capsid forms its shell from a conserved eight-stranded antiparallel β-barrel core motif, with the β-barrel strands connected by flexible loops that shape the capsid exterior. A conserved loop at the C-termini of VP capsids and a highly conserved α-helix located at the capsid interior mediate subunit interaction [241].

The loops shaping the capsid exterior are variable between AAV serotypes. There are nine variable regions on the capsid surface as determined by structural comparison of AAV2 with AAV4 (Figure 1.7) [242]. Other, more closely related AAV serotypes differ in fewer of these variable regions, such as AAV2 and AAV8 in the three variable regions I, II, IV [243]. The surface loops determine AAV capsid surface topology and therefore influence cell surface receptor binding. Certain motifs within the variable regions are associated with different functions (Figure 1.7).

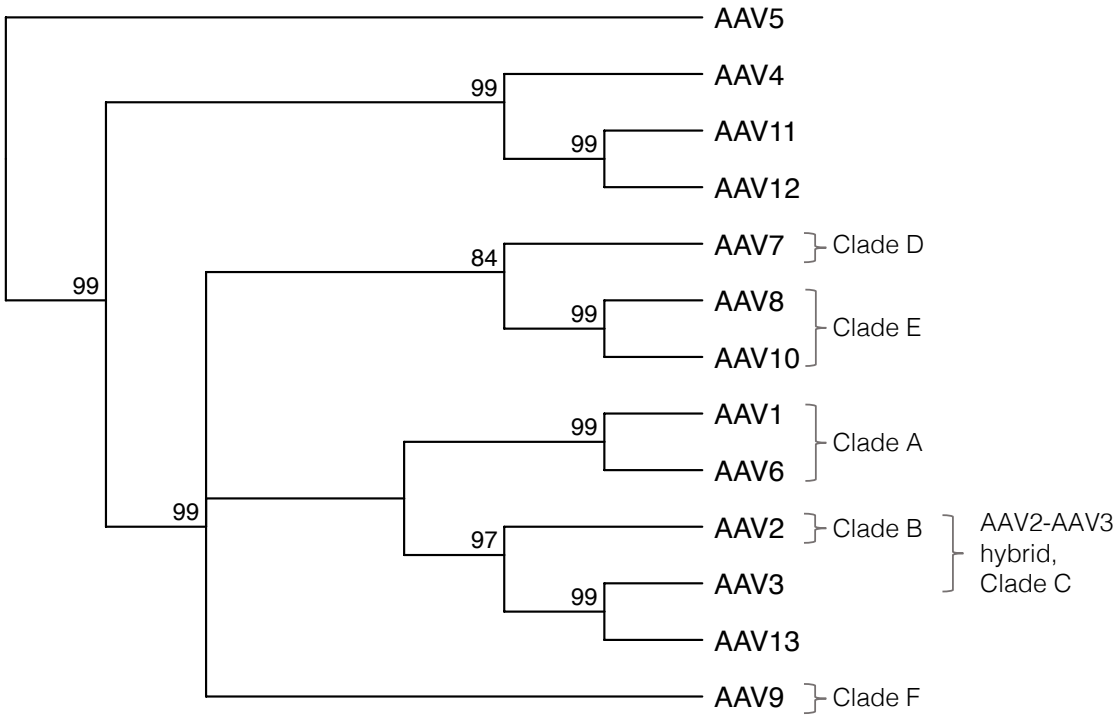


Figure 1.6 Phylogenetic tree of AAV serotypes 1-13

Neighbouring phylogenies of AAV1-13; generated by outgroup rooting and uncorrected p-distances using MacVector on the basis of VP1 capsid amino acid sequences alignment (as shown for rAAV1-9, below). Bootstrap values (n= 1000) are indicated at the branching nodes. The indicated clades were established in an analysis comparing sequences and antigen reactivity of AAV serotypes 1-9 and over 200 isolates [221]. AAV variants can be assigned to clades as shown in [221, 244]. AAV4 and 5 are most distinct from serotypes in clade A-F. AAV11 and AAV12 are most related to AAV4. Clade C originated from one recombination event of Clade B (AAV2) and AAV3. AAV13 is most related to AAV capsids in clade C. AAV1 and 6 share the same clade because of high sequence homology and antibody cross-reactivity. Therefore, according to the definition of 'serotype' as serologically distinct, AAV1 and 6 are not true serotypes, but the initially adopted numbering was not changed by the field.

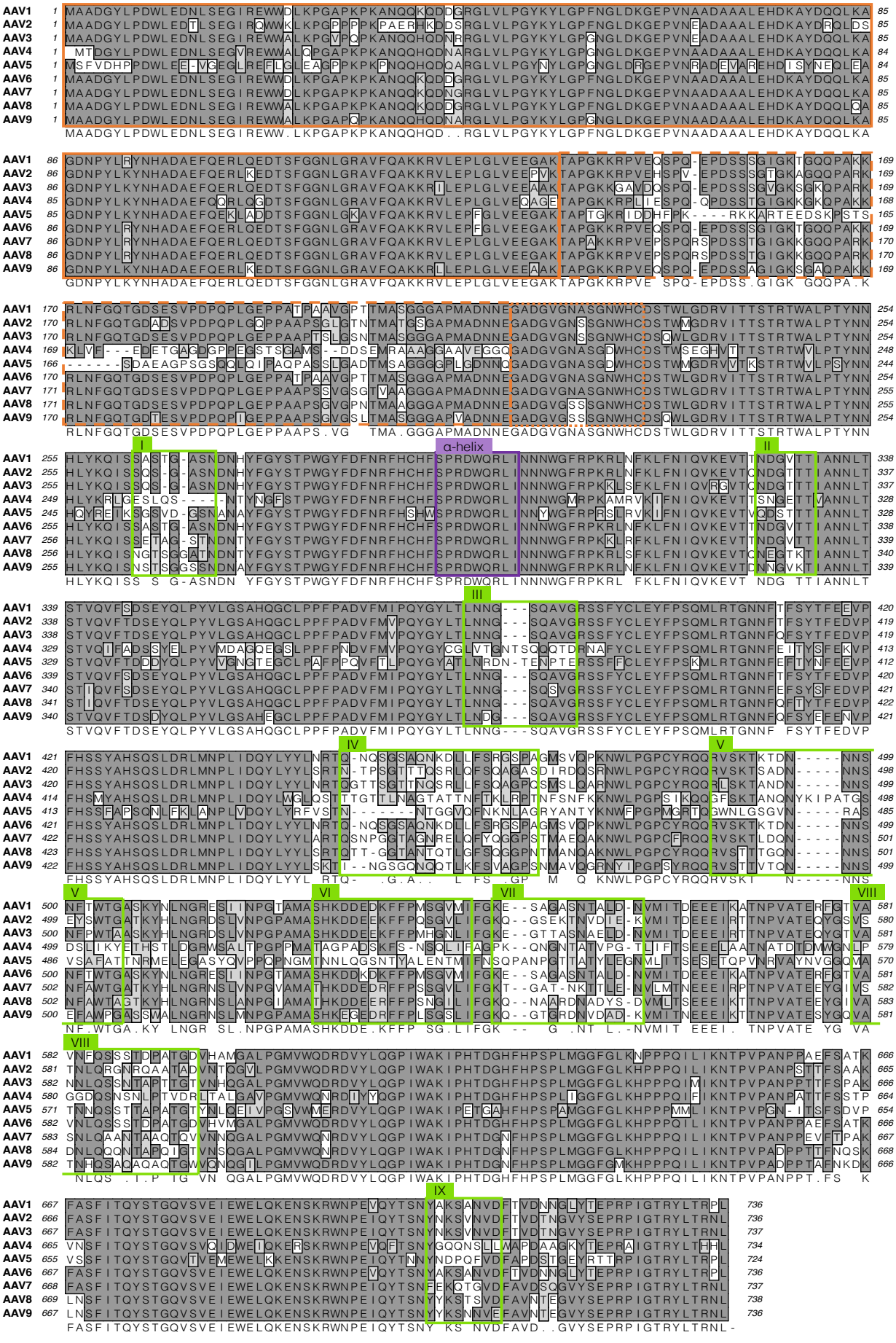

AAV1
AAV2
AAV3
AAV4
AAV5
AAV6
AAV7
AAV8
AAV9
I
α-helix
II
III
IV
V
VI
VII
VIII
IX

Figure 1.7 Alignment of the VP1 sequences of AAV1-9

The VP1 amino acid sequences of selected AAV serotypes were aligned and the sequence consensus is shown in dark grey text. Conservative amino acid changes are shown in light grey shading and deviations are shown with white background. VP1-3 N-terminal regions are marked in orange; the VP1 unique region with a continuous line, VP1/VP2 N-terminal shared region with a discontinuous line and the VP3 N-terminal region with a dotted line. The loops that are variable between AAV serotypes and shape AAV serotypes surfaces are marked in green [242]. The variable regions I-X are associated with AAV entry receptor binding, but also with adaptive immune response. Variable regions V, VI, VIII, IX have been implicated in heparan sulphate binding for heparan sulphate binding serotypes [238, 245, 246], R585 and R588 in region VIII are particularly crucial for heparan sulphate binding [222]. Variable region IV, V and VIII have been associated with sialic acid binding [247, 248]. The α-helix motif that is highly conserved within the Parvoviridae family is marked in purple.

The transduction specificity of rAAV vectors is mediated by the expression of viral entry receptors on the cell membrane and intracellular host factors. For many AAV serotypes protein co-receptors have been identified, some of which are indispensable, such as the universal AAV receptor KIAA0319L (AAVR, [249]). More recently, GPR108 was identified as a highly conserved AAV entry factor [250]. However, we probably do not know all receptors involved in AAV cellular entry and entry receptor expression throughout the human lung is also not yet fully understood.

Because viral cell entry is host specific, investigations of rAAV tropisms in animal models often do not translate well to other species and (see Section 1.4.3), even if it does translate, it remains unclear whether this is due to the same or different entry receptors. Even selection in animal models thought to have a highly similar distribution of entry receptors compared with human lung, such as ferret, can lead to unrepresentative results for human [251]. Further, the primary entry receptor sialic acid which is relevant to many serotypes is variable between airways and lung parenchyma as well as between the apical and basolateral lung surfaces [183, 248, 252]. Therefore, AAV serotypes must be carefully selected in appropriate models with the distribution of entry receptors in humans in mind. As we likely do not know all entry receptors and cellular factors in viral entry, information on viral entry should be used for negative rather than positive selection. For instance, AAVR is not expressed on the apical surface of human airway epithelia, which is an argument against the use of AAVR-dependent serotypes to target human airways, such as AAV2 used in cystic fibrosis clinical trials (Section 1.4.3).

AAV serotypes 1-13 and their entry receptors are shown in Table 1.1. Amongst AAV serotypes, AAV1, 2, 4, 5, 6, 8, and 9 are the best characterised and, for all of these, clinical studies are underway. Most AAV serotypes transduce many tissues to some

degree and, due to the natural tropism of AAV, most serotypes transduce liver and spleen to some extent. A selection of tropisms strongly associated with the vector is shown in Table 1.1, with a focus on human and NHP data where available, with clinical trial numbers and approved drugs also indicated where applicable. To identify potential candidates for SP-B gene therapy, references to lung tropism are listed separately with the tested species and their administration method. The rAAV2.CFTR trials have spiked interest in rAAV airway tropism, where the disease associated CFTR transmembrane protein is located [6], but only a few studies investigated rAAV tropism to the lung parenchyma and these were not conducted in humans or NHP (Table 1.1). In NHP studies, lung transduction was explicitly assessed for airways or more generally assessed for lung tissue by methods like luciferase imaging or deposition of viral DNA [253, 254], without fully clarifying transduction of airway versus parenchymal cell types [255], therefore remained inconclusive. Because of differential distribution of entry receptors between airways and lung parenchyma, and between species, predictions about rAAV tropism to the lung parenchyma remain challenging.

Table 1.1 Naturally occurring AAV serotypes, their receptors and tropisms

	Glycan Receptor (Primary Receptor)	Co-receptor	General tropism	Reference to lung tropism
AAV1	α-2,3-N-linked SA α-2,6-N-linked SA [256, 257]	AAVR [249]	CNS [258] retina (Glybera) skeletal muscle [259] heart [260] (NCT01643330) lung	Airway, HAE, apical [181] Lung homogenates, murine, luminal [181] Airway and lung parenchyma, pig, luminal (nebuliser) [261] Airway, NHP, luminal [254] Lung, NHP, luminal [255] Lung, perinatal sheep, luminal [262]
AAV2	HS [263]	AAVR [249] HGFR [264] FGFR [265] LR [266] Integrin [267] CD9 tetraspanin [268]	retina (Luxturna) skeletal muscle [269, 270]	Airway, HAE, apical [181, 248] Airway, clinical trial, luminal [178] Airway and lung, NHP, luminal [253]
AAV3	HS [271]	AAVR [249] LR [266] FGFR [272] HGFR [273]	less well characterised because inefficient transduction in many tissues [159]. hepatocellular carcinoma [274]	-
AAV4	α-2,3-O-linked SA [275]	unknown (not AAVR) [276]	retina [277] (NCT01496040)	Lung homogenate, murine, luminal [159]
AAV5	α-2,3-N-linked SA [275, 278]	AAVR [249, 279] PDGFR [280]	CNS [281] retina [277] lung	Airway, HAE, apical [181, 248, 282] Lung homogenates, murine, luminal [181] Lung parenchyma, murine, luminal [218, 283]
AAV6	α-2,3-N-linked SA,	AAVR [249]	neuronal [285]	Lung parenchyma, murine, luminal [283]

	α-2,6-N-linked SA, HS [238, 256, 257]	EGFR [284]	lung heart [286]	Airway, HAE, apical [287]
AAV7	Unknown	Unknown	CNS [288] liver [289]	-
AAV8	Unknown	AAVR [249] LR [266]	liver [289], (NCT03173521 & NCT00979238) skeletal muscle [290, 291]	Lung, perinatal sheep, luminal [262]
AAV9	N-linked Galactose [292]	AAVR [249] LR [266]	broad tropism, including: skeletal muscle [290], (Zolgensma) CNS [293, 294] (NCT02725580)	Airway, HAE, apical [248] Lung, perinatal sheep, luminal [262] Lung parenchyma, murine, luminal [283]
AAV10	Unknown	Unknown	skeletal muscle [295, 296] CNS [297]	-
AAV11	Unknown	Unknown	(less well characterised) skeletal muscle (mouse, IM) [295] kidney, spleen, lung, heart, stomach (mouse, IV) [295]	-
AAV12	Unknown	Unknown	(less well characterised) muscle (mouse, IM) [298]	-
AAV13	HS [299]	-	unknown	-

Abbreviations: SA, sialic acid; HS, heparan sulphate; AAVR, AAV receptor; EGFR, epidermal growth factor receptor; FGFR1, fibroblast growth factor receptor 1; HAE, human airway epithelial culture; HGFR, hepatocyte growth factor receptor; PDGFR, platelet-derived growth factor receptor; CNS, central nervous system; IV, intravenous; IM, intramuscular.

AAV serotype tropisms with other foci such as clinical trials and other animal models are reviewed in [202, 214, 215, 300-303]. AAV co-receptors are reviewed in [304, 305].

1.5.5 AAV capsid engineering

Aside from the natural discovery of AAV serotypes from human, NHP and other species, AAV capsids can be engineered to improve transduction of target tissues, increase transgene expression and evade pre-existing immunity. One approach is rational serotype design. The crystal structures and the amino acids shaping vector topology (as in Figure 1.7, above) are known for most AAV serotypes. Amino acids essential for glycan receptor binding have been identified through mutational studies. This knowledge has made rational design possible. One common approach is exchanging elements between AAV serotypes with the hope to combine the favourable characteristics of two serotypes. The combination of serotype 2 with five capsid mutations from serotype 1 resulted in improved muscle transduction over AAV2 and reduced antigen cross-reactivity with the parental serotypes (AAV2.5, [306]). The exchange of a hexapeptide sequence in AAV2 which is associated with heparan sulphate (HS) binding with AAV8 capsid sequence, retargeted the vector to skeletal muscle (AAV2i8, [307]). Similarly, for AAV6 which performed well in mouse airway and human airway epithelial culture, singleton changes from AAV1 were made to improve transduction efficiency. These studies resulted in the generation of serotype AAV6.2 (AAV6 with one singleton change F129L, [244, 283]).

Most efforts aim to improve binding to the cell surface of the target cells, but as discussed in chapter 1.5.2, cellular entry is not the only component crucial for successful transduction, and sequences thought to be involved in vector stability, cellular trafficking and efficient AAV genome transcription also influence transduction efficiency. This was shown for the AAV2 ‘dead zone’, in which mutations can decrease transduction more than 1000-fold [308, 309]. One rational design example to improve intracellular trafficking is the replacement of tyrosines that are prone to phosphorylation, causing alternative trafficking, ubiquitination and proteasomal degradation. The replacement of two tyrosines with

phenylalanine (Y445F and Y731F) in a rAAV6.2 vector resulted in improved transduction of murine lungs and muscle and rapid onset of expression compared to AAV6 [310].

Bioinformatic *in silico* capsid design can aid the rational approach. *In silico* analysis identified amino acid positions suitable for mutation by employing information about the VP variable regions, shuffling mutations from 150 AAV variants and evaluation using a 3D AAV model [311]. Further, *in silico* analysis of naturally occurring serotypes and variants lead to the phylogenetic reconstruction of an ancestral AAV, Anc80, which efficiently transduced liver, muscle and retina in mice after *in vivo* delivery and showed limited cross-reactivity to common AAV serotypes [312].

Directed evolution is a capsid design approach that exploits natural selection in an experimental setting. Random AAV capsid libraries are generated to produce a multitude of different rAAV vectors and natural selection occurs via transduction of a model system [313]. This process is repeated until suitable vector candidates emerge. The advantage of directed evolution is that it can generate capsids beyond the current, still limited understanding of rAAV, as library generation aims to be random. Popular methods of library generation are capsid shuffling, in which fragments of AAV capsids are combined, and error-prone PCR [313, 314]. These methods can also be combined to create a library as close as possible to random distribution. Bocavirus from the *Parvoviridae* family was shown to cross-package AAV2 genomes and was combined with AAV capsid sequences in DNA shuffling to diversify the library [315]. Directed evolution resulted in the generation of the serotype rAAV2.5T which transduced human airway epithelial cultures more efficiently than naturally occurring AAV serotypes [248]. The models employed for capsid selection must be carefully chosen as eventual adaptation tends to be highly specific. The vector rAAV2.5T did not effectively transduce lung airways in a porcine model of cystic fibrosis after selection in human airway epithelial cultures and *vice versa*; subsequent

directed evolution for the porcine airway was not able to transduce human airway epithelial cultures [316, 317]. Moreover, the enhanced transduction of the CNS by vector rAAV.PHP.B in C57BL/6J mice, was not even translatable to a different mouse strain [226].

In summary, AAV serotypes can be engineered via rational and directed evolution approaches to evade the immune system and enhance specific transduction and expression in target tissues. Host factors for transduction, such as viral entry receptors, vary between species and between cell types of one organ, therefore, the model for capsid selection should be carefully chosen and verification of isolates in an additional model considered.

1.6 Previous studies on gene therapy for SP-B deficiency

1.6.1 Studies *in vivo*

As SP-B deficiency is a monogenic disease and surfactant replacement therapy proved to be ineffective, gene therapeutic approaches are being considered. In 1995, Korst *et al.* delivered SP-B and SP-A expressed from rAdV, under the control of the CMV promoter, to the lungs of wildtype cotton rats resulting in detection of SP-B mRNA in the lung homogenates via Northern blot analysis [318]. Yei *et al.* performed a similar experiment to deliver SP-B with rAdV using a different constitutive promoter and showed SP-B mRNA expression *in vivo* by Northern blot and *in situ* hybridisation and SP-B protein expression detected by radiolabelling in transduced cells [319]. Both of these studies show successful viral gene transfer for SP-B *in vivo*, but are unsuitable for therapeutic long-term expression, largely because the rAdV vector is strongly immunogenic leading to clearance of transduced cells by the immune system thus preventing long-term expression.

More recent approaches to SP-B gene therapy *in vivo* have utilised the conditional SP-B knock-out mouse model of SP-B deficiency developed by Melton/Whitsett and

colleagues [320]. This mouse model allows for breeding and *in vivo* investigation of SP-B deficiency which is lethal in neonates. The SP-B$^{(-/-)}$ mice express murine SP-B cDNA controlled by the conditional tetracycline promoter such that the activator protein (reverse tetracycline transactivator, rtTA) is expressed specifically in ATII cells under the human *SFTPC* promoter. In these mice, SP-B is expressed in ATII cells as long as doxycycline is added to their food, but upon doxycycline withdrawal the mice die after ca. 7 days [320].

Using the SP-B deficiency mouse model, Barnett *et al.* improved survival by electroporating human SP-B cDNA in plasmid form to the lungs with the SP-B transgene expressed from the CMV or UBC promoter [321]. The CMV.SP-B plasmid had the greater impact on survival, extending it by five-fold (median survival 25 days). A gene therapy approach using chemically modified mRNA for SP-B expression also increased survival [322]. Intratracheal dosing twice a week with naked chemically modified mRNA restored SP-B mRNA expression levels to 71 % compared with wildtype, with the majority of mice surviving until the predetermined endpoint of the study at 28 days after doxycycline withdrawal (average life span 26 days). Mahiny *et al*. targeted the inducible SP-B expression cassette in the Melton/Whitsett mouse model to introduce the constitutive CAG promoter allowing for doxycycline independent expression [323]. Delivery of ZNFs with chemically modified mRNA and CAG promoter donor with rAAV6 for HDR significantly prolonged survival compared with control groups.

These various studies, conducted in the Melton/Whitsett mouse model, are important proof-of-concept studies showing that gene therapy can correct the SP-B deficiency phenotype and prolong survival *in vivo*. However, the delivery methods presented in their current form are not transferrable to the clinic since plasmid electroporation to the lungs of neonates, especially if requiring repeated delivery, will be difficult to translate clinically. The mRNA administration was performed as twice-weekly dosing of anaesthetised

animals via intratracheal high-pressure spraying, likely required for the lifetime of the animal; and again this would be difficult to translate to patients. The delivery strategy used by Mahiny *et al*. [323] could be feasible in a clinical setting and, theoretically, would not require repeat administration, but the genetic target is engineered and not present in individuals with SP-B deficiency.

1.6.2 Studies in organoids

Organ-like 3D cell cultures known as organoids are emerging models in basic and applied research. Organoids can be grown from embryonic stem cells (ESC), induced pluripotent stem cells (iPSC) or adult organotypic stem cells. Jacob *et al*. corrected the 121ins2 mutation in iPSCs from a subject homozygous for this mutation by electroporation of Cas9 and sgRNA as a plasmid and ssDNA donor. The Cas9 induced a DSB which was corrected via HDR using the ssDNA donor encoding the wildtype sequence [324]. Subsequently, 3D monolayered 'alveolospheres' were grown from 121ins2-corrected and uncorrected iPSC cell lines showing SP-B expression and characteristic LBs. In a similar study, iPSC from a subject homozygous for 121ins2 were transduced with rLV vector expressing SP-B using the CMV promoter [325]. The transduced and untreated iPSC cell lines were differentiated into more complex lung organoids containing cell populations of the proximal and distal lung. Both of these studies highlight lung organoids as a powerful human lung model capable of modelling hallmarks of SP-B deficiency and showing phenotype correction by gene therapy. Neither study, however, is able to model the delivery challenges of *in vivo* gene therapy, which will require delivery to the lung of neonates. Further, gene therapy for SP-B deficiency is particularly specific, as it requires delivery to ATII cells of the lung parenchyma, which are the only cell types equipped for efficient processing of SP-B to its mature form and package it in LBs to produce functional surfactant. The studies discussed above are important steps in the development of a gene therapy for SP-B deficiency, but have not yet resulted in a truly translatable approach.

1.7 Aims

Surfactant Protein B deficiency is an ultrarare genetic disease that is invariably fatal, with no treatment option other than lung transplantation or compassionate care. To improve prognosis, we propose rAAV-based gene therapy to the parenchyma of the neonatal lung. This is a truly translatable approach, as rAAV has been used safely in clinical studies in the lung previously. The overall goal to develop a rAAV based gene therapy for SP-B deficiency was split in three aims:

1. The first aim of this study is to identify a suitable rAAV-based vector for SP-B deficiency gene therapy in mice, by testing AAV serotypes suitable for transgene delivery to the surfactant-producing ATII cells and identifying SP-B expression cassettes for high levels of SP-B expression in the target ATII cells (Chapter 3).
2. The second aim is the investigation of the gene therapy serotypes and transgene cassettes in human primary cell models to verify the rAAV vector selected in mice (Chapter 4).
3. The third aim is the creation of a human disease model for SP-B deficiency with the potential to evaluate rAAV-based gene addition and gene editing strategies (Chapter 5).

Chapter 2 – Materials and Methods

2.1 Common reagents and solutions

All buffers were prepared using Milli-Q reverse osmosis ultrapure water (Merck Millipore) unless stated otherwise. All incubations, unless stated otherwise, were performed at room temperature (RT). During wash steps the previous solution was always first removed before new buffers were added unless stated otherwise. Where concentrations are stated as percentages [%], these refer to weight per volume (w/v) ratios unless otherwise specified.

2.2 Plasmids and molecular cloning

Plasmids used in this study are listed in Table 2.1, Section 2.2.7. New plasmids were constructed either via ligation of restriction fragments from existing plasmids, from PCR amplified products or from synthesised fragments obtained from Gene Art Services (Thermo Fisher Scientific). Procedures for plasmid construction and amplification of starting material for molecular cloning (Sections 2.2.1-2.2.4), transfection (Section 2.4.2) and rAAV production (Section 2.5.1) are detailed in the following sections. A typical molecular cloning workflow consisted of digestion of the plasmid backbone and DNA insert with the same or compatible restriction enzymes to create sticky ends (Section 2.2.2), followed by purification via gel extraction (Section 2.2.3) and ligation (Section 2.2.4). The ligation product was used to transform bacteria which were plated on selective agar plates (Section 2.2.5). Finally, plasmids from bacterial clones were amplified and analysed for the correct DNA sequence (Section 2.2.6).

2.2.1 Polymerase chain reaction

DNA was amplified via the polymerase chain reaction (PCR) using Phusion HotStart II DNA Polymerase (Thermo Fisher Scientific). Typically, a reaction included 200 μM dNTPs, 0.5 μM of forward and reverse primer and 0.01-0.02 U/μL polymerase, diluted in 1x Phusion HotStart II buffer. For amplification of genomic DNA, 50-100 ng template were added, while for plasmid DNA templates 10 ng were added. Primers were designed using an online tool (primer3.ut.ee) and screened for potential off-targets (www.ncbi.nlm.nih.gov/tools/primer-blast). Primer melting temperatures were calculated using the T_M calculator online tool matching the polymerase (Thermo Fisher Scientific). Cycling conditions were designed according to the manufacturer's instruction, taking into consideration the calculated primer T_M and size of the desired PCR product. PCR products were examined via agarose gel electrophoresis (Section 2.2.3).

2.2.2 Restriction enzyme digest

Restriction enzyme digests were performed for molecular cloning or for diagnostic digests. All restriction enzymes were purchased from New England Biolabs and used according to the manufacturer's instructions regarding buffer, incubation time and temperature. Typically, between 1-5 μg DNA in a volume of 20-50 μL were digested for 2-3 h with 10-20 units of enzyme.

2.2.3 Gel electrophoresis and extraction

DNA was analysed via agarose gel electrophoresis. Agarose was dissolved by boiling in 1x Tris-acetate-EDTA buffer (Thermo Fisher Scientific). SYBR Safe DNA gel stain (Thermo Fisher Scientific) was added to the 1-2 % agarose/buffer solution, which was poured into a suitable mould containing a removable comb to create sample wells, before being left to cool and set. Agarose gels were placed in a suitable submerged electrophoresis tank filled with 1x Tris-acetate-EDTA buffer. DNA samples were mixed

with loading dye (New England Biolabs), pipetted into sample wells and electrophoresed at 80-100 V for 45-60 min. DNA bands were visualised via illumination at 302 nm using the iBright FL1500 imaging system (Thermo Fisher Scientific). Fragment size was estimated in relation to a 1 kb or 100 bp DNA ladder (New England Biolabs).

For purification of DNA fragments after restriction digests, the desired fragments were excised from the gel using a scalpel while visualising them with the iBright FL1500 imaging system (Thermo Fisher Scientific) at a wavelength of 490-520 nm. Fragments were extracted and purified using the Qiaquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions. The same kit was also used for purification of PCR products, where separation of the PCR product on a gel was optional.

2.2.4 Ligation

Purified DNA fragments were ligated using T4 DNA ligase (New England Biolabs). Each reaction consisted of vector backbone, insert, ligase and its supplied buffer. Control reactions omitted the addition of insert. Typically, the ratio of vector to insert ranged from 1:3 to 1:7 with 1-10 ng/μL DNA per 10 μL reaction. The ligation reaction was incubated for 10 min and subsequently transformed in chemically competent *E. coli* (see Section 2.2.5).

2.2.5 Plasmid amplification

Plasmids were amplified from isolated bacterial colonies picked from selective agar plates and grown in Luria Bertani broth (LB broth), both supplemented with appropriate antibiotics (100 μg/mL Ampicillin or 50 μg/mL Kanamycin). LB broth consisted of an autoclaved solution of 1 % Tryptone (VWR), 1 % NaCl and 0.5 % granulated yeast extract (Merck) in water. To prepare antibiotic selective agar plates, LB broth was supplemented with 1.5 % of Select Agar (Invitrogen) before autoclaving and 100 μg/mL Ampicillin or 50 μg/mL Kanamycin was added after autoclaving and cooling.

Plasmids or ligation products were first transformed using One Shot™ Stbl3™ Chemically Competent *E. coli* (Thermo Fisher Scientific) according to manufacturer's instructions. For plasmid retransformation, less material (ca. 0.1-10 ng) was used compared with ligation reaction products. A single colony was picked from a selective agar plate to inoculate 5 mL or 400 mL LB broth dependent on desired yields. Alternatively, 5 mL cultures served as a starter culture for 400 mL cultures. Bacterial cultures were incubated overnight at 37 °C in LB broth supplemented with the appropriate antibiotic in an orbital shaking incubator (180-250 rpm). Plasmids were purified either using the EndoFree Plasmid Mega Kit (Qiagen) or Wizard Plus SV Miniprep kit (Promega), according to manufacturer's instructions. Purified plasmids were diluted in nuclease-free water to a concentration of 2 mg/mL and stored at - 20 °C.

2.2.6 Plasmid analytics

Following plasmid amplification, products were visually inspected after gel electrophoresis. Typically, multiple restriction digests were used to assess fragment lengths and to confirm the identity of a plasmid. In newly constructed plasmids, the expression cassette was subsequently sequenced in Sanger sequencing reactions (see Section 2.3). The NanoDrop Spectrophotometer 2000 (Thermo Fisher Scientific) was used for assessment of plasmid concentration.

2.2.7 Plasmids in this study

Plasmids used for rAAV and rLV production, as well as for transgene expression, are summarised in Table 2.1.

Table 2.1 Plasmids

Plasmid name	Plasmid number	Promoter	Transgene	Source
rAAV genome plasmids for transgene expression				
pAAV2.hCEFI.EGFP	pGM505	hCEFI	EGFP	Gift from Dr Tiong Kit Tan
pAAV2.CMV.EGFP	pGM546	CMVie	EGFP	Gift from Dr Barnabas Williams
pAAV2.CMV.SP-B	pGM705	CMVie	SP-B	this study
pAAV2.CMV.soSP-B	pGM706	CMVie	soSP-B	this study
pAAV2.hCEFI.SP-B	pGM707	hCEFI	SP-B	this study
pAAV2.hCEFI.soSP-B	pGM708	hCEFI	soSP-B	this study
pAAV2.SP-B MG	pGM709	SP-B	SP-B	this study
pAAV2.soSP-B MG	pGM710	SP-B	soSP-B	this study
Accessory plasmids for rAAV production				
pAd Delta F6	-		Adenoviral replication genes E4, E2a and VA	Penn Vector Core
pAAV2/X	-	p5	AAV2 *rep* gene and *cap* gene from AAV serotype 1, 2, 5, 6, 8 or 9	Penn Vector Core
pAAV2/6.2	-	p5	AAV2 *rep* gene and *cap* gene from AAV serotype 6 with mutation F129L	[244] Distributed by Penn Vector Core
pAAV2/6.2FF	-	p5	AAV2 *rep* gene and *cap* gene from AAV serotype 6 with mutation F129L, Y445F and Y731F	[310] Gift from Omar Habib
Plasmids for rLV production				
pCBA.SIV.GagPol	pGM297	CBA	*Gag* and *pol* genes from SIV	Gift from Mary Connolly
pCAG.Rev	pGM299	CAG	*Rev* gene from SIV	Gift from Ian Pringle

pCAG.SeV.Fct4	pGM301	CAG	Sendai virus F envelope protein	Gift from Mary Connolly
pCAG.SeV.EnvctHN	pGM303	CAG	Gene for Sendai virus HN envelope protein	Gift from Mary Connolly
pSIV.CMV.EGFP	pGM311	CMV	EGFP	Gift from Mary Connolly
pSIV.hCEF.EGFP	pGM357	hCEF	EGFP	Gift from Mary Connolly

2.3 Sanger sequencing

Sanger sequencing of purified PCR products or plasmids was performed by Source BioScience. Plasmid templates were provided at 100 ng/μL and PCR products at 10 ng/μL, 5 μL per sequencing reaction. Primers for the sequencing reaction were designed according to Source Bioscience's requirements and supplied at 3.2 pmol/μL, 5 μL per rection. In GC-rich rich amplicons, dGTP chemistry was typically used to improve the quality of sequencing data.

2.4 Mammalian Cell Culture

2.4.1 Lines and maintenance

Mammalian cell lines were cultured in adherent cell culture conditions in media as listed in Table 2.2 in humidified incubators (5 % CO_2, normoxic, 37 °C). Immortalised cell lines were passaged between 1:5 and 1:15 biweekly at ca. 80-90 % confluence. Human HEK293T cells [326] were detached by incubation in 1x PBS at 37 °C for 5 min. For culture of human A549 cells [327], 0.25 % Trypsin (Gibco) in 1x PBS was used for cell detachment. For cryopreservation of immortalised cells, cells were washed and detached as described above, resuspended in culture media supplemented with 10 % DMSO and frozen in freezing containers at - 80 °C. After a minimum of 16 h, vials were moved to liquid

nitrogen tanks. Typically, immortalised cell lines were cultured up to passage 30 before discarding.

Human RUES2 cells [328] were passaged between 1:5 and 1:20, ca. biweekly at 80-90 % confluence. Colonies showing signs of differentiation were manually removed by scratching from the culture plate before passaging. For standard culturing, cells were washed in 1x PBS and incubated with 1x Dispase (Stem Cell Technologies) at 37 °C for 6 min. Cells were washed in the culture plate for three times in 1x DPBS. On the final step new culture medium was added, cells removed using a cell scraper and clumps broken-up to ca. 20-50 cells by gentle pipetting. Finally, the cell suspension was distributed in Matrigel coated wells. For Matrigel-coating, tissue culture plates were incubated with 0.5 mg Matrigel (Matrigel Growth Factor Reduced Basement Membrane Matrix, #356230, Corning) diluted in 6 mL DMEM/F-12 at 37 °C overnight; 1 mL per well in a 6-well plate. Accutase (Stem Cell Technologies) was used to passage very large colonies (ca. 2 mm diameter) or achieve cells suspension with smaller clumps (10-20 cells) or single cell suspension. Cells were washed one in 1x DPBS and incubated with 1x Accutase at 37 °C for 8 min. The cell suspension was diluted with culture media (1:2), centrifuged (120 g/5 min) and the pellet resuspended in new culture media. For cryopreservation, cells were frozen in mFreSR™ medium (Stem Cell Technologies) according to manufacturer's instructions. RUES2 cells were tested for mycoplasma every 6 months using a mycoplasma PCR testing kit by Rebecca Dean (Applied Biological Materials Inc.).

Table 2.2 Cell lines

Name	Type	Source	Media
A549	Lung carcinoma cell line (immortalised)	ATCC (CCL-185)	F-12K media (Gibco), 10 % heat-inactivated FBS, 100 U/mL penicillin, 100 mg/mL streptomycin, 5 mL L-glutamine (1x final)
HEK293T	Human embryonic kidney containing the SV40 T-antigen cell line (immortalised)	ATCC (CRL-11268)	High glucose DMEM (Gibco), 10 % heat-inactivated FBS, 100 U/mL penicillin, 100 mg/mL streptomycin, 5 mL L-glutamine (1x final)
RUES2	Rockefeller University embryonic stem cell line 2 (primary)	WiCell hPSCregID RUESe0002-A	mTeSR/mTeSR Plus (Stem Cell Technologies), Primocin (1:1000, Invivogen)

2.4.2 Transfection

For transient transfection with plasmid DNA, adherent mammalian cells were seeded in 6-well plates. At 80-90 % cell confluence, 2 μg plasmid DNA complexed with 6 μL Lipofectamine 3000 (Thermo Fisher Scientific) was added per well, both diluted in OptiMEM (Gibco) and complexes prepared according to manufacturer's instructions. Cells were analysed 24-48 h after transfection. For transient transfection of plasmid DNA for viral vector production see Section 2.5.

For chemical transfection of Cas9 RNP and sgRNA (both Synthego), HEK293T cells were similarly treated as above, but using the Lipofectamine™ CRISPRMAX™ reagent (Thermo Fisher Scientific) according to 'CRISPR Editing of Immortalized Cells with RNPs using Lipofection Protocol' by Synthego. Briefly, Cas9 (*Streptococcus pyogenes* origin) and sgRNA were complexed according to manufacturer's instructions. Cells were transfected at 80 % confluence by reverse transfection with 3.9 pmol sgRNA and 3 pmol Cas9 per well in a 24-well format and genomic DNA extracted at day 4 post transfection for downstream analysis.

2.4.3 Creation of a CRISPR/Cas9-edited stem cell line

2.4.3.1 Electroporation of stem cells for Cas9 RNP delivery

Cas9 RNPs were electroporated in RUES2 embryonic stem cells (see Section 2.4.1) to achieve targeted gene editing in *SFTPB* using the P3 Primary Cell 4D-Nucleofector™ X Kit S and 4D-Nucleofector set up, including X unit (all Lonza). RUES2 cells growing in small colonies (ca. 50-200 cells per colony) were selected for use on the day of electroporation, as these were typically found to efficiently generate the single cell suspension necessary for the procedure.

First, sgRNA was rehydrated and subsequently complexed with Cas9 protein (*Streptococcus pyogenes* origin) for 10 min according to manufacturer's instructions (both Synthego). Electroporation of plasmid DNA expressing GFP, supplied with P3 Primary Cell 4D-Nucleofector™ X Kit S (Lonza) served as positive control for transfection. The nucleofector solution (Lonza) was prepared by mixing nucleofection solution with the supplement according to manufacturer's instructions. A single-cell suspension of RUES2 cells was prepared using Accutase (Stem Cell Technologies, see Section 2.4.1). If a single cell solution was not achieved after 5-8 min of Accutase treatment, the cells were discarded and where possible smaller colonies were used. Cells were counted and centrifuged (115 g / 3 min) to facilitate complete removal of all media/Accutase. Cells were resuspended in the prepared nucleofector solution (Lonza), mixed with CRISPR editing reagents, swiftly added to Nucleocuvette™ vessels and electroporated using the CA137 programme in the 4D-nucleofector. One reaction included 300 000 cells resuspended in 20 μL nucleofector solution, 300 pmol sgRNA, 40 pmol Cas9 protein and, in case of HDR gene editing in exon 4, 300 pmol donor DNA (G*G*GTGGGCACAGGGGCCTGCGT GGGGAGGCAGGCAGGAGTGAGCTTGCAGCCCTCACAATCTGGTTCTGGAAGTAGT CGATGACCAGGGTTCGAAGTAGTCGTCAAGCACTTGGTTGCACTGGGGC*A*T). The

127 bp donor DNA was designed complimentary to the targeted strand with asymmetric homology arms (38 and 89 bp from cut site). When more than one sgRNA was electroporated, equal parts of each sgRNA were added summing to 300 pmol total. The duration of cellular exposure to the nucleofection solution was kept to a minimum, ideally less than 8 min elapsed from resuspension in nucleofection solution to recovery in culture media. For resting cultures, cells were cultured in a Matrigel-coated 24-well plate with prewarmed mTeSR1 medium supplemented with 10 μM Y-27632 (Rho-associated, coiled-coil containing protein kinase 'ROCK' inhibitor) until cells were dense enough for passaging (24-96 h).

Table 2.3 Single guide RNAs for editing in *SFTPB*

Name	Target	Sequence	PAM
exon 4_121ins2	*SFTPB*, exon 4	GCUUGACGACUACUUCCCCC	TGG
exon 8_40fw	*SFTPB*, exon 8	UCAUGGUCAGGGCCCGUUCA	AGG
exon 7_126rev	*SFTPB*, exon 7	GGCAAGGCUAUUCACAAAUA	TGG
exon 8_91fw	*SFTPB*, exon 8	CAGCCAGAGACACCUAGGAU	GGG
exon 7_41rev	*SFTPB*, exon 7	GCCAGCUUCAGACCUACGUC	TGG

2.4.3.2 Analysis of editing efficiency

Pools of edited RUES2 stem cells were analysed for editing efficiency upon the first passage after recovery culture. DNA from a proportion of treated cells was extracted using the QIAamp DNA Mini Kit according to the manufacturer's instructions. Regions of interest were amplified using PCR with primers from Table 2.4 and purified via gel extraction (see Section 2.2.3). Amplicons were analysed via Sanger sequencing (see Section 2.3) using the Synthego ICE v2 analysis tool (https://ice.synthego.com). The *Interference of CRISPR edits* (ICE) tool is a bioinformatics algorithm based on *Tracking Indels by Decomposition*

(TIDE) that predicts frequency and position of edits which have been shown to strongly correlate with next generation sequencing data [329].

2.4.3.3 Isolation of a gene edited stem cell clone

Single cell clones were isolated by limiting dilution. Edited RUES2 cells were dissociated to single-cell solution using Accutase (see Sections 2.4.1 and 2.4.3.1). Cells were diluted to 0.7 cells/100 µL in mTeSR1 media with CloneR supplement (Stem Cell Technologies) and seeded 100 µL/well in a Matrigel-coated 96 well plate. Medium was changed to standard mTeSR medium after 48-96 h and then every 2-3 days until colonies, clearly visible without magnification, emerged after 10-14 days.

Subsequently, clones were expanded and genotyped. Each clone was passaged using Accutase and 20 % of the cells seeded in 48-well and the remaining was used for genotyping. DNA was extracted using QuickExtract™ DNA Extraction Solution (Lucigen). Briefly, 50 µL of QuickExtract™ solution were added to each cell pellet and heated to 65 °C for 6 min, followed by 98 °C for 2 min in a PCR machine. Samples were kept at - 20 °C until analysis.

For genotyping, the regions of interest were amplified via PCR using AmpliTaq Gold™ 360 2x Master Mix (Thermo Fisher Scientific). Each 50 µL reaction contained 25 µL AmpliTaq Gold™ 360 Master Mix, 0.2-1 µL template and 0.5 µM of each amplification primer. Amplification primers are shown in Table 2.4. In case of analysis for the large excision by gene editing between exons 7 and 8, the exon 7 forward and exon 8 reverse primer were used. Clones were screened for the large deletion by gel electrophoresis. For editing of 121ins2 in exon 4, successful gene editing created a BstB1 restriction site. Amplicons were analysed by restriction fragment length analysis. Briefly, ca. 20 % of the PCR product was digested using 0.5 µL BstB1 enzyme diluted in CutSmart buffer (both New England Biolabs) at 65 °C for 4 h. An amplicon of genomic DNA with an endogenous BstB1

restriction site served as control for successful and complete digest (see 'Chr2p 84749902', Table 2.4). Sanger sequencing confirmed the successful editing of clones showing complete digestion of the amplicon. During expansion, clones were sequenced again and finally cryopreserved using mFReSR™ or CryoStor® CS10 cryopreservation medium (both Stem Cell Technologies).

Table 2.4 Primers for analysis of gene editing in *SFTPB*

Region	Amplification forward	Amplification reverse	sequencing
SFTPB, exon 4	GGTAATGATGCC CAGATCCTG	TTTGTGTCTGGCC GGCTT	CACAGCCAGAGATGGTC TT
SFTPB, exon 7	GATGCATTTGGG AGGTGGG	CACAGAGGGAGAG GGTCATG	For:AACAGAGAGCTCCCT CCTG Rev:ATGACGGAGTAGCG CTCA
SFTPB, exon 8	CATGACCCTCTC CCTCTGTG	CCCTCTCCTTTTGC CAATGG	For:CCTAATGCTCACATC TCCAG Rev:AGATAACCAGGCCTC TGAAG
Chr2p 84749902	TACCCTCTGCAG TGTGAGTG	AAGTGCTAGGCTT ACAGGGG	-

2.5 Viral vector production

2.5.1 Recombinant adeno-associated viral vector production

Recombinant AAV vectors were produced with the triple transfection method described previously [198, 330]. Adherent HEK293T cells were cultured in 1720 cm^2 HYPERflasks (Corning). Two HYPERflasks were typically used for the preparation of one rAAV vector batch. For seeding, HEK293T cells were grown in T175 flasks to 80 % confluence (ca. 3.4E7 cells/flask). One T175 flask at 80 % confluence was needed to seed one HYPERflask. On day 3 after seeding, when usually 80 % confluence was reached in the HYPERflask, the cells were transfected with a total of 500 μg pAAV2 genome, *rep*/*cap* and

helper plasmid at 1:1:1 molecular ratio (Table 2.1). All rAAV genome plasmids and *rep* gene were based on AAV2. Linear Polyethylenimine (PEI, 25 kDa, Polysciences) served as transfection agent. To perform the transfection, the plasmids and 1.125 mL of PEI 1mg/mL were each diluted in 10 mL of 150 mmol NaCl, then combined by adding the PEI solution dropwise to the plasmid solution. The transfection mix was added to the cells in HEK293T medium (Table 2.2) with reduced FBS (2 % instead of 10 %). After 72 h, cells were harvested by shaking of the HYPERflasks and subsequent centrifugation (500 g/10 min). The cell pellet from each HYPERflasks was resuspended in 15 mL lysis buffer (1 M Tris(hydroxymethyl) aminomethane, 'Tris' and 150 mM NaCl) supplemented with half of a cOmplete mini EDTA-free protease inhibitor cocktail tablet (Roche). Cells were lysed by four freeze-thaw cycles (- 80 °C / 37 °C). After the final thaw, benzonase was added (final concentration of 50 U/mL, Merck Millipore), incubated at 37 °C for 45 min and cellular debris removed by centrifugation (3700 g / 20 min).

The rAAV vectors were isolated by iodixanol density centrifugation. Typically, three gradients were required for the purification of lysate from two HYPERflasks. Solutions for the iodixanol gradients were prepared and volumes layered in Optiseal tubes (Beckman Coulter) as shown in Table 2.5. Starting with the 15 % fraction in the tube, the 25 % fraction was layered underneath and repeated for all fractions with the lysate added on top of the gradient, resulting in the layers 60 %, 40 %, 25 %, 15 %, lysate from tube bottom to top. Gradients were centrifuged at 59 000 rpm (ca. 272 000 g) for 90 min using a 70Ti rotor in an Optima XPN-80 Ultracentrifuge (both Beckman Coulter). After centrifugation, the 40 % layer containing the rAAV vector was extracted using a 18G needle by puncturing the tube. Subsequently, buffer exchange and concentration were performed using 100 kDa cut-off Amicon Ultra-15 filter tubes (Merck Millipore) by three wash steps with 1x PBS buffer. The rAAV vectors were concentrated to ca. 200 μL, aliquoted and stored at - 80 °C.

Vectors were titrated to determine the number of DNase-resistant genome copies (GC) per mL. Purified rAAV was prepared for titration by DNase I digest for 20 min at 37 °C to remove any residual DNA followed by boiling for 10 min, to inactivate the DNase and release viral genomes. The released viral genomes were diluted 1:10, 1:100 and 1:1000 and analysed in triplicate by adding 4 μL of viral DNA to 6 μL of Taqman Fast Universal PCR Master Mix. To determine GC/mL, samples were titrated against a WPRE standard curve (gift from Dr Tiong Kit Tan) ranging from 1E2-1E8 DNA copies. The qPCR was performed at 95 °C for 120 seconds followed by 40 cycles of 95 °C for 1 second and 60 °C for 20 seconds using the following primers and probes: forward: TGGCGTGGTGTGCACTGT; reverse: CCCGGAAAGGAGCTGACA (both primers 300 nm final); probe: FAM-TTGCTGACGCAACCCCCACTGG-TAMRA (100 nm final).

The rAAV vectors were examined by SDS (sodium dodecyl sulphate)-PAGE (polyacrylamide gel electrophoresis) for integrity and purity. Of each vector, 1E10 GC were incubated with Laemmli sample buffer at 95 °C for 10 min under reducing conditions, loaded on precast SDS-gels (Bolt™ 12 %, BisTris, Life Technologies) and electrophoresed at 200 V (25 V/cm) until the loading dye front was within ca. 1cm of the end of the gel. Coomassie stain was used to visualise AAV capsid proteins VP1, VP2 and VP3 and any impurities.

Table 2.5 Iodixanol gradient preparation for one virus purification or 3 gradients

Fraction	Iodixanol	5 M NaCl	PBS-MK buffer	H_2O	Phenol Red	Volume added to each gradient
15 %	6 mL	4.8 mL	4.8 mL	8.4 mL	-	7.2 mL
25 %	7.5 mL	-	3.6 mL	6.9 mL	30 μL	4.8 mL
40 %	10.05 mL	-	3 mL	1.95 mL	-	4 mL
60 %	15 mL	-	-	-	30 μL	4 mL

PBS-MK buffer: 1x PBS with 5 mM $MgCl_2$ and 12.5 mM KCl; Phenol red: 0.5 % in 1x PBS, Sigma-Aldrich; Iodixanol: OptiPrep density gradient medium, Sigma-Aldrich.

2.5.2 Recombinant lentiviral vector production

Recombinant lentiviral vectors were produced and titrated by the Virus Production Team of the Gene Medicine Research Group (NDCLS, University of Oxford). Briefly, the rLV vectors were produced by transient transfection of LV-Max™ cells (Thermo Fisher Scientific) cultured in their manufacturer supplied media in suspension culture. Cells were transfected with LV-MAX transfection reagent at the ratio 20:9:6:6:6 with genome plasmid : SIV Gag-Pol plasmid : SIV Rev protein plasmid : Sendai virus F envelope protein plasmid : Sendai virus HN envelope protein plasmid (details Table 2.1). The cell supernatant was harvested 48 hours post-transfection and the F protein was activated by enzymatic digest. Vectors were purified from filtered cell culture supernatant by column anion exchange chromatography and tangential flow filtration (TFF) for diafiltration and concentration. Viral vectors were titrated by determining the transducing units (TU) in LV-MAX™ cells by qPCR for the genome-integrated WPRE target after transduction against a standard curve with the same primers and probe as in Section 2.5.1.

Table 2.6 Viral vectors used in this study

Vector name	Serotype/ Pseudotype	Transgene cassette	Source
rAAV1.CMV.EGFP	1	pAAV2.CMV.EGFP (pGM546)	This study/in collaboration with Omar Habib
rAAV2.CMV.EGFP	2	pAAV2.CMV.EGFP (pGM546)	This study/in collaboration with Omar Habib
rAAV5.CMV.EGFP	5	pAAV2.CMV.EGFP (pGM546)	This study/in collaboration with Omar Habib
rAAV6.CMV.EGFP	6	pAAV2.CMV.EGFP (pGM546)	This study/in collaboration with Omar Habib
rAAV6.2.CMV.EGFP	6.2	pAAV2.CMV.EGFP (pGM546)	This study/in collaboration with Omar Habib

rAAV6.2.hCEFI.EGFP	6.2	pAAV2.hCEFI.EGFP (pGM505)	This study
rAAV6.2FF.CMV.EGFP	6.2FF	pAAV2.CMV.EGFP (pGM546)	This study/in collaboration with Omar Habib
rAAV8.CMV.EGFP	8	pAAV2.CMV.EGFP (pGM546)	This study/in collaboration with Omar Habib
rAAV9.CMV.EGFP	9	pAAV2.CMV.EGFP (pGM546)	This study/in collaboration with Omar Habib
rAAV5.CMV.SP-B	5	pAAV2.CMV.SP-B (pGM705)	This study
rAAV5.hCEFI.SP-B	5	pAAV2.hCEFI.SP-B (pGM707)	This study
rAAV5.SP-B MG	5	pAAV2.SP-B MG (pGM709)	This study
rAAV6.2.CMV.SP-B	6.2	pAAV2.CMV.SP-B (pGM705)	This study
rSIV.F/HN.CMV.EGFP (vGM010)	F/HN	pSIV.CMV.EGFP (pGM311)	Virus Production Team, wave 111
rSIV.F/HN.hCEF.EGFP (vGM107)	F/HN	pSIV.hCEF-EGFP (pGM357)	Virus Production Team, wave 109

2.6 Animal studies

2.6.1 Maintenance

All animal experiments were carried out under United Kingdom Home Office approved project licence (PPL 30/3259) in accordance with the Animals (Scientific Procedures) Act 1986. Wildtype, pathogen-free BALB/c mice (Envigo) were used for all animal experiments. Mice were housed by the University of Oxford Biomedical Services and fed on standard chow and water *ad libitum.* Mice were aged 8-9 weeks at the start of

experiment and allowed to acclimatise for 1 week before start of experiments. At the experimental end point, animals were sacrificed by cervical dislocation.

2.6.2 Intranasal dosing

Animals were dosed with 1E11 GC of rAAV vector diluted in 100 μL. For dosing, mice were induced with 5 %/5 % isoflurane/O_2 (IsoFlo, Zoetis) in a prefilled induction chamber. Upon loss of righting reflex, the isoflurane supply was adjusted to 2.5 %. Mice were removed individually from the chamber for dosing. Depth of anaesthesia was monitored by muscle relaxation, foot pad pinch reflex and breathing rate. While holding the mouse upright, the diluted rAAV vector was pipetted onto the nose while holding the mouth shut to allow for vector inhalation. Naive mice remained untreated and mock treated mice were dosed intranasally with vector diluent (PBS buffer). No more than four mice were induced at once. Recovery was monitored until normal active behaviour was observed. Recovery started usually within 2 min. Mice were checked regularly in the days after dosing.

2.6.3 Lung harvest

After cervical dislocation, the reflex to the foot pad pinch was tested. The jugular vein was cut to disrupt circulation and confirm death and the skin was removed to expose rib cage and trachea. A small incision was made underneath the larynx, a canula attached to a 21G needle and a 2 mL syringe was inserted to inflate the lung with 4 % PFA to fix the tissue. The trachea was secured to the canula by a suture and removed from the carcass. Lungs were stored in a tube filled with 4 % PFA overnight and not more than 24 h. Next, the lungs were inflated with 30 % sucrose in 1x PBS for dehydration and incubated in that solution as in the previous step. On the third day, lungs were inflated with a 1:2 solution of dehydration solution and OCT (optimal cutting temperature) embedding medium (Thermo Fisher Scientific). Lungs were separated into the left lobe, superior lobe, middle lobe,

inferior lobe and post-caval lobe by careful dissection. Each lobe was embedded in OCT medium in cryomoulds and stored at - 80 °C until sectioning.

2.6.4 Bronchoalveolar lavage

If bronchoalveolar lavage fluid (BALF) was collected, it was harvested from the same animals as for lung harvest. Mice were euthanised, death confirmed and an intratracheal tube installed as in Section 2.6.3. Lungs were flushed with 1 mL of 1x PBS supplemented with 0.1 % BSA and 50 mM EDTA (BALF collection solution) by inflation and collection of the solution using a syringe attached to the intratracheal tube. This process was repeated three times to receive more concentrated BALF samples and was performed before fixation. The BALF samples were centrifuged to remove cellular debris or blood impurities (16 000 g / 20 min / 4 °C). Samples were aliquoted and stored at - 80 °C until use.

2.7 Precision cut lung slices

2.7.1 Generation and culture of precision cut lung slices

Human PCLS were generated from fresh human lung tissue donated after surgical resection. Only tissue assessed as non-cancerous by a pathologist and kept at 4 °C for less than 8 h was used in experiments. The tissue was received under the ethics of the Oxford Radcliffe Biobank (ORB, research ethics committee (REC) reference: 19/SC/0173, ORB reference: OCHRE 18/A104). Low-melting point agarose (Sigma) was brought to the boil, equilibrated to 42 °C and supplemented with 100 U/mL pen/strep. The lung tissue was inflated with a blunt needle and kept on ice until the agarose solidified. The tissue was cut in long cylinders using a scalpel and subsequently cut traversal in 8 mm diameter sections with 500 μm thickness using a microtome (VF-300-0Z, Precisionary Instruments). After cutting, the sections were washed three times during a 24 h period in culture medium on a rotating wheel at 37 °C. Culture medium consisted of DMEM supplemented with 4mM L-glutamine, 20 μg/mL gentamicin, 100 U/mL pen/strep, 1.25 μg/mL amphotericin B.

Finally, intact sections were selected and placed in 24-well plates. For maintenance, PCLS were kept in culture media (1 mL/well). The media was changed every 2-3 days.

A tissue live/dead (LD) stain was performed to monitor the viability of PCLS using the LIVE/DEAD™ Viability/Cytotoxicity Kit, for mammalian cells according to the manufacturer's instructions (Thermo Fisher Scientific) and as used for PCLS elsewhere [331]. Briefly, the PCLS were transferred to DMEM supplemented with 4 μM calcein and 4 μM ethidium homodimer and incubated at 37 °C for 45 min. After incubation, PCLS were washed in DMEM and imaged for the calcein staining (metabolically active cells, 'live') and ethidium homodimer staining (DNA stained if the cell membrane is compromised, 'dead') using an EVOS fluorescence microscope. PCLS treated with 0.3 mM Triton-X-100 (T-X-100) at 37 °C for 1 h served as negative control for live stain and positive control for the dead stain.

2.7.2 Vector dosing to precision cut lung slices

On day 7 after generation, PCLS placed in 500 μL culture media were transduced with 4E10 GC with rAAV vector diluted in in 50 μL 1x PBS. For rLV vectors, maximum feasible dose was applied in TSSM buffer (20 mM Tris, 100 mM NaCl, 10 mg/mL sucrose, 10 mg/mL mannitol). On day 1 after transduction, the media was increased to 1 mL/well, and then the usual schedule for culture maintenance resumed.

Onset of EGFP expression was monitored by *en face* imaging using an EVOS fluorescence microscope. On day 8 after transduction (day 15 after generation), PCLS were frozen to process in fixed-frozen sections. Briefly, PCLS were fixed in 4 % PFA for 15 min, and subsequently dehydrated in 30 % sucrose in PBS buffer at 4 °C overnight. PCLS were embedded in OCT medium and kept at - 80 °C until sectioning. PCLS were sectioned in 7 μm sections and kept at 4 °C with desiccants (MiniPax® absorbent packets, Sigma) until analysis. For analysis of native EGFP expression, slides were mounted with

an autofluorescence reducing mountant (ProLong Diamond antifade, Thermo Fisher Scientific) and imaged using an EVOS fluorescence microscope.

2.8 Lung bud organoids

2.8.1 Generation and culture of lung bud organoids

Lung bud organoids were generated from the human ESC line AND-2 which was obtained from the “Biobanco de células madre de Granada” (ISCIII, Spain) at passages 29-35. LBOs were generated by sequential differentiation as visualized in Figure 4.7 and described in detail in [332] (first published in [333]). The protocol for generation for LBOs used in chapter 5 was essentially as described Meyer-Berg, *et al.* 2020 [332] except that the LBOs in chapter 5 were generated from the human ESC line RUES2 or from gene edited cell lines generated, while LBOs generated in chapter 4 were grown from the human ESC line AND-2. Further, the first step in the differentiation protocol was omitted, because the RUES2 cell line was cultured in media not requiring feeder cells (see Section 2.4.1). Additionally, Accutase (Stem Cell Technologies) for ESC dissociation and TrypLE Select Enzyme (Life Technologies) in all later cell dissociation steps, replaced Trypsin. The optional growth factor EGF during maturation was omitted in these organoids.

2.8.2 Microinjection

LBO microinjections of rAAV vectors were performed as described in detail previously [332], except that for LBOs injected in chapter 5 pre-pulled injection needles with 5 μm tip were used (TIP5TW1-L, World Precision Instruments). Further, the injection equipment consisted of a FemtoJet microinjector (Eppendorf), a manual micromanipulator (World Precision Instruments) and microscope (Brunel microscopes) and ca. 10 % phenol red (Sigma-Aldrich) was added to the rAAV vector solutions to visualise successful injection.

2.8.3 Cas9 microinjection for gene editing in lung bud organoids

Cas9 protein (*Streptococcus pyogenes* origin) and sgRNA were rehydrated and complexed according to the manufacturer's instructions (both Synthego). For transfection of four LBOs, a total of 300 pmol sgRNA and 40 pmol Cas9 were complexed and then 300 pmol donor DNA (Table 2.7) and 1 μL of the Cas9 reagent (Lipofectamine™ CRISPRMAX™ transfection kit, Thermo Fisher Scientific) were added. Separately, 1.5 μL CRISPRMAX™ transfection reagent was mixed with 4.5 μL OpitMEM and 1 μL phenol red solution (0.5 % in 1x PBS, Sigma-Aldrich). The RNP complexes were immediately added to the transfection reagent mix, incubated for 5 min, and microinjected in LBOs within 30 min. LBOs were returned to the incubator with fresh media supplemented with 10 μM Y-27632. Genomic DNA was extracted using QuickExtract™ DNA Extraction Solution (Lucigen) on day 3 post transfection and region of interest analysed using Sanger sequencing and the Synthego ICE v2 analysis tool (see Section 2.4.3.2).

Table 2.7 Sequences of CRISPR/Cas9 gene editing tools to correct 121ins2 mutation in *SFTPB* exon 4

Name	Sequence	PAM
sgRNA exon4_121ins2_ correction	TTGACGACTACTTCGAACCC	TGG
donor	G*G*GTGGGCACAGGGGCCTGCGTGGGGAGGCAGGC AGGAGGTGAGCTTGCAGCCCTCACAATCTGGTTCTGG AAGTAGTCGATGACCAGGGGGAAGTAGTCGTCAAGCA CTTGGTTGCACTGGGGC*A*T	-

2.9 Preparation of fixed-frozen sections

Fixed-frozen sections were prepared using the Cryostar NX50 Cryostat (Thermo Fisher Scientific). The OCT medium embedded tissue or culture was removed from storage at - 80 °C and allowed to equilibrate to - 20 °C in the cryostat chamber. The tissue block was

frozen to the specimen chuck using OCT medium and trimmed at 30 μm until the block's region of interest was reached, after which, 7 μm sections were cut and mounted on SuperFrost Plus slides (VWR). After drying for 10 min, slides were stored at 4 °C with desiccants (MiniPax® absorbent packets, Sigma) until analysed.

2.10 Immuno- and lectin histochemistry

Fixed-frozen sections were stained for various targets as listed in Table 2.8 and Table 2.9. All incubations were performed in a humidified staining chamber in the dark. A wash step consisted of flicking off excess buffer and submerging slides in respective wash buffer.

Sections from LBOs and PCLS underwent target antigen retrieval, with the exception of SP-B staining. For heat target antigen retrieval, sections were first washed once with wash buffer (1x PBS, 0.1 % T-X-100) and subsequently incubated in citrate buffer (10 mM sodium citrate, 0.05 % Tween-20, pH 6.0) in a Coplin jar at 96 °C for 15 min. Subsequently, the slides were washed once in wash buffer and the tissue marked on the slide with hydrophobic ink pen. When heat antigen retrieval was omitted, tissue was circled with a hydrophobic ink pen followed by one wash step before continuing to blocking. Staining for heparan sulphate included a permeabilisation step (1x PBS / 1 % T-X-100 / 15 min).

For immunohistochemistry, sections were first blocked for 2 h in a humidified staining chamber with the buffer indicated in Table 2.9. After one brief wash step, the primary antibody was applied at 4 °C overnight. Subsequently, the slides were washed in wash buffer four-times for 5 min. All secondary antibodies were applied in 1:500 in wash buffer with 1 % BSA for 1 h. After washing the slides for four times in wash buffer for 5 min, sections were mounted with ProLong™ Gold antifade mountant with DAPI (Thermo Fisher Scientific).

For lectin histochemistry, heat antigen retrieval was first performed as described above. The wash buffer for lectin stains was 1x TBS. For biotinylated lectins a streptavidin/biotin blocking kit was used (SP-2002, Vector Labs) and each blocking solution was applied to the sections for 15 min. Sections were blocked with blocking buffer (1x TBS with 1 % BSA) for 20 min. Lectins were applied at 10 μg/mL in blocking buffer at 4 °C (see Table 2.8). Sections were washed twice for 10 min and where applicable the streptavidin-coupled fluorophore (Thermo Fisher Scientific) was applied for 2 h in blocking buffer. After the final wash step, slides were mounted with ProLong™ Gold antifade mountant with DAPI (Thermo Fisher Scientific). Sections were stored at 4 °C in the dark and imaged within 2 weeks.

A parallel staining control omitting the primary antibody served as a negative staining control in immunohistochemistry. For lectin stains, the glycan receptors heparan sulphate and sialic acid were digested by heparinase III (AMS.HEP-ENZ III-S, amsbio) and sialidase A (Agilent Technologies), respectively, just before lectin or primary antibody incubation. Glycan digests were performed at 37 °C (HybEZ II oven, Advanced Cell Diagnostics). Heparinase III was applied at 4 mUI/mL in reaction buffer (100 mM sodium acetate, 3 mM $CaCl_2$ and 0.2 % BSA) for 3 h. Sialidase A was incubated at 5 UI/mL in reaction buffer B (Agilent Technologies) for 2 h.

Table 2.8 Lectin histochemistry

Target	Lectin	conjugate	Catalogue number, Supplier
α-2,3-linked-sialic acid	*Maackia amurensis* lectin I (MAA I)	Fluorophore (fluorescein)	FL-1311, Vector Laboratories
α-2,3-linked-sialic acid	*Maackia amurensis* lectin II (MAA II)	Biotin	B-1265, Vector Laboratories
α-2,6-linked-sialic acid	*Sambucus nigra* lectin (SNA)	Fluorophore (fluorescein)	FL-1301, Vector Laboratories

Table 2.9 Antibodies for immunohistochemistry

Target	Blocking	Primary antibody	Primary antibody Dilution	Secondary antibody
SP-C (LBOs and PCLS)	1x PBS, 1 % BSA, 5 % goat serum, 0.1 % T-X-100	α-prosurfactant protein C (ab3786, Merck)	1:200	Goat anti-rabbit, Alexa Fluor 555 (A21428, Life Technologies)
SP-C (murine lung)	1x PBS, 1 % BSA, 5 % goat serum, 0.1 % T-X-100	α-prosurfactant protein C (ab3786, Merck)	1:2500	Goat anti-rabbit, Alexa Fluor 594 (A11012, Life Technologies)
SP-B	1x PBS, 5 % BSA, 0.1 %T-X-100	α-surfactant protein B (sc-133143, Santa Cruz Biotech)	1:200	Donkey anti-mouse, Alexa Fluor 555 (A31570, Life Technologies)
Heparan sulphate	1x PBS, 1 % BSA, 0.1 % T-X-100	α-heparan sulphate, clone F58-10E4 (370255, Amsbio)	1:100	Goat anti-Mouse IgM, (A21426, Life Technologies)
AAVR	1x PBS, 1 % BSA, 5 % goat serum, 0.1 % T-X-100	α-KIAA0319L (PA5-67257, Invitrogen)	1:75	Goat anti-rabbit, Alexa Fluor 555 (A21428, Life Technologies)
ACE2	1x PBS, 2 % BSA, 0.1 % T-X-100	α-ACE2 (AF933, R&D Systems)	1:500	Donkey anti-goat (A21432, Life Technologies)
TMPRSS2	1x PBS, 2 % BSA, 0.1 % T-X-100	α-TMPRSS2 (ab92323, abcam)	1:500	Goat anti-rabbit, Alexa Fluor 555 (A21428, Life Technologies)

2.11 Immunocytochemistry

Protein expression in cells was investigated using immunocytochemistry (ICC). In all wash steps 1x PBS was used. Adherent cells were seeded on cover slips which were coated according to the cell line's culture conditions. After a minimum of 24 h culture, cells were briefly washed and then fixed in 4 % PFA in 1x PBS for 15 min. Cells were washed once and stored in 1x PBS for up to 2 weeks if not proceeding immediately to permeabilsation (1 % saponin in 1x PBS) for 15 min. Cells were blocked in 3 % BSA in 1x PBS for 30 min followed by primary antibody incubation. Primary antibodies were diluted as indicated in Table 2.10 in blocking buffer and incubated at 4 °C for 3 h. After washing three times, the secondary antibody was applied in blocking buffer for 1 h (Table 2.10). Finally, cells were washed for three times and coverslips were mounted using ProLong™ Gold antifade mountant with DAPI (Thermo Fisher Scientific). Slides were stored at 4 °C in the dark and imaged within 1 week.

Table 2.10 Antibodies for immunocytochemistry

Target	Antibody	Species	Catalogue number	Supplier	Dilution
OCT4	anti-OCT4	rabbit	A24867	Invitrogen	1:200
SSEA4	anti-SSEA4 IgG3	mouse	A24866	Invitrogen	1:100
TRA-1-60	anti-TRA-1-60 IgM	mouse	A24868	Invitrogen	1:100
SOX2	anti-SOX2	rat	A24759	Invitrogen	1:100
FOXA2	Human HNF-3 beta/ FOXA2 antibody	goat	AF2400-SP	RD systems	2.5 µg/mL
rat	Alexa Fluor™ 488 donkey anti-rat	donkey	A24876	Invitrogen	1:250
mouse IgG3	Alexa Fluor™ 488 goat anti-mouse IgG3	goat	A24877	Invitrogen	1:250
rabbit	Alexa Fluor™ 555 donkey anti-rabbit	donkey	A24869	Invitrogen	1:250
mouse IgM	Alexa Fluor™ 555 goat anti-mouse IgM	goat	A24871	Invitrogen	1:250
goat	Alexa Fluor™ 555 donkey anti-goat	donkey	A21432	Invitrogen	1:250

Antibodies partially available as set from PSC 4-Marker Immunocytochemistry Kit together with ICC reagents (A24881, Invitrogen)

2.12 Western blot analysis

Western blot analysis was used to analyse protein expression levels in BALF of mice (Section 2.6.4) and lysates of immortalised cell lines and LBOs. Cell lysates for Western blot analysis were obtained by lysing cells in 1x RIPA buffer (Merck Millipore) supplemented with cOmplete mini EDTA-free protease inhibitor cocktail tablet (Roche) on ice for 20 min. Lysates were homogenised using QIAshredder spin columns (Qiagen) and cellular debris was removed by centrifugation (18 000 g / 5 min / 4 °C). Samples were aliquoted and stored at - 80 °C or used immediately.

A BCA assay (DC protein assay, Bio-Rad) was used to determine total protein concentration in both BALF samples and cell lysates by titration against a seven step BSA standard curve run in triplicate. Samples were diluted as appropriate (typically up to 1:4)

and run in duplicate according to the manufacture's microplate protocol. Absorbance was measured at 750 nm using a SpectraMAX i3x plate reader and SoftMax Pro software v7.0 (both Molecular Devices).

Samples normalised to total protein content were incubated at 70 °C for 10 min with Laemmli sample buffer under non-reducing conditions (Bio-Rad). Samples were electrophoresed on precast SDS gels (Bolt™ 12 %, BisTris, Life Technologies) at 200 V (25 V/cm) in 1x NuPAGE™ MES SDS Running Buffer (Life Technologies). Protein sizes were estimated relative to Kaleidoscope protein marker (Bio Rad). After completion of SDS-PAGE, the gel was blotted with a 0.2 μm nitrocellulose membrane (Bio-Rad) between filter papers and sponges. Wet protein transfer was performed in 1x Bolt Transfer buffer with 10 % methanol (Life tech) at 200 mA for 90 min in a Trans-Blot Cell (Bio-Rad) with frozen cooling packs inserted into the transfer buffer to reduce buffer heating. The transfer efficiency and accurate normalisation was examined by Ponceau S staining of the nitrocellulose membrane. Subsequently, the membrane was destained with 0.1 M NaOH and washed once with ddH_2O and wash buffer (PBS-T: 1x PBS with 0.1 % Tween 20).

For immunoblotting, all antibodies were applied in wash buffer with 5 % skimmed milk powder (Blocking grade blocker, Bio-Rad) and on a rocking platform. Primary antibodies were diluted as indicated in Table 2.11 and incubated with the membranes at 4 °C overnight. After four washes for 5 min, the secondary antibody was applied for 1 h, diluted as in Table 2.11. The membrane was washed again (4x 5 min) and briefly rinsed with ddH_2O. The blot was developed using the Clarity ECL substrate (Bio-Rad) and signal recorded using an iBright FL1000 system (Thermo Fisher Scientific).

In some instances, typically to confirm equivalence of samples, sequential antibody staining of the same membrane was performed. To facilitate this sequential staining, existing antibody complexes were removed by incubated twice in stripping buffer (200 mM

glycine, 0.1 % SDS, 1 % Tween 20, pH 2.2) for 10 min. After a wash step, membranes were reprobed with the sequential primary antibody by starting with the blocking step.

For Western blot analysis, band intensities were quantified using the iBright analysis software (Thermo Fisher Scientific). Bands of interest were normalised against an endogenous protein as loading control.

Table 2.11 Antibodies for immunoblotting

Target	Antibody	Species	Catalogue number	Supplier	Dilution
SP-B	Anti-Pro + Mature Surfactant Protein B antibody	rabbit	ab40876	Abcam	1:5000 in Chapter 3, 1:3000 in Chapter 5
GAPDH	Anti-GAPDH antibody	rabbit	G9545	Sigma	1:30 000
SP-A	Anti-SP-A/PSAP antibody	rabbit	ab115791	Abcam	1:1000
Rabbit IgG	Goat anti-Rabbit IgG (H+L) Secondary Antibody, HRP	goat	A16096	Invitrogen	1:10 000

2.13 Flow cytometry

Cells for flowcytometry analysis were harvested with TrypLE Select Enzyme (Life Technologies) to achieve a single cell solution while also minimising digestion of the target epitope by avoiding long incubations. Cells were briefly washed in 1x PBS, fixed for 15 min (2 % PFA), washed twice and stored in 1x PBS for up to 2 weeks. Cells were blocked for 30 min (2 % BSA in 1x PBS) and subsequently incubated with fluorophore-coupled antibodies (Table 2.12). Antibodies were incubated in 1x PBS with 0.5 % BSA for 1 h. After three washes, flow cytometry was recorded using an LSRII flow cytometer and data analysed using FlowJo X software (both BD Biosciences).

Table 2.12 Antibodies for flow cytometry

Target	Antibody	Catalogue number	Supplier	Dilution
CXCR4	CD184 (CXCR4) Monoclonal Antibody (12G5), APC	17-9999-42	Invitrogen	1:100
c-Kit	CD117 (c-Kit) Monoclonal Antibody (104D2), FITC	11-1178-42	Invitrogen	1:100

2.14 RNAscope

2.14.1 RNAscope pretreatment and assay

RNAscope® was used as an *in situ* hybridisation (ISH) method to visualise and quantify mRNA expression. RNAscope® H2O2 and Protease Reagents (Advanced Cell Diagnostics) and RNAscope Multiplex Fluorescent Reagent kit v2 (Advanced Cell Diagnostics) were used to perform the assay. The manufacturer's instructions of both kits were followed according to support document 323100-USM, 2017 (available at https://acdbio.com) with additional pretreatment steps specified below. All incubations were performed in an HybEZ II oven (Advanced Cell Diagnostics) which allowed for controlled temperature in a (optionally humidified) staining chamber. All wash steps up to probe hybridisation were performed in ddH_2O, all wash steps after probe hybridisation in RNAscope® wash buffer (Advanced Cell Diagnostics). Wash steps included submerging the slides for 2 min.

On the day before the assay, fixed-frozen murine lung sections (prepared as in 2.9) were pretreated to prevent tissue detachment. This included baking at 37 °C for 60-90 min, post-fixation step (30 min/cold 4 % PFA in 1x PBS), followed by dehydration (70 % and 100 % IMS, 1 min each) and finally baking at 60 °C for 30 min. Slides were kept with desiccants overnight.

Slides were treated with RNAscope® Hydrogen Peroxide treatment for 10 min and washed twice before target antigen retrieval. Slides were incubated for exactly 5 min at 98-102 °C in a beaker filled with RNAscope® Target Retrieval Reagent (Advanced Cell Diagnostics). Slides were dehydrated and hydrophobic barrier drawn. Next, slides were incubated with RNAscope® Protease III solution at 40 °C for 30 min.

Sections were stained using RNAscope probes (Table 2.13). To hybridise probes, sections were incubated with probes at 40 °C for 2 h. Slides were washed twice before proceeding to signal amplification. AMP1 and AMP2 were each applied for 30 min at 40 °C, followed AMP3 for 15 min at 40 °C. After each amplification step slides were washed twice.

Opal dyes 520 and 570 (1:1500 in TSA buffer, Perkin Elmer) were used to visualise the *Sftpb* and WPRE signal, respectively. Signal for each target ('channel') was developed sequentially. For signal development, slides first incubated at 40 °C with HRP-C1 reagent for 15 min, followed by incubation with Opal dye for 30 min and incubation with HRP blocker for 15 min. After each step slides were washed twice. Finally, sections were mounted using ProLong™ Gold antifade mountant with DAPI (Thermo Fisher Scientific). Slides were stored at 4 °C in the dark and imaged within one week.

2.14.2 Nuclease pretreatment to uncouple RNA and DNA signal

As an optional pretreatment step, slides were treated with RNase or DNase to uncouple RNA and DNA signal, respectively. This was performed as the final step before application of RNAscope probes for the RNAscope assay (Section 2.14.1). For RNase pretreatment, RNase A (7000 UI/mL, Qiagen) was diluted in TBS-T (1x TBS with 0.05 % Tween). Each slide was digested with 3.5 UI in 50 μL/slide at 37 °C for 45 min. For DNase pretreatment, RNase-free DNase 1 (New England Bio Labs) was applied in its supplied buffer at 2 UI in 50 μL/slide at 37 °C for 45 min. In double digests, first DNase and then RNase was applied. After each nuclease digestion step slides were washed in ddH_2O (4x 5 min).

Undigested slides were incubated with the respective buffer without inclusion of the nuclease. In all continuing steps, slides from each nuclease digestion condition were kept separately to avoid cross-contamination. DNase pretreatment as described above was applied in all RNAscope assays performed for subsequent quantification of mRNA expression to prevent cross-reaction of the WPRE probe with WPRE vector DNA. When DNase pretreatment was applied, RNase as well as RNase/DNase double digests were run as controls in parallel.

Table 2.13 RNAscope probes

Probe name	mRNA target, channel	Catalogue name	Catalogue number
WPRE probe	WPRE (C1)	WPRE-O1	450261
Sftpb probe	*Sftpb* (C2), not cross-reactive with human *SFTPB* mRNA	Sftpb-noX-human	539421-C2
Positive control probe	Murine housekeeping genes *Polr2a* (C1), *Ppib* (C2), *Ubc* (C3)	3-plex Positive Control Probe-Mm	320881
Negative control probe	*DapB* of *Bacillus subtilis* (C3)	3-plex Negative Control Probe	320871

All probes were purchased from Advanced Cell Diagnostics. *Polr2a*, RNA polymerase (II) polypeptide A; *Ppib*, Peptidyl-propyl cis-trans isomerase B; *Ubc*, Ubiquitin C.

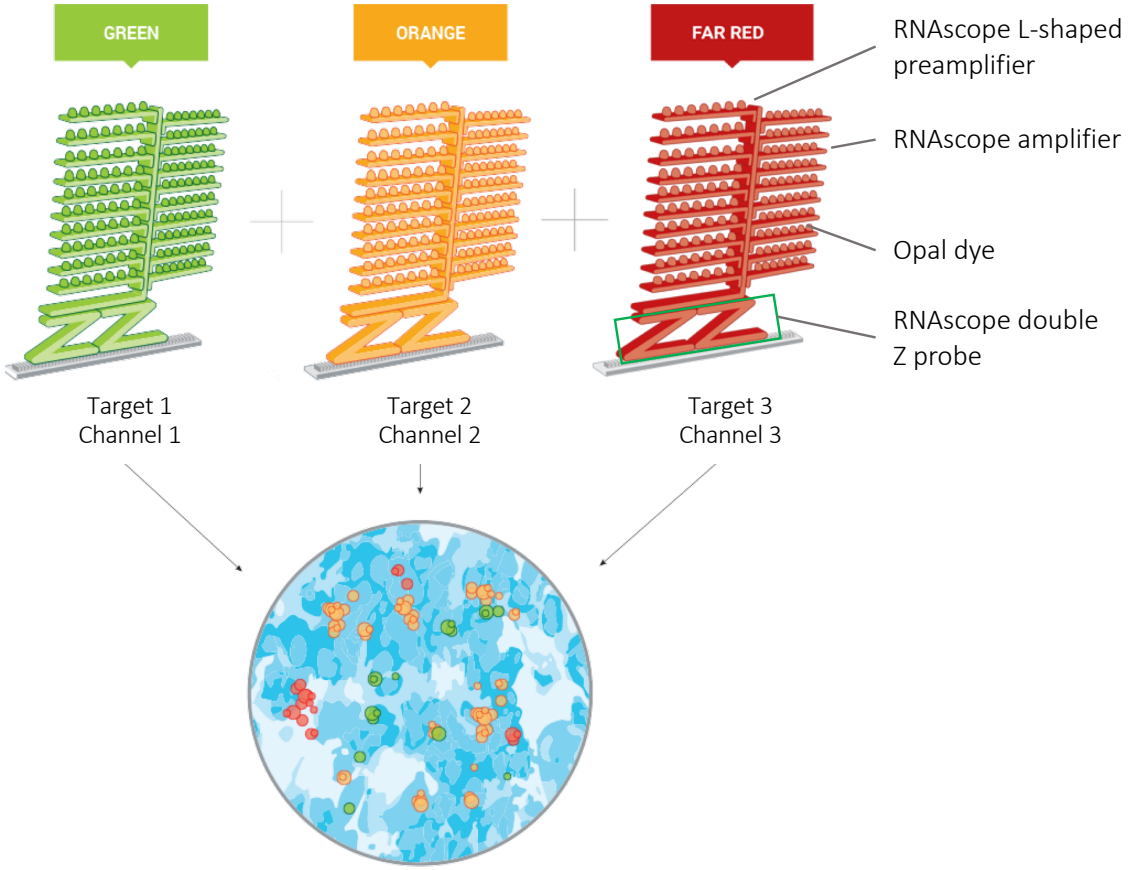


Figure 2.1 RNAscope detection of nucleic acid targets

The RNAscope double Z probes bind to their target nucleic acids. For each target several double Z probes are designed (usually 20 double Z probes, minimum 3). The nucleic acid preamplifier and 20 amplifiers hybridise to the double Z probe in their respective channel. Each amplifier can bind 20 channel-specific label probes, leading to a 20x20x20 signal amplification ensuring that every mRNA molecule becomes visible as one dot. HRP label probes catalyse Opal fluorophore labelling by tyramide signal amplification [334]. Importantly, size or intensity of the dot is only a measure amplification efficiency and not quantitative measure for number of nucleic acids.

Figure modified from https://acdbio.com/rnascope®-fluorescent-multiplex-assay.

2.15 Microscopy and image analysis

Samples were imaged using a EVOS FL Auto 2 ('EVOS', Thermo Fisher Scientific) widefield fluorescence microscope, Leica DM4B (Leica Microsystems) fluorescence microscope or, where indicated, Zeiss LSM-780 inverted confocal microscope. To take images for subsequent quantification, images were taken in a randomized fashion. Where certain morphological structures within one section needed to be imaged (e.g. lung parenchyma or airways), imaging locations were selected using the DAPI channel only. When the transduction rate was very low, the section was scanned systematically to record total number of transduced cells found in each section, rather than randomized imaging.

Images were analysed using Fiji ImageJ Software (version 1.0). Additionally, the ImageJ 'adjustable_watershed' plugin (2012-May-08 Michael Schmid) was installed to digitally differentiate between overlapping dots. The ImageJ software allowed for brightness intensity analysis and counting of dots above brightness threshold, both in whole images and select areas. Macros were written to accelerate image analysis and as a record of image processing and multistep analysis during this study. Macros allowed for channel subtraction to correct for autofluorescence in EVOS fluorescence microscope recorded images (see Table 2.14), automated counting of all cells by counting DAPI-stained nuclei (Table 2.15), counting of transduced EGFP-positive cells (Table 2.16) and assisted counting of RNAscope signal dots (Table 2.17).

Table 2.14 ImageJ macro for channel subtraction

```
function Openmyfilesd0d2(path, filename){
    open(path+filename+"d0.TIF");
    open(path+filename+"d2.TIF");
};

function Closemyfilesd0d2(filename){
    selectWindow(filename+"d0.TIF");
    close();
    selectWindow(filename+"d2.TIF");
    close();
};

function Channelseparation (filename){
    imageCalculator("Subtract create", filename+"d0.TIF",filename+"d2.TIF");
};

function Setbrightnessandgrey(min, max){
    selectWindow(filename+"d2.TIF");
    run("32-bit");
    setMinAndMax(min, max);

};

function AdjustedChannelSubtraction(path, filename,min, max){
    Openmyfilesd0d2(path, filename);
    Setbrightnessandgrey(min, max);
    Channelseparation(filename);
    Closemyfilesd0d2(filename);
};

function Main(min, max){
    path = getDirectory("Choose a Directory");
    myfilelist = getFileList(path);
    for (i = 0; i < myfilelist.length; i++)
    {
        if(lengthOf(myfilelist[i])>6)
        {
            if(substring(myfilelist[i],(lengthOf(myfilelist[i])-6),lengthOf(myfilelist[i]))=="d0.TIF")
            {
                if(indexOf(myfilelist[i], "TD") != -1)
                {
                    filename = substring(myfilelist[i],0,lengthOf(myfilelist[i])-6);
                    print("processing: "+filename);
                            AdjustedChannelSubtraction(path, filename,min,max);
                };
            };
        };
    };
};

Main(-10,245);
```

Table 2.15 ImageJ macro to count nuclei

```
// Colour Thresholding-------------
run("Color Threshold...");
min=newArray(3);
max=newArray(3);
filter=newArray(3);
a=getTitle();
run("HSB Stack");
run("Convert Stack to Images");
selectWindow("Hue");
rename("0");
selectWindow("Saturation");
rename("1");
selectWindow("Brightness");
rename("2");
min[0]=0;
max[0]=255;
filter[0]="pass";
min[1]=0;
max[1]=255;
filter[1]="pass";
min[2]=56;
max[2]=255;
filter[2]="pass";
for (i=0;i<3;i++){
  selectWindow(""+i);
  setThreshold(min[i], max[i]);
  run("Convert to Mask");
  if (filter[i]=="stop")  run("Invert");
}
imageCalculator("AND create", "0","1");
imageCalculator("AND create", "Result of 0","2");
for (i=0;i<3;i++){
  selectWindow(""+i);
  close();
}
selectWindow("Result of 0");
close();
selectWindow("Result of Result of 0");
rename(a);
// Adjustable Watershed and counting particles-------------
run("Adjustable Watershed", "tolerance=0.05");
run("Analyze Particles...", "size=8-Infinity pixel summarize");
```

Table 2.16 ImageJ macro to count transduced cells.

```
// Colour Thresholding
run("Color Threshold...");
min=newArray(3);
max=newArray(3);
filter=newArray(3);
a=getTitle();
run("HSB Stack");
run("Convert Stack to Images");
selectWindow("Hue");
rename("0");
selectWindow("Saturation");
rename("1");
selectWindow("Brightness");
rename("2");
min[0]=0;
max[0]=255;
filter[0]="pass";
min[1]=0;
max[1]=255;
filter[1]="pass";
min[2]=41;
max[2]=255;
filter[2]="pass";
for (i=0;i<3;i++){
  selectWindow(""+i);
  setThreshold(min[i], max[i]);
  run("Convert to Mask");
  if (filter[i]=="stop")  run("Invert");
}
imageCalculator("AND create", "0","1");
imageCalculator("AND create", "Result of 0","2");
for (i=0;i<3;i++){
  selectWindow(""+i);
  close();
}
selectWindow("Result of 0");
close();
selectWindow("Result of Result of 0");
rename(a);
// Count particles
run("Analyze Particles...", "size=7-Infinity pixel show=Overlay clear summarize");
run("Invert");
```

Table 2.17 ImageJ macro to count RNAscope signal dots in selected image area

```
run("Duplicate...", " ");
run("Make Inverse");
setBackgroundColor(0, 0, 0);
run("Clear", "slice");
run("Color Threshold...");
min=newArray(3);
max=newArray(3);
filter=newArray(3);
a=getTitle();
run("HSB Stack");
run("Convert Stack to Images");
selectWindow("Hue");
rename("0");
selectWindow("Saturation");
rename("1");
selectWindow("Brightness");
rename("2");
min[0]=0;
max[0]=255;
filter[0]="pass";
min[1]=0;
max[1]=255;
filter[1]="pass";
min[2]=95;
max[2]=255;
filter[2]="pass";
for (i=0;i<3;i++){
  selectWindow(""+i);
  setThreshold(min[i], max[i]);
  run("Convert to Mask");
  if (filter[i]=="stop")  run("Invert");
}
imageCalculator("AND create", "0","1");
imageCalculator("AND create", "Result of 0","2");
for (i=0;i<3;i++){
  selectWindow(""+i);
  close();
}
selectWindow("Result of 0");
close();
selectWindow("Result of Result of 0");
rename(a);
// Count particles in selected area.
run("Adjustable Watershed", "tolerance=0.1");
run("Analyze Particles...", "size=2-Infinity pixel show=Ellipses display summarize");
```

2.16 Statistical analysis

Statistical analysis was performed using GraphPad Prism (version 8). Non-Parametric tests were used in this study when normal distribution necessary for parametric test could not be assured due to data values set between fixed limits (e.g. percentages). Differences between compared groups was deemed significant for p-values of < 0.05. Significance was indicated in figures as appropriate using asterixis (*= $p< 0.05$, ** = $p< 0.01$, *** = $p< 0.001$, **** = $p< 0.0001$).

Chapter 3 – Selection of a candidate rAAV vector in the murine lung *in vivo*

3.1 Introduction

The gene therapy vector rAAV is available with several naturally occurring serotypes and engineered variants that exhibit various cell and tissue tropisms (discussed in Section 1.5.4). As ATII cells of the lung parenchyma are the only cells that contain organelles for production of mature, active surfactant (discussed in Section 1.2.2), these are the target cells for SP-B gene therapy. The studies described in this chapter aimed to identify a suitable rAAV vector for SP-B deficiency using wildtype mice as a model organism, starting with the identification of AAV serotypes with tropism to ATII cells.

The choice of model for viral gene transfer is crucial for the investigation of vector tropism. Here, intranasal delivery to the mouse lung allows for investigation of AAV serotypes *in vivo*, where vector is exposed to the polarised lung epithelium from the luminal surface, as would be anticipated in patients. Further, several of the physiologic barriers to vector delivery to the lung, such as mucus, surfactant and alveolar macrophages are also present in the mouse lung. Lastly, mice are a prevalent and well-studied model for drug development supporting the selection of mice as the first model organism to investigate rAAV gene therapy for SP-B deficiency.

Previously, Limberis and colleagues investigated multiple AAV serotypes in murine airways and alveoli after intratracheal delivery (Section 1.4.1) where the results were reported by attributing scores based on transduction efficacy observed from LacZ reporter gene expression [283]. Serotypes 5, 9, 6 and 6.2 (6.2 > 6 = 5 > 9) were identified as

achieving highest levels of transduction of alveoli [283]. The rAAV6.2 vector (AAV6 + F129L, [244]) achieved the highest score of all serotypes tested [283], however, alveoli consist of both ATI and ATII cells and therefore the transduction efficiency of ATII cells specifically could not be deduced from these studies. Here, AAV serotypes 5, 9 and 6.2, which were reported to be highly efficient in transducing the murine lung parenchyma [283] were investigated for their ability to transduce ATII cells. Serotype 6 was not selected in this chapter because serotype 6.2 consistently outperformed serotype 6 across human airway epithelial culture, murine airway, murine parenchyma and secretory transgene expression as a measure for whole lung transduction [283]. Therefore, rAAV6.2 was seen as a more efficient version of serotype 6.

3.2 Results

3.2.1 Investigation of AAV serotypes suitable for transduction of the murine lung parenchyma *in vivo*

To determine which AAV serotype transduces the greatest number of ATII cells in the murine lung, mice were dosed intranasally (Section 2.6.2) with 1E11 GC of rAAV5, rAAV6.2 or rAAV9 vector expressing EGFP from the CMV promoter (Table 2.6). On day 28 after transduction, mice were culled and the lungs were processed in fixed-frozen sections (Section 2.6.3, Section 2.9). Two whole lung sections per mouse were scanned for native EGFP fluorescence (Section 2.15) to determine overall transduction efficiency and tropism for transduction of airway versus parenchyma (Figure 3.1). Lung tissue is highly autofluorescent, with elastin and collagen being common autofluorophores [335], helpfully, the ratio of signal to autofluorescence was improved by channel subtraction of the unoccupied RFP channel signal from the GFP channel images (Section 2.15). Representative images of transduction of the murine lung parenchyma and airways are shown in (Figure 3.1a). For AAV serotypes 6.2 (n= 4) and 9 (n= 3), transduction in both the parenchyma and the airways was detected, whereas for AAV serotype 5 (n= 3) only transduction of the parenchyma was observed. Whole lung sections were analysed and the number of transduced cells was first calculated as a percentage of total cells (Figure 3.1d). Serotype 5 achieved the highest transduction rate in the lung which was significantly different from naive mice (mean= 3.06 %, p= 0.023). AAV serotypes 9 and 6.2 showed only low transduction rates (mean= 0.32 % and mean= 0.34 %, respectively) that were not significantly different from naive (mean= 0.08 %).

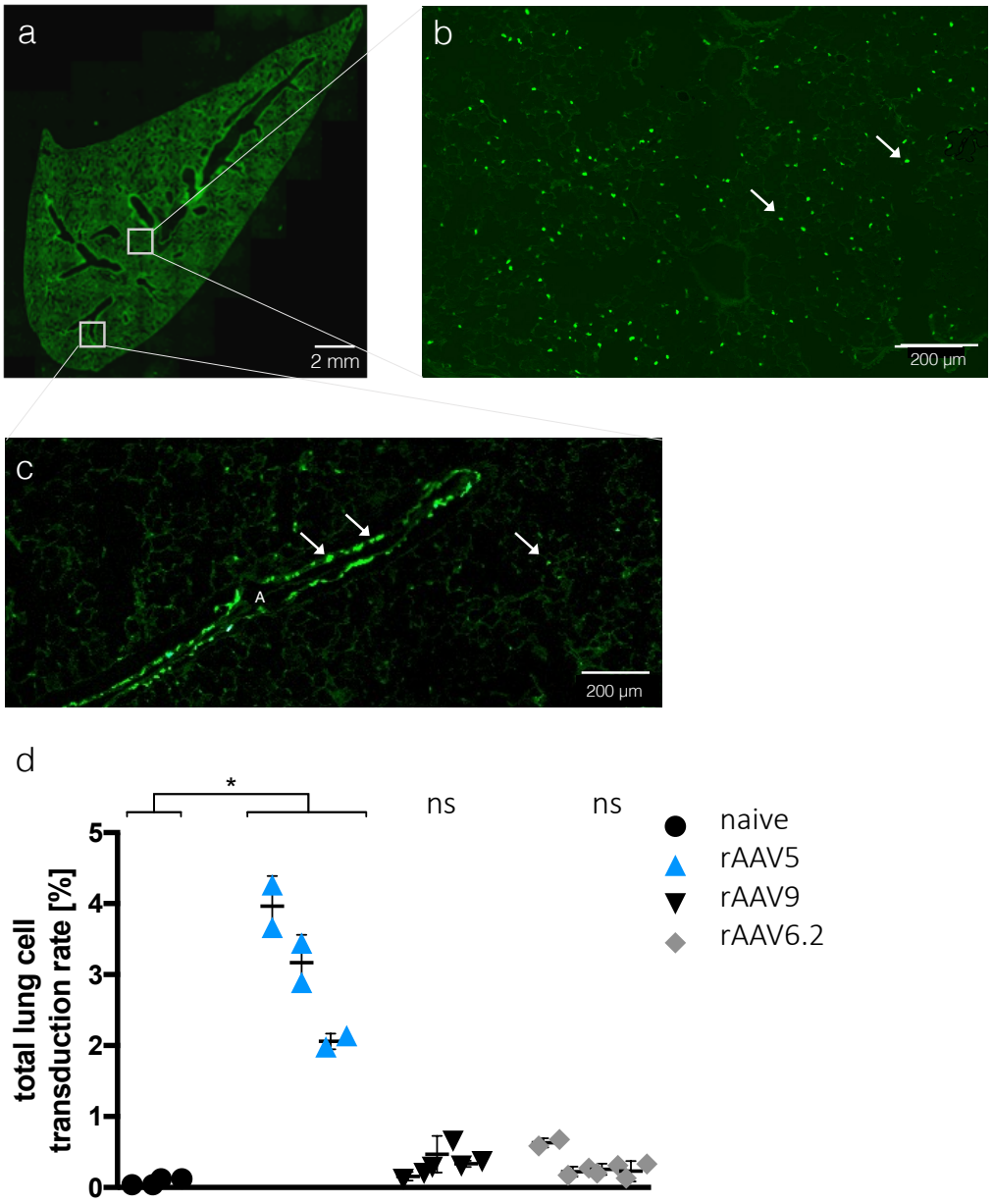


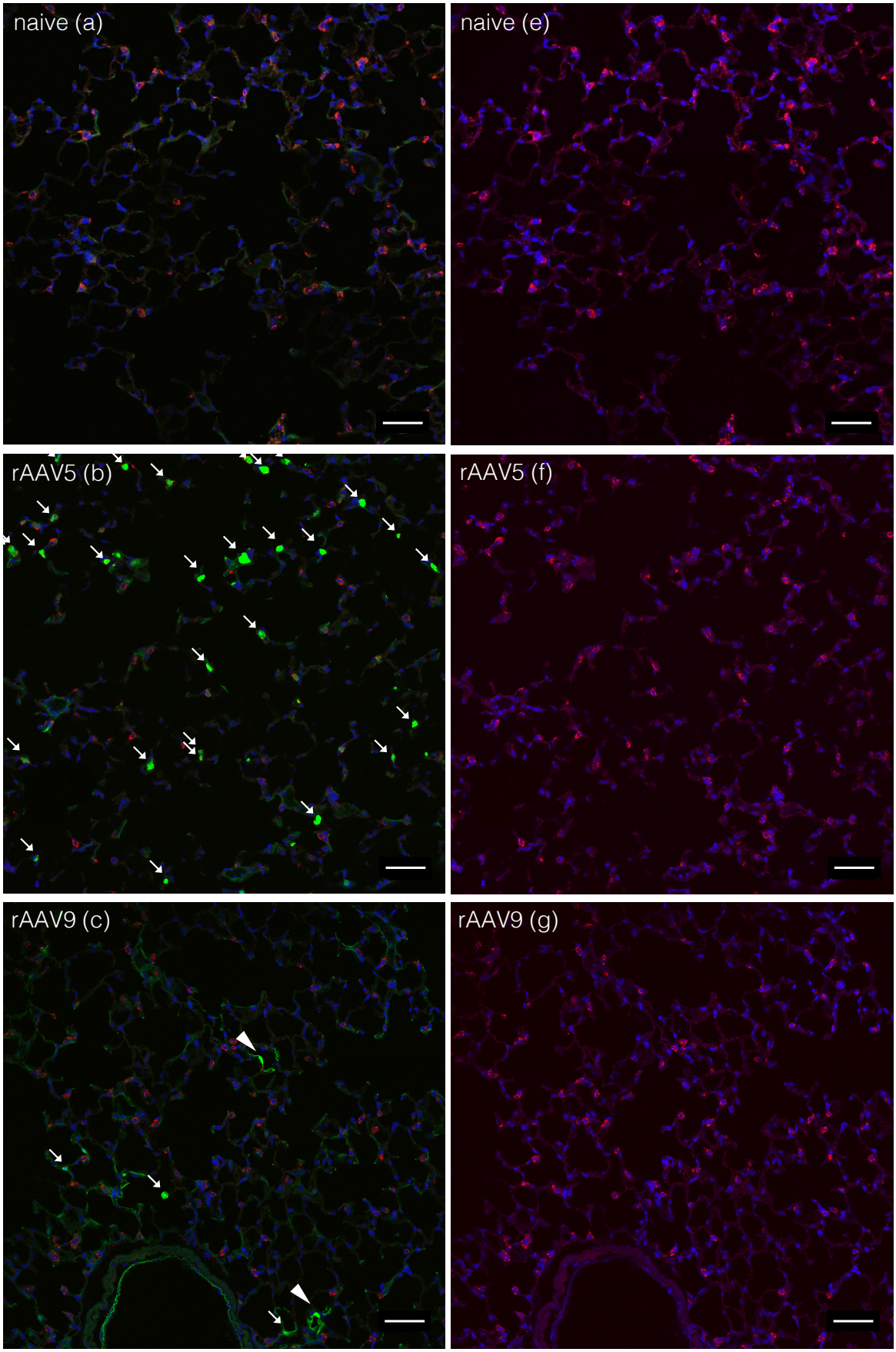
Figure 3.1 Transduction in murine lungs after intranasal delivery of AAV serotypes 5, 9 and 6.2

Mice were intranasally dosed with 1E11 GC of rAAV5.CMV.EGFP (n= 3), rAAV6.2.CMV.EGFP (n= 4), rAAV9.CMV.EGFP (n= 3) or remained naive to treatment (n= 2) and 28 days later the mice were culled the lungs processed to fixed-frozen sections. A typical autofluorescence-corrected whole lung section images of the left lobe of a mouse transduced with rAAV vector and imaged for EGFP native fluorescence using a EVOS FL Auto 2 fluorescence microscope is shown in (a). Image details of the lung parenchyma after transduction with rAAV5 (b). An airway (marked 'A') after transduction

with rAAV9 (c). White arrows indicate examples of transduced cells. Quantification of total lung cell transduction expressing EGFP-positive cells as a percentage of stained nuclei (DAPI) using an ImageJ macro (d, Section 2.15). Each symbol represents one image analysed. Two images were analysed per mouse; each mean (bar) and error symbol (SD) represents the data from the two images associated with each mouse. Kruskal-Wallis test with Dunn's post-hoc test for multiple comparisons analysing means of each animal. (*) $p < 0.05$, (ns) non-significant. Scale bars are 2 mm or 200 μm as indicated.

To determine the specific transduction rate in ATII cells, lung sections were subsequently stained for the ATII cell marker pro-SP-C using immunohistochemistry (Section 2.10). Images were collected in a blinded fashion by first only using the DAPI channel to select image locations (Section 2.15) before subsequently imaging the GFP and red channels for EGFP and pro-SP-C visualisation, respectively. Representative confocal microscopy images are shown in Figure 3.2. For rAAV5, native EGFP fluorescence indicated a relatively high number of ATII cells transduced while transduction of other cell types was not observed. For rAAV9 and rAAV6.2, fewer cells were transduced overall, although in the alveolar region, cells negative for pro-SP-C with morphology similar to ATI cells did also appear to be transduced (Figure 3.2c, arrowheads).

The transduction rate in ATII cells was calculated by choosing 30 cells stained with ATII-cell marker at random in an image and then quantifying the proportion of transduced cells based on colocalisation with EGFP fluorescence (Figure 3.3). In the four randomly taken images per animal, all contained ATII cells transduced by rAAV5, which resulted in the highest transduction rate in ATII cells and was significantly different from naive (mean= 19.4 %, p= 0.014). The transduction rates of rAAV6.2 and rAAV9 were not significantly different from naive and rAAV5 transduced tissues (mean= 4.0 %, p= 0.99 and mean= 4.5 %, p= 0.99, respectively). In summary, rAAV5 achieved the highest overall transduction in the murine lung, as well as the highest transduction rate in ATII cells. Moreover, rAAV5 was most specific for ATII cell transduction with no other transduced cell types observed. Therefore, AAV serotype 5 was identified as candidate serotype for efficient gene transfer to murine ATII cells.


naive (a)
naive (e)
rAAV5 (b)
rAAV5 (f)
rAAV9 (c)
rAAV9 (g)

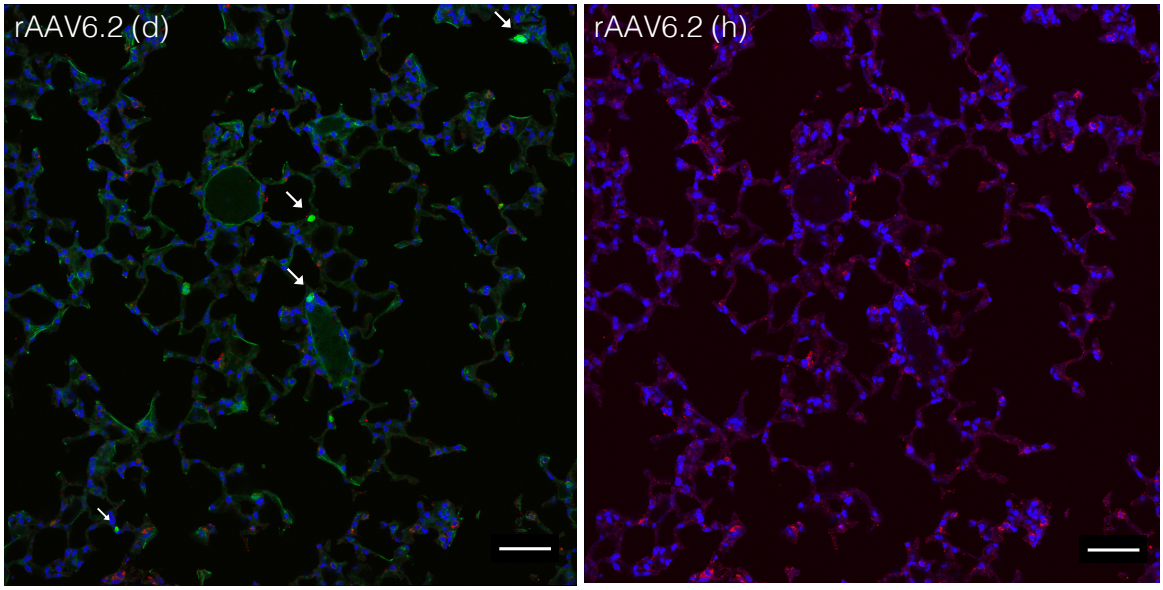


Figure 3.2 Representative images of lung sections from naive mice and mice transduced with AAV serotypes 5, 9, and 6.2

Mice were intranasally dosed with 1E11 GC of rAAV5.CMV.EGFP (n= 4), rAAV6.2.CMV.EGFP (n= 4), rAAV9.CMV.EGFP (n= 4) or remained naive to treatment (n= 2). On day 28 after transduction, the lungs were processed to fixed-frozen sections, immunostained for the ATII cell marker pro-SP-C and imaged using confocal fluorescence microscopy. Merged channel images showing native fluorescence of transduced cells (GFP, green), pro-SP-C immunostaining (Alexa 594, red) and nuclei (DAPI, blue; a-d) as well as pro-SP-C immunostaining and nuclei only (e-h) of left lung lobes from naive mice (a+e) or mice transduced with rAAV5 (b+f), rAAV9 (c+g) or rAAV6.2 (d+h). White arrows indicate colocalisation of native EGFP fluorescence and ATII cells. White arrowheads indicate colocalisation of native EGFP fluorescence with cells morphologically reminiscent of ATI cells. Scale bar 50 µm.

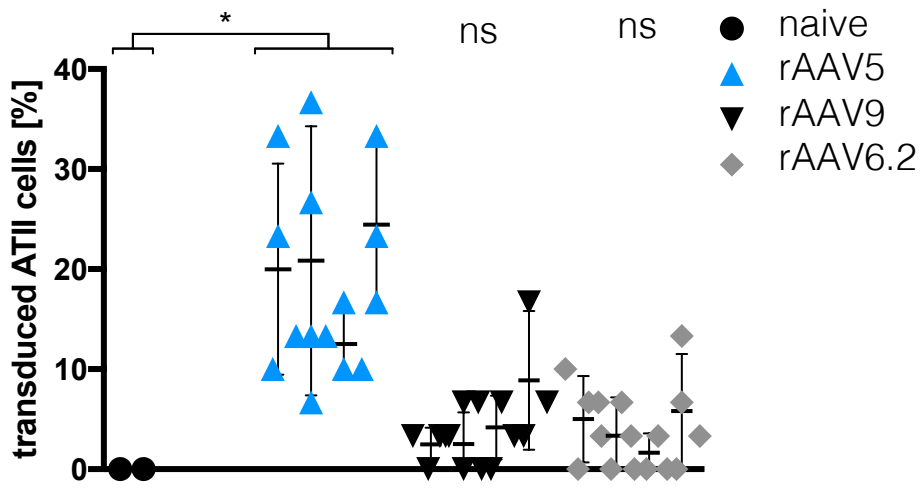


Figure 3.3 Quantification of transduction rate in ATII cells after intranasal dosing of mice with rAAV5, rAAV9 and rAAV6.2

Mice were intranasally dosed with 1E11 GC of rAAV5.CMV.EGFP (n= 4), rAAV6.2.CMV.EGFP (n= 4), rAAV9.CMV.EGFP (n= 4) or remained naive (n= 2). On day 28 after transduction, the lungs were processed to fixed-frozen sections, immunostained for the ATII cell marker pro-SP-C and imaged using confocal fluorescence microscopy. For each animal, a total of 3-4 images were captured and analysed. In each analysed image, 30 ATII cells were chosen randomly and then the proportion of transduced cells determined. Each symbol represents the proportion of transduced cells in one analysed image. Each mean (bar) and error symbol (SD) represents one mouse. Kruskal-Wallis test with Dunn's post-hoc test for multiple comparisons analysing means of each animal. (*) $p < 0.05$, (ns) non-significant.

3.2.2 Design and validation of rAAV expression cassettes for SP-B expression and viral vector production

Several rAAV transgene cassettes were designed and constructed (Section 2.2) for therapeutic SP-B expression, each consisting of a promoter sequence, SP-B transgene and termination sequence (Figure 3.4). The SP-B cDNA was designed based on the human *SFTPB* sequence published by Pilot-Matias *et al.* [336] and was congruent with sequences found on NCBI data base (NM_000542). In the SP-B cDNA sequence, two restrictions sites for ApaI, one for KpnI and one for NheI were removed to avoid interference with molecular cloning. A Kozak sequence ('CCRCC', here 'CCACC', [337]) and NheI restriction site were added before the start codon to insert the SP-B cDNA sequence in a rAAV genome cassette as depicted in Figure 3.4a. Additionally, a synthetic-optimised version of the SP-B cDNA (soSP-B) was designed (rationale for sequence optimisation discussed in Section 1.3.1.2). The SP-B cDNA was codon-optimised for codon usage frequency in humans using an algorithm made publicly available by Gene Art (Thermo Fisher Scientific). Further, all CpG sites and all splice sites with a score > 0.5 as identified with CBS intron analysis tool were removed (http://www.cbs.dtu.dk/services/NetGene2). To achieve this, codons were replaced with the next most common codon in humans that allowed removal of the CG dinucleotides and splice acceptor sites. In the present study, the SP-B or soSP-B sequence was paired with either the CMV promoter with its reported ubiquitous, high-level expression, including in murine ATII cells [321] or hCEFI promoter due its demonstrated long-term expression potential in the murine lung ([97], Figure 3.4a).

The SP-B and soSP-B cDNA sequences as described above were the foundation for the design of the SP-B/soSP-B minigene ('SP-B MG/soSP-B MG', Figure 3.4b), where the entire endogenous human *SFTPB* promoter sequence (from -1040 of start codon as

described in Pilot-Matias [336]) was used to drive transgene expression to allow endogenous transcriptional regulation as in *SFTPB*. Instead of a synthetic promoter proximal intron, the small promoter proximal intron 1 (ca. 300 bp) found in *SFTPB* was included at its endogenous site between exon 1 and 2 (Figure 3.4b). Promoter proximal introns have been shown to facilitate down-stream processing (Section 1.3.1.1), which is thought to be particularly relevant for complex proteolytic processing of mature SP-B. After intron 1, the minigene sequence continued with the SP-B/soSP-B cDNA sequence described above. In the SP-B minigene the endogenous Kozak sequence was retained (GTGCC) and in the soSP-B minigene replaced with CCACC.

After construction, expression from the rAAV genome cassettes was tested in the ATII cell-like A549 carcinoma cell line by transient transfection of the rAAV genome plasmids for SP-B expression using Lipofectamine 3000 (Section 2.4.2). Plasmids encoding the SP-B or soSP-B transgene, paired with the different promoters were transfected and lysates analysed by Western blot (Section 2.12) using the Anti-Pro + Mature Surfactant Protein B antibody (Table 2.11). A549 cells do not express SP-B without stimulation [338], but after transfection are expected to process the SP-B proprotein to the processing intermediate and mature SP-B homodimer (Figure 3.5a). Lysates from cells transfected with pAAV2.CMV.SP-B (pGM705) or pAAV2.CMV.soSP-B (pGM706, Table 2.1) were analysed in duplicate (Figure 3.5b). For both plasmids, three bands specific to transfection were observed (Figure 3.5b, arrows) when compared with 'mock' transfected cells, transfected with pAAV2.CMV.EGFP (pGM546, Table 2.1). The sizes of the observed bands (ca. 42, 22 and 16 kDa) align with sizes reported in the literature ([9], Figure 1.3) of various SP-B processed forms. For cells transfected with pAAV2.hCEFI.SP-B (pGM707) or pAAV2.hCEFI.soSP-B (pGM708, Table 2.1) three bands were observed in a similar pattern (Figure 3.5c, arrows). For cells transfected with the SP-B/soSP-B minigene (pAAV2.SP-B MG, pGM709 or pAAV2.soSP-B MG, pGM710), only a band corresponding

to the size of the mature SP-B homodimer was observed (Figure 3.5c). All plasmids enabled expression of the mature SP-B homodimer, but there was no consistent trend between expression levels, neither from plasmids with SP-B transgene (pGM705, pGM707, pGM709) and plasmids with soSP-B transgene (pGM706, pGM708, pGM710) nor between the ratio of band intensity of mature SP-B, processing intermediate and proprotein. In conclusion, in A549 cells grown in submerged culture, SP-B expression was observed from a range of rAAV genome configurations incorporating an SP-B cDNA or minigene, but there was no consistent or pronounced difference in expression levels between plasmids expressing SP-B and soSP-B transgenes.

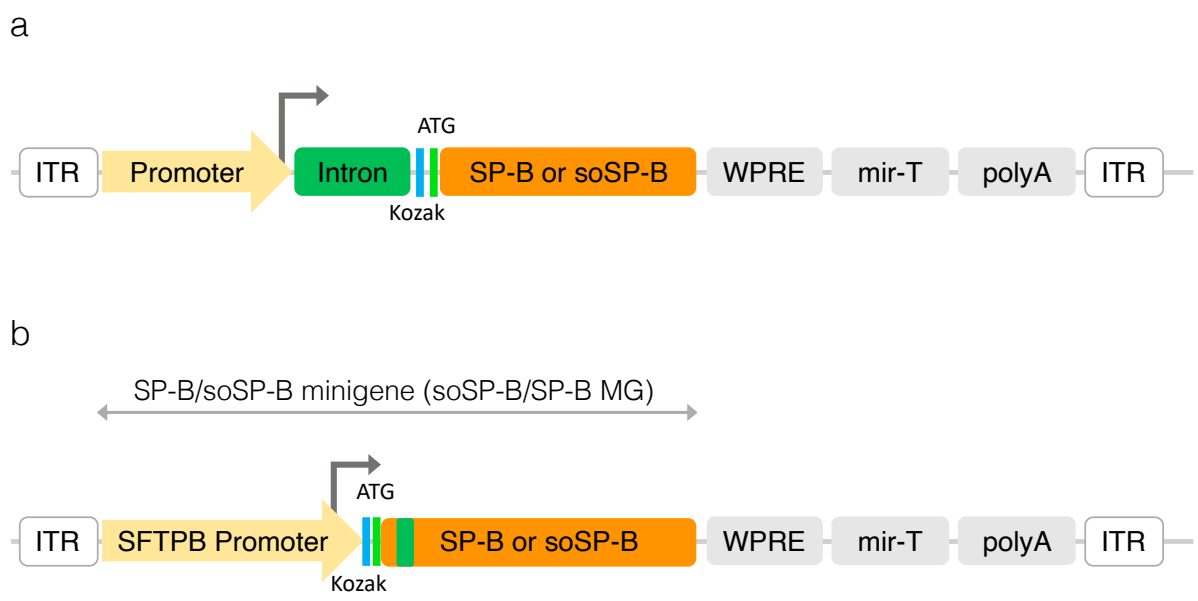


Figure 3.4 Schematic of rAAV expression cassettes for SP-B transgene delivery

The rAAV expression cassette encoded a promoter sequence (yellow), a transgene (orange) and a termination sequence (grey). The transgene was either the human cDNA sequence of SP-B, or a synthetic optimised (soSP-B) version. The promoter directing gene expression was either the CMV (pAAV2.CMV.SP-B and pAAV2.CMV.soSP-B) or hCEFI promoter (pAAV2.hCEFI.SP-B and pAAV2.hCEFI.soSP-B, a). The SP-B minigene (SP-B/soSP-B MG) cassette consists of the endogenous, human *SFTPB* promoter followed by the human SP-B cDNA and the endogenous promoter proximal intron between exon 1 and 2 as in *SFTPB* (pAAV2.SP-B MG and pAAV2.soSP-B MG, b). In the synthetic optimised version of the minigene, the Kozak sequence was altered and the cDNA was codon optimised in the same fashion as in soSP-B. ITR, inverted terminal repeats (white); polyA, poly adenylation signal; WPRE, wood chuck hepatitis virus response element (grey); mirT, microRNA target (grey).

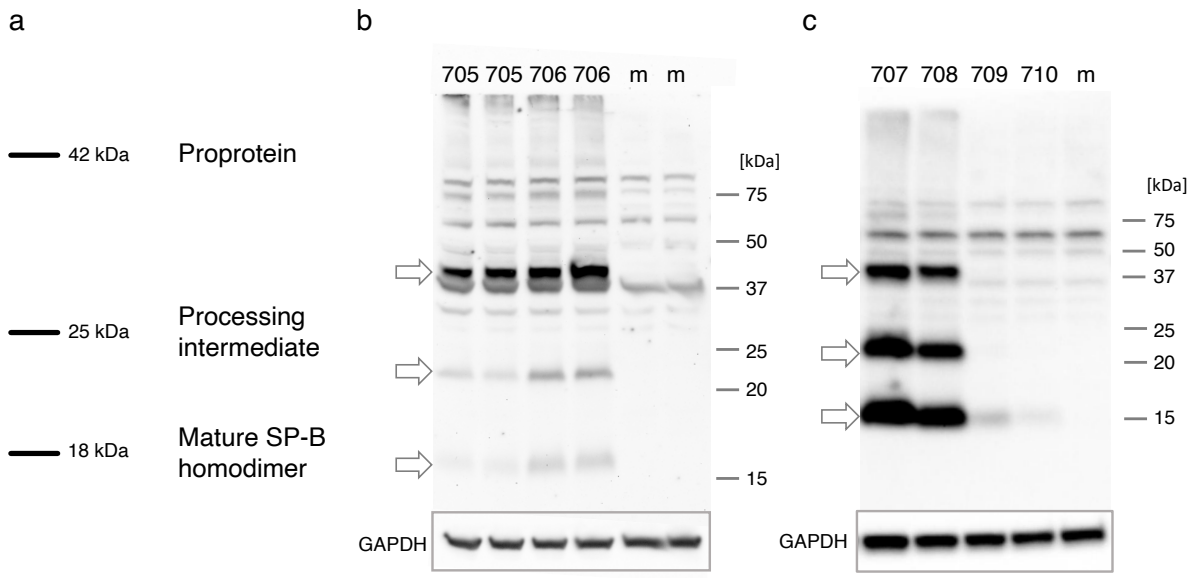


Figure 3.5 Western Blot of A549 cells transfected with plasmids expressing SP-B and soSP-B transgenes

A549 cells cultured in a 6-well format were either mock transfected or transfected at 80 % confluency with 2 µg of rAAV genome plasmids expressing SP-B or soSP-B under different promoters (pGM705-710). At 48 h post transfection, lysates were harvested, normalised to total protein (20 µg, b; 11 µg, c) and analysed for SP-B expression in Western blot analysis using the Anti-Pro + Mature Surfactant Protein B antibody. The reported sizes of the SP-B proprotein, processing intermediate and mature homodimer are indicated (a). Western blot (b) showing expression from plasmids pAAV2.CMV.SP-B (pGM705) and pAAV2.CMV.soSP-B (pGM706). Western Blot (c) showing expression from plasmids pAAV2.hCEFI.SP-B (pGM707), pAAV2.hCEFI.soSP-B (pGM708), pAAV2.SP-B MG (pGM709) and pAAV2.soSP-B MG (pGM710). Arrows indicate bands specific to transfection with SP-B/soSP-B plasmids. Subsequent immunoblotting for Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was included as control to demonstrate consistent sample loading.

The rAAV expression cassettes (pGM705-pGM710) were packaged into AAV5, which was identified as the serotype with the highest expression levels in murine ATII cells (see Section 3.2.1). After triple transfection in HEK293T cells in parallel and purification of all vectors (Section 2.5.1), the titre and yield of each production was determined (Figure 3.6). Yields were plotted grouped by promoter. The highest vector yields were achieved in cells transfected with genome plasmids including the *SFTPB* promoter (pGM709-710), followed by genome plasmids using the hCEFI promoter (pGM707-708). Cells transfected with genome plasmids including the CMV promoter (pGM705-706) achieved lowest yields. The promoter in the genome plasmid strongly impacted the mean yield of each vector group by ca. 10-fold, but all preparations yielded sufficient vector for further experiments. Due to a lack of statistical power, the differences observed in rAAV vector yield failed to reach significance, though the magnitude of the observed trend (approximately 2-logs between CMV and *SFTPB* promoter) is highly suggestive of an important production limitation with the CMV vector.

In summary, two versions of human SP-B cDNA for rAAV gene transfer were designed and each incorporated into rAAV expression cassettes paired with promoters CMV, hCEFI, or *SFTPB* in a 'minigene' version. The ability of each pAAV2 genome plasmid to produce mature SP-B was confirmed in A549 cells, and packaging of each plasmid in rAAV5 resulted in vector yields sufficient for animal experiments. For work described in the next Sections/Chapters, experiments used the SP-B rather than the soSP-B expression cassette.

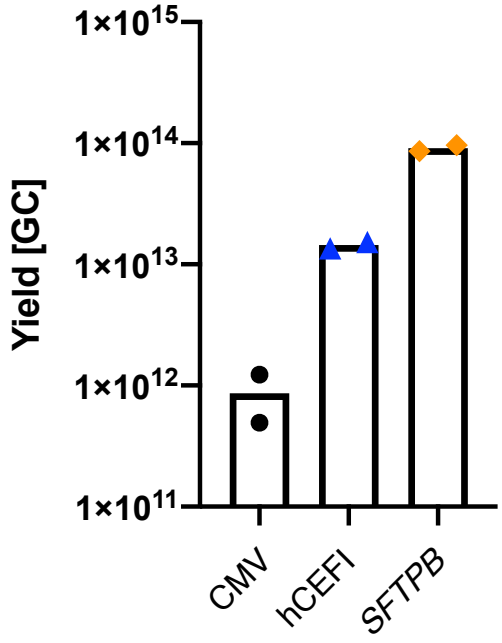


Figure 3.6 Production yields of rAAV5 vectors containing SP-B and soSP-B expression cassettes

Recombinant AAV5 vectors were produced by triple plasmid transfection of adherent HEK293T cells in 1720 cm^2 HYPERflasks using genome plasmids encoding the CMV promoter (pAAV2.CMV.SP-B, pGM705 and pAAV2.CMV.soSP-B, pGM706), the hCEFI promoter (pAAV2.hCEFI.SP-B, pGM707 and pAAV2.hCEFI.soSP-B, pGM708) or encoding the *SFTPB* promoter (pAAV2.SP-B MG, 709 and pAAV2.soSP-B MG, pGM710). Vectors were purified using Iodixanol gradient centrifugation, concentrated and rAAV vector concentration (Genome copies (GC)/mL) determined using qPCR detecting WPRE for calculation of vector yields. Vector production yields of vectors produced for SP-B or soSP-B transgene expression were plotted grouped by the promoter (CMV: pGM705-706; hCEFI: pGM707-708; *SFTPB*: pGM709-710) used for their expression. Vector yields (GC) are plotted with mean on a logarithmic scale. Kruskal-Wallis test was non-significant.

3.2.3 Development of RNAscope to quantify therapeutic transgene expression after rAAV delivery

The aim for vector selection was to determine which promoter confers the highest SP-B expression level in the target ATII cells *in vivo* via a quantitative method. The focus lies solely on ATII cells as expression in this cell type only is expected to have a therapeutic effect [42]. Western blot is a semiquantitative method suitable for analysis of protein expression levels, however, SP-B is highly conserved across species and available antibodies cannot differentiate between human and murine SP-B. As an alternative, qPCR may be exploited to differentiate between endogenous and transgene expression, e.g. by targeting a post-regulatory element of transgene mRNA. However, qPCR cannot differentiate between expression from ATII cells and another cell type. Therefore, the ISH-based method RNAscope, which combines quantification of expression with localisation information was chosen to compare expression levels from transgene cassettes using CMV, hCEFI and *SFTPB* promoters (pGM705-710, Table 2.1). The RNAscope technique is able to visualise each mRNA molecule as one 'dot' in a tissue section by signal amplification (Figure 2.1). The dot size or brightness is a measure of efficiency of signal amplification only, and does not reflect the quantity of mRNA. Section 2.14.1 describes the details of the RNAscope method used here.

The vector-targeting RNAscope probe was designed to hybridise to WPRE, because it provides flexibility as it is a common element in all rAAV vectors used and because of the high (64 %) sequence homology between *Sftpb* and *SFTPB*. To establish the method RNAscope for detection of WPRE expression after rAAV transgene delivery, mice (n= 3) were dosed intranasally with 1E11 GC of rAAV5.CMV.SP-B (Table 2.6) or remained naive (n= 3). After 14 days, the lungs were processed in fixed-frozen sections and a post-fixation protocol established to avoid detachment of tissue sections from the slide during the

assay (Section 2.14.1). Sections were stained with a negative control probe (Table 2.13) targeting the bacterial gene *DapB* to estimate the frequency of false negative RNAscope signal; as shown in Figure 3.7a, no RNAscope signal was detected. To test mRNA quality of the sections and the suitability for mRNA detection, sections were stained with a positive control probe (Table 2.13) targeting the mRNA of the highly expressed endogenous *Ubc* gene (Figure 3.7b). High levels of RNAscope signal was observed as multiple 'dots' throughout the section, confirming the retention of mRNA at sufficient quality and functionality for the RNAscope assay.

The WPRE-targeting RNAscope probe WPRE-O1 ('WPRE probe', Table 2.13), but not the WPRE-mut6 probe (#524721), was experimentally identified as suitable for WPRE detection from the manufacturer's portfolio (data not shown). Sections from naive and rAAV5.CMV.SP-B dosed mice were costained with the WPRE probe and the *Sftpb* probe (Table 2.13) for staining murine endogenous *Sftpb* expression as a ATII cell marker (Figure 3.7c-f). The *Sftpb* probe used a custom design specifically targeting regions in *Sftpb* that are different from SP-B cDNA to ensure no cross-reaction with expression of human SP-B.

No WPRE RNAscope signal was observed in sections from naive mice stained with the WPRE probe (Figure 3.7c) and for the simultaneously stained endogenous murine *Sftpb* mRNA, high levels of RNAscope signal were detected (Figure 3.7d). No individual dots were observed, but instead the *Sftpb* mRNA signal appeared to fill entire cells, likely due high expression of *Sftpb* mRNA at this level of image resolution which was observed in parenchymal cells with a shape reminiscent of ATII cells and observed at the expected frequency for ATII cells; these cells were therefore surmised to be ATII cells. In sections from rAAV5 dosed mice, high levels of WPRE RNAscope signal (dots) were observed in

cells located in the lung parenchyma (Figure 3.7e). This shows that the RNAscope assay allowed for specific staining of WPRE in rAAV-dosed compared with naive mice.

In sections from rAAV5 dosed mice, individual WPRE signal dots were observed (Figure 3.7e, magnification), but often individual dots could not be observed at this high signal level and image resolution (Figure 3.7e, arrows). WPRE RNAscope signal was predominantly concentrated in 'hotspot' cells in the murine lung parenchyma that were identified as ATII cells by *Sftpb* coexpression (Figure 3.7f). However, scattered signal dots were observed throughout the section. Due to the rarity of false-positive signal in either the negative-staining control, or WPRE staining in naive mice, a false signal seemed unlikely. Theoretically, the scattered signal dots could stem from individual cells expressing individual WPRE mRNA, or from vector DNA of extracellular rAAV vectors, implying that the method used costains WPRE mRNA and WPRE vector DNA.

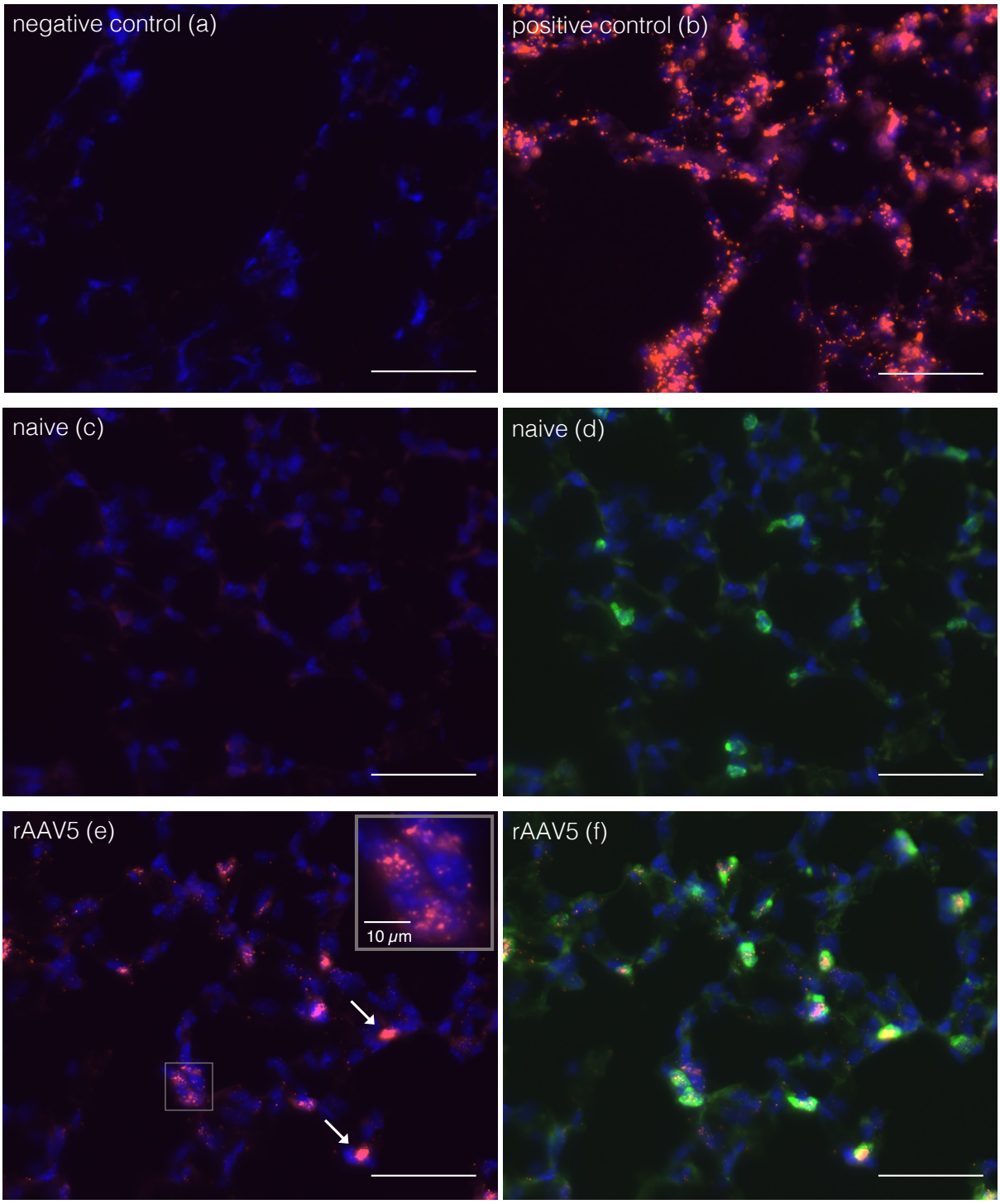


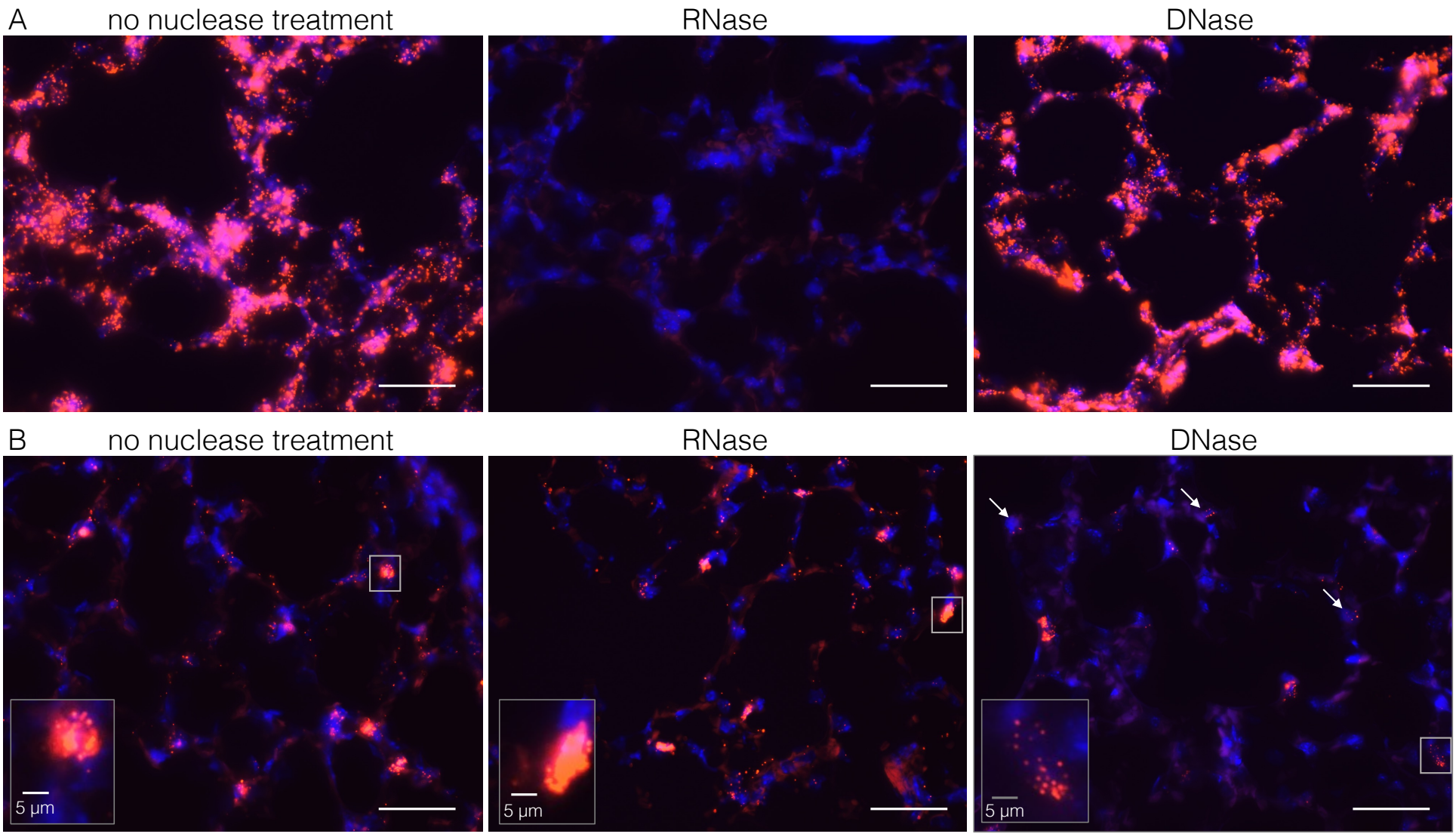
Figure 3.7 RNAscope in murine lung sections of naive mice and mice dosed with rAAV5.CMV.SP-B

Mice were dosed intranasally with 1E11 GC of rAAV5.CMV.SP-B (n= 3) or remained naive (n= 3). After 14 days, mice were culled and left lung lobes were processed in fixed-frozen sections. After a post-fixation and RNAscope pretreatment protocol, sections were stained using various RNAscope probes for RNAscope assay: Merged image of section

stained for nuclei (DAPI, blue) and with negative control probe (3-plex Negative Control Probe, #320871, red) targeting the mRNA expressed from bacterial gene *DapB* (a). Merged image of section stained for nuclei (DAPI, blue) and with positive control probe (3-plex Positive Control Probe-Mm, #320881, red) for the murine endogenous *Ubc* mRNA expression (b). Sections of naive mice (c, d) or rAAV5.CMV.SP-B dosed mice (e, f) were stained for vector transgene mRNA expression using the WPRE probe (WPRE-O1, #450261, red) and for murine, endogenous *Sftpb* expression using the *Sftpb* probe (Sftpb-noX-human, #539421-C2, green). Merged images of nuclei (DAPI, blue) and WPRE RNAscope signal (red) are shown in c and e. Merge of RNAscope signal for WPRE RNAscope signal (red), murine *Sftpb* RNAscope signal (green) and nuclei (DAPI, blue) are shown in d and f. Magnified image detail and examples of cells with RNAscope signal not discernible as single dots (white arrows) are shown (e). Scale bar 50 µm except where stated otherwise.

Sections were treated with nucleases just before application of the RNAscope probes (Section 2.14.2) in order to test the hypothesis that the WPRE probe stains both mRNA and vector DNA after rAAV transduction. Nuclease digests were first tested in sections stained for endogenous *Ubc* expression (Figure 3.8A). Digestion of slides with RNase removed virtually all the RNAscope signal, with only 0-2 dots per image remaining, indicating that the assay conditions were suitable for mRNA digestion and that the RNAscope exclusively stained mRNA from endogenous genes with no other targets such as genomic DNA. Digestion of similar slides with DNase left the sections unchanged, indicating that DNase digestion did not interfere with the RNAscope assay.

Nuclease digests were therefore performed in parallel in sections stained with WPRE probe (Figure 3.8B). In contrast to endogenous *Ubc* expression, RNase digestion did not cause a visible reduction of WPRE RNAscope signal, confirming that the WPRE probe targets nucleic acids other than mRNA. Digestion with DNase led to a marked reduction of WPRE RNAscope signal, with the majority of cells showing fewer WPRE signal dots and the absence of dots scattered across the section (Figure 3.8B). To confirm the remaining signal is mRNA, and that there is no residual DNA, slides were digested with both RNase and DNase (Figure 3.8C), which removed virtually all signal (0-2 dots per image). This confirmed that DNase digestion was suitable for DNA removal and that the remaining signal after DNA digestion was mRNA. Most likely, the signal removed by DNase digestion was vector DNA. In conclusion, a DNase digestion step is necessary to uncouple the vector DNA signal from mRNA transgene expression after rAAV transduction and by addition of a DNase digestion step (Section 2.14.2), a novel method was established suitable for exclusive staining of transgene mRNA expression and accurate quantification after rAAV vector transduction.


A
no nuclease treatment
RNase
DNase
B
no nuclease treatment
RNase
DNase
5 µm
5 µm
5 µm

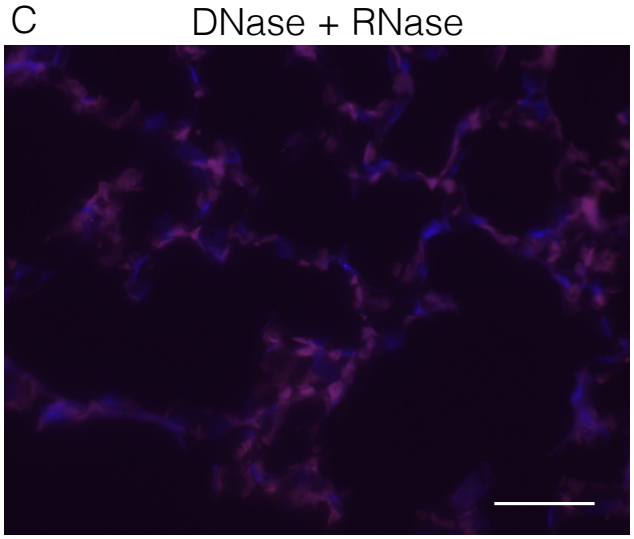


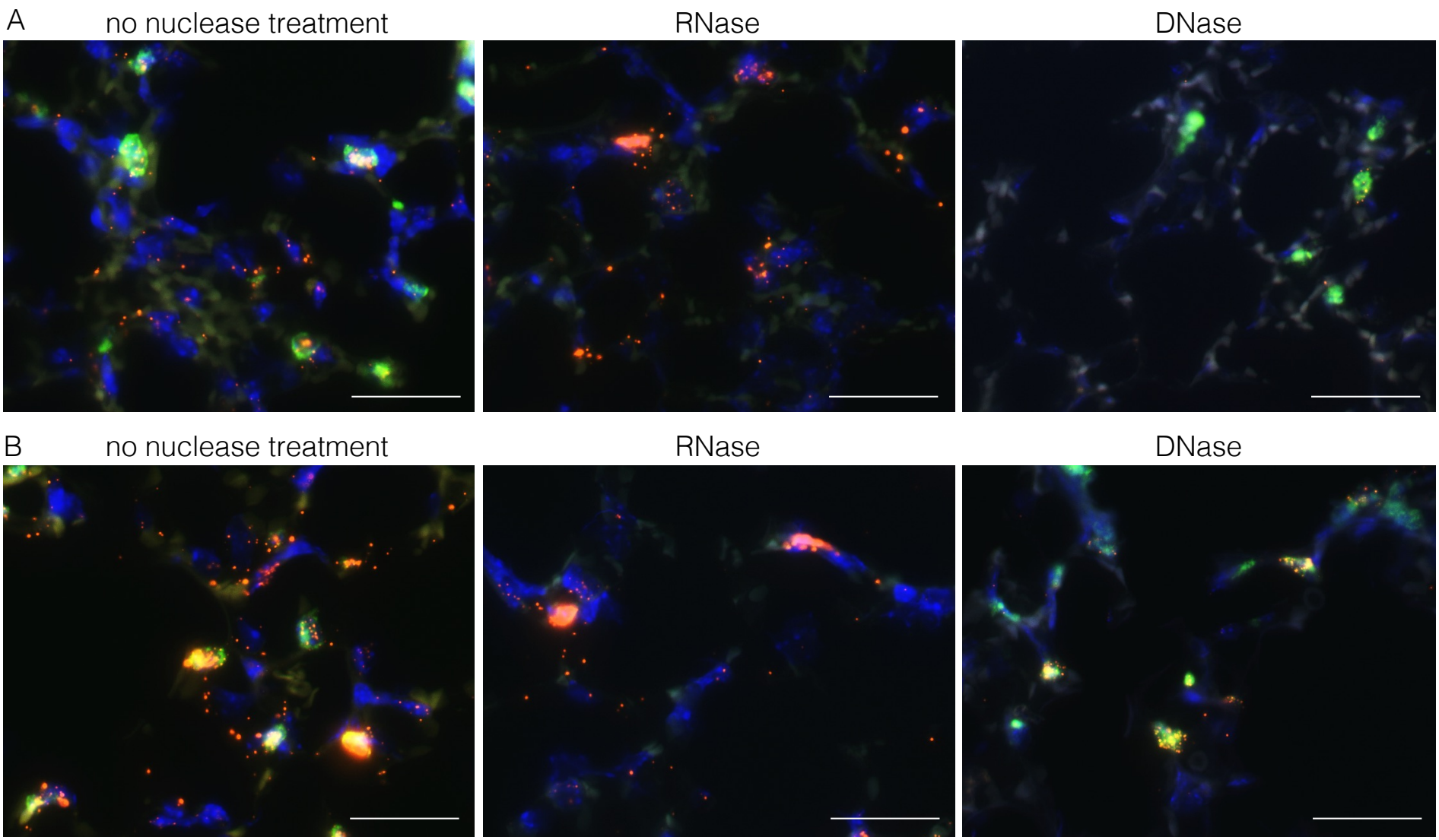
Figure 3.8 RNAscope in lung sections of rAAV-dosed mice using RNase and DNase digestion to understand the origin of RNAscope

Mice were dosed intranasally with 1E11 GC of rAAV5.CMV.SP-B (n= 3) and after 14 days, mice were culled and left lung lobes were processed in fixed-frozen sections. After post-fixation and RNAscope pretreatment, sections either remained untreated with nuclease or were treated with RNase, DNase or both. RNAscope assay was performed using the WPRE probe (WPRE-O1, #450261) or positive control probe (3-plex Positive Control Probe-Mm, #320881) targeting the murine endogenous *Ubc* mRNA expression. Merged images of nuclei (DAPI, blue) and RNAscope signal (red) are shown of sections treated with nucleases as indicated and stained with positive control probe (panel A) or WPRE probe (panels B, C). Magnified image details and white arrows indicate cells with punctate, individual signal dots in panel B. Representative images are shown. Scale bar 50 μm except where stated otherwise.

3.2.4 Identification of promoters driving high SP-B mRNA expression levels in murine ATII cells

Mice (n= 4) were dosed intranasally with 1E11 GC of rAAV5.CMV.SP-B, rAAV5.hCEFI.SP-B or rAAV5.SP-B MG (Table 2.6) with the aim to identify which promoter mediates the highest level of expression in murine ATII cells using the established RNAscope method including DNase pretreatment as established in Section 3.2.3 above.

Firstly, the reproducibility of the established RNAscope method including DNase pretreatment (Section 2.14.2) was examined across animal experiments by nuclease digestion of sections of one animal from each rAAV vector group with RNase or DNase (Figure 3.9). As observed in the previous animal experiment (section 3.2.3 above and Figure 3.8), RNase digestion had no effect on WPRE RNAscope signal in sections from mice dosed with rAAV5.CMV.SP-B, but DNase digestion reduced the signal within cells and removed scattered signal dots across the section (Figure 3.9A); this confirmed reproducibility across animal experiments with the same vector and RNAscope probes. The same effect was observed following nuclease digestion of sections from mice dosed with rAAV5.hCEFI.SP-B and rAAV5.SP-B MG (Figure 3.9B and C). Double digestion with DNase and RNase confirmed that the remaining signal after DNase digestion is RNA (Figure 3.9D) and demonstrated that the effect of nuclease digestion is reproducible. In conclusion, DNase pretreatment was a necessary step in the protocol for RNAscope staining and quantification of mRNA expression, in order to avoid detection of vector DNA as a confounding factor after rAAV delivery.


A
no nuclease treatment
RNase
DNase
B
no nuclease treatment
RNase
DNase

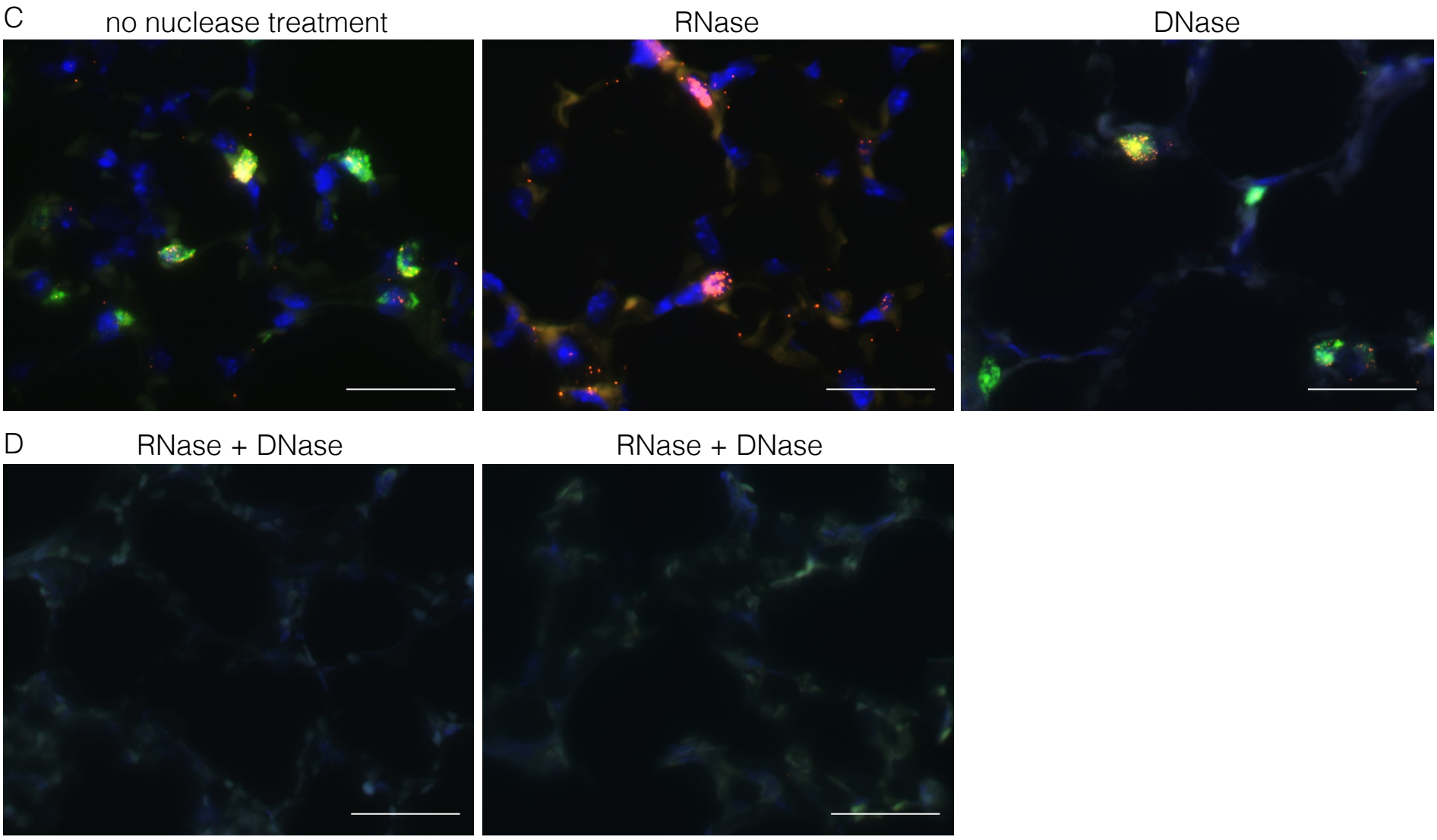

C
no nuclease treatment
RNase
DNase
D
RNase + DNase
RNase + DNase

Figure 3.9 Uncoupling of the WPRE mRNA expression signal from DNA in sections of mice dosed with rAAV vectors expressing SP-B under various promoters

Mice (n= 4) were dosed intranasally with 1E11 GC of rAAV5.CMV.SP-B, rAAV5.hCEFI.SP-B or rAAV5.SP-B MG. On day 14 after dosing, mice were culled and lungs processed in fixed-frozen lung sections. After post-fixation and RNAscope pretreatment, left lung lobe sections either remained nuclease untreated or were treated with RNase, DNase or both. RNAscope assay was performed using the WPRE probe (WPRE-O1, #450261) together with the *Sftpb* probe (Sftpb-noX-human, #539421-C2). Merged images of nuclei (DAPI, blue) WPRE RNAscope signal (red), and *Sftpb* RNAscope signal (green) showing sections from mice doses with of rAAV5.CMV.SP-B (A), rAAV5.hCEFI.SP-B (B) or rAAV5.SP-B MG (C). Double digests (D) were performed on sections of the same animal as in A (left) and B (right). All images in one panel (A-C) are from the same animal. Scale bar 50 µm.

Once the reproducibility of the RNAscope staining protocol including DNase pretreatment was established, one section from each mouse was stained using this protocol (Section 2.14.2), including controls for RNase and DNase digestion. Sections were imaged in a systematic fashion. For analysis, the area of interest was marked and dots counted using an ImageJ macro (Table 2.17), providing a lower limit brightness cut-off for signal. Only cells with ≥ 2 signal dots were counted as transduced and only transduced cells, in image focus, identified as ATII cells (parenchymal cells with positive *Sftpb* RNAscope signal) were quantified. Examples of WPRE mRNA signal quantification are shown in Figure 3.10. Distinct RNAscope signal dots were observed in all treatment groups, but not in naive mice. On rare occasions the signal was so high that it accumulated to uniformly fill the entire cell making it difficult to quantify. This was the case for four analysed cells in the rAAV5.CMV.SP-B treatment group, 15 cells for rAAV5.CEFI.SP-B and six cells for rAAV5.SP-B MG. This, together with the likelihood that signal dots overlap and not all cell planes were imaged, suggest that the mRNA molecules quantified per cell represent a minimum number and are likely underestimated in all groups.

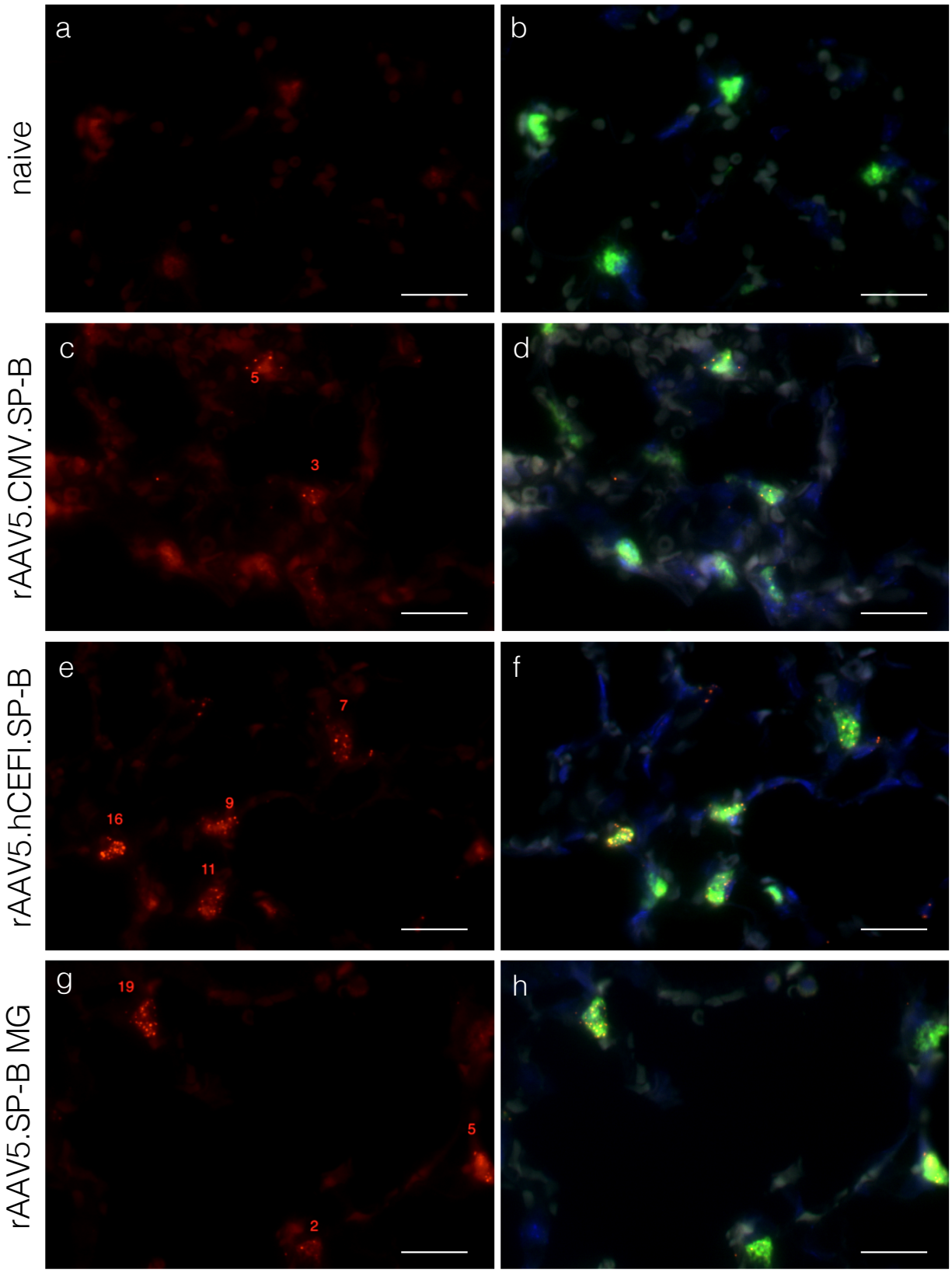


Figure 3.10 Quantification of SP-B transgene mRNA expression in transduced ATII cells after transduction of mice with rAAV vectors expressing SP-B under various promoters

Mice (n= 4) either remained naive or were dosed intranasally with 1E11 GC of rAAV5.CMV.SP-B, rAAV5.hCEFI.SP-B or rAAV5.SP-B MG. On day 14 after dosing, mice were culled and lungs processed in fixed-frozen lung sections. After post-fixation and RNAscope pretreatment, the RNAscope DNase pretreatment was applied to decouple mRNA signal from vector DNA. RNAscope assay was performed in one section of each animal using the WPRE probe (WPRE-O1, #450261) together with the *Sftpb* probe (Sftpb-noX-human, #539421-C2). Using the scanning function on the EVOS FL Auto 2, every two image frames one image was taken, after setting the RNAscope signal in focus. If there was no transduced cell apparent, no image was taken. Cells with ≥ 2 signal dots were counted as transduced and only cells in focus, with positive *Sftpb* mRNA stain and counted as transduced, were quantified. Only signal dots within the cell as determined by *Sftpb* mRNA stain were counted. The cell perimeter was marked and RNAscope signal dots counted using an ImageJ macro, providing a lower limit brightness cut-off for signal. The successful differentiation of individual RNAscope signal dots by the macro was assessed manually. Examples of images quantified for each treatment group are shown for naive mice (a, b) and mice dosed with rAAV5.CMV.SP-B (c, d), rAAV5.hCEFI.SP-B (e, f) or rAAV5.SP-B MG (g, h). Left column images show the WPRE RNAscope signal only (a, c, e, g) together with determined values for WPRE mRNA signal. Right column images are merged images of nuclei (DAPI, blue), WPRE RNAscope signal (red), and *Sftpb* RNAscope signal (green). Scale bar 25 µm.

In each animal 55-59 cells were quantified with the method described above using an ImageJ macro (Table 2.17) and as demonstrated in Figure 3.10. The identified number of WPRE mRNA molecules per cell were plotted grouped for each animal in Figure 3.11a. There was little variance between animals within a group. To analyse the effect of the promoters on mRNA expression level, all cells from all animals within each experimental group (n= 4 per group) were combined (Figure 3.11b). WPRE mRNA expression levels were significantly highest in ATII cells of mice transduced with rAAV5.hCEFI.SP-B with a mean expression of 11.1 transgene mRNA copies per analysed cell ($p= 0.044$ versus rAAV5.SP-B MG and $p< 0.0001$ versus all other groups). Expression from rAAV5.SP-B MG achieved a mean expression level of 8.9 transgene mRNA copies per cell, significantly lower than the hCEFI promoter ($p= 0.044$) and significantly higher ($p< 0.0001$) than the CMV promoter, with a mean of 5.1 transgene mRNA copies per cell. All rAAV5 dosed groups were significantly different ($p< 0.0001$) from naive where no transduced cells were observed. The data are summarised as the mean expression level in each animal (Figure 3.11c). In summary, the promoter expression levels were hCEFI > SP-B minigene > CMV in murine ATII cells *in vivo*. Thus, the hCEFI promoter would be the promoter of choice for correction of SP-B deficiency in the murine disease model.

The RNAscope signal in DNase-untreated sections was often too high for quantification. For an estimate of the effect of DNase treatment on quantification, the percentage of ATII cells positive for WPRE RNAscope signal (% *Sftpb*-positive cells with ≥ 2 signal dots) was determined (Figure 3.12). The positive WPRE RNAscope signal rate in groups not treated with DNase was greater than in DNase-treated sections across experimental groups. The mean positive WPRE RNAscope signal rate determined in untreated sections ranged from 95-77 %, while mean positive WPRE RNAscope signal rate in DNase-treated groups was 78 % for hCEFI, 46 % for SP-B minigene and 22 % for CMV. This showed that DNase digestion had a distinct effect on quantification of transgene mRNA expression levels and

large proportions of the signal detected in sections not treated with DNase can be attributed to vector DNA.

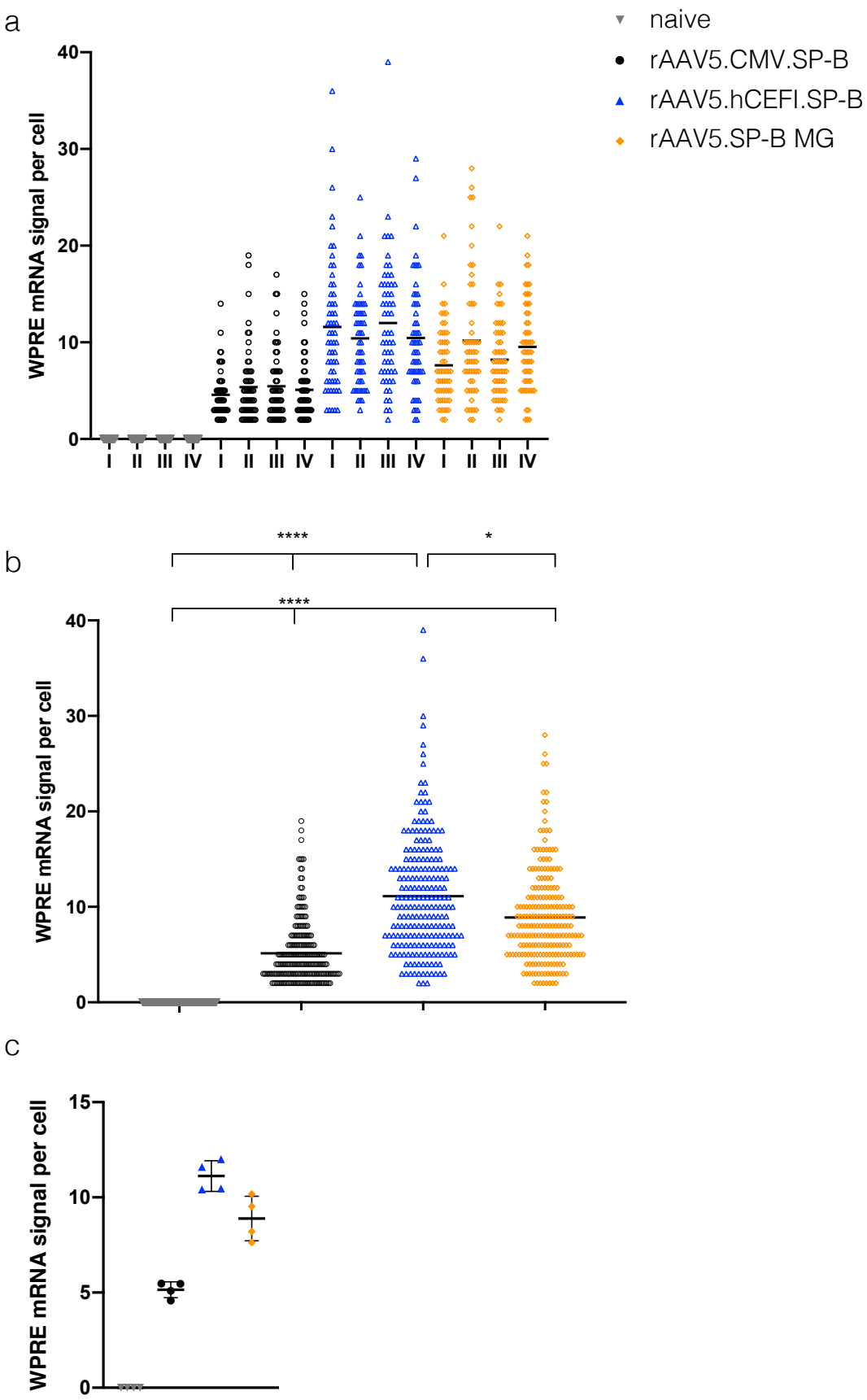

a
naive
rAAV5.CMV.SP-B
rAAV5.hCEFI.SP-B
rAAV5.SP-B MG
WPRE mRNA signal per cell
40
30
20
10
0
I II III IV I II III IV I II III IV I II III IV
b
****
*
****
WPRE mRNA signal per cell
40
30
20
10
0
c
WPRE mRNA signal per cell
15
10
5
0

Figure 3.11 Quantification of SP-B transgene mRNA expression in transduced *Sftbp* positive cells after rAAV gene transfer

Mice (n= 4) either remained naive or were dosed intranasally with 1E11 GC of rAAV5.CMV.SP-B, rAAV5.hCEFI.SP-B or rAAV5.SP-B MG. On day 14 after dosing, mice were culled and lungs processed in fixed-frozen lung sections. After post-fixation and RNAscope pretreatment, the RNAscope DNase pretreatment was applied to decouple mRNA signal from vector DNA. RNAscope assay was performed in one section of each animal using the WPRE probe (WPRE-O1, #450261) together with the *Sftpb* probe (Sftpb-noX-human, #539421-C2). Sections were imaged in a systematic fashion and WPRE RNAscope signal dots within the cell perimeter quantified by an ImageJ macro. Cells with ≥ 2 signal dots were counted as transduced and only cells in focus, identified as ATII cells (parenchymal cells with positive *Sftpb* mRNA stain) and counted as transduced, were quantified. In each animal, 55-59 cells were quantified. The quantified WPRE signal dots in each analysed cell are shown grouped for each animal (a) or grouped for each experimental group (b) and are summarised as average WPRE mRNA signal per cell found in each animal (c). Each symbol represents one cell (a, b) or mean with SD of one animal (c). Kruskal-Wallis test with Dunn's post-hoc test for multiple comparisons comparing means of each group was performed in (b). (*) $p < 0.05$; (****) $p < 0.0001$.

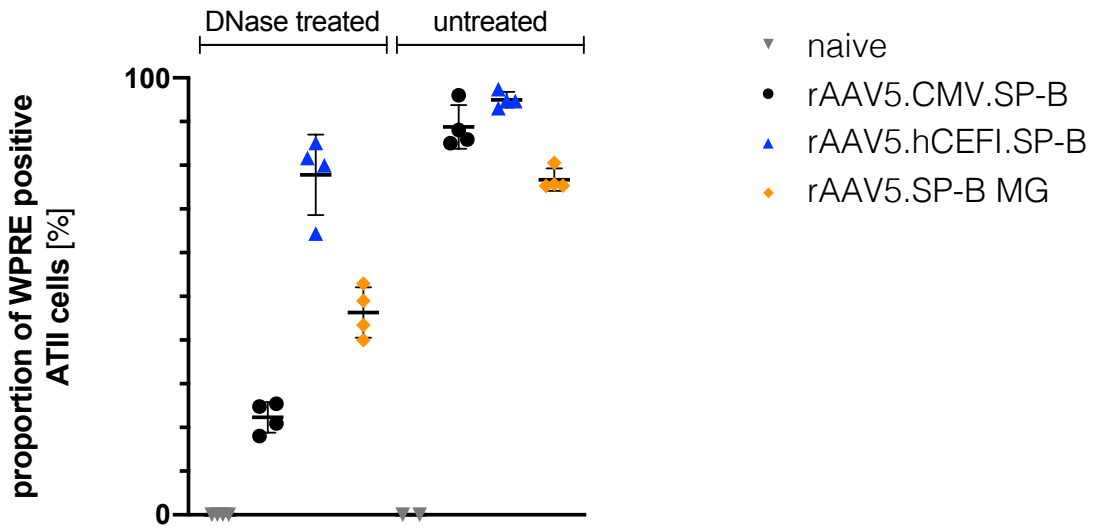


Figure 3.12 Percentage of transduced ATII cells in DNase treated and untreated murine lung sections

Mice (n= 4) either remained naive or were dosed intranasally with 1E11 GC of rAAV5.CMV.SP-B, rAAV5.hCEFI.SP-B or rAAV5.SP-B MG. On day 14 after dosing, mice were culled and lungs processed in fixed-frozen lung sections. After post-fixation and RNAscope pretreatment, sections either remained untreated or RNAscope DNase pretreatment was applied. RNAscope assay was performed in one section of each animal (n= 4, exception naive untreated n= 2) using the WPRE probe (WPRE-O1, #450261) together with the *Sftpb* probe (Sftpb-noX-human, #539421-C2). Sections were systematically scanned and images taken and analysed for transduced ATII cells (cells in the murine lung parenchyma with positive *Sftpb* RNAscope signal and with ≥ 2 WPRE RNAscope signal dots). During systematic imaging, only images with ≥ 1 transduced cell were included in analysis. For each analysed animal 71-81 ATII cells were randomly selected and the proportion of transduced cells determined. Each symbol represents the transduction rate in ATII cells observed in one animal. Kruskal-Wallis test with Dunn's post hoc test for multiple comparisons comparing mean transduction rates determined in untreated versus DNase treated section for each vector was significant for rAAV5.CMV.SP-B ($p < 0.0312$) and non-significant for all other groups.

3.2.5 SP-B protein expression levels in murine BALF after rAAV delivery *in vivo*

BALF from mice used to investigate transgene mRNA expression (Section 3.2.4, above) was examined for SP-B protein expression using Western Blot analysis to determine if the observed differences in mRNA expression translated into differences in protein expression. Figure 3.13a shows Western blot (Section 2.12) for SP-B on BALF samples (Section 2.6.4) from two naive mice. One primary antibody-specific protein band migrating at ca. 21 kDa was detected. As only mature SP-B protein is expected in the BALF (Section 1.2.3), it was concluded that this ca. 21 kDa band most likely corresponds to the mature SP-B homodimer, which appeared to migrate with a slightly higher apparent molecular weight than the reported 16-18 kDa for murine SP-B [42, 321].

To compare SP-B protein expression levels, BALF was analysed for SP-B as well as both GAPDH and SP-A as loading controls (Figure 3.13b). GAPDH is a commonly used endogenous protein for normalisation, however, during BALF collection blood can contaminate the samples. Here, red coloration of the samples was observed for mice naive mouse III, rAAV5.CMV.SP-B mouse III and IV, as well as rAAV5.hCEFI.SP-B-dosed mouse I and to lesser extent mice II-IV. Therefore, SP-A, which is a protein present in pulmonary surfactant was chosen as a more reliable control for total protein normalisation. SP-A was observed to migrate at ca. 130 kDa which is attributed to the SP-A trimer (105 kDa reported, [339, 340]). SP-A and GAPDH band intensities were variable across animals with no specific trend amongst experimental groups. SP-B band intensities appeared fairly consistent, with variable elevations in mice dosed with rAAV5.hCEFI.SP-B vector. BALF was analysed in two Western blot technical replicates and the SP-B band intensity normalised to the SP-A loading control (Figure 3.13c). There was no significant difference in SP-B expression levels between experimental groups or in comparison with naive animals. The SP-B expression levels in rAAV5.hCEFI.SP-B dosed mice appeared

slightly elevated compared to the other treatment groups, but this trend was not statistically significant, nor consistent across technical Western blot replicates.

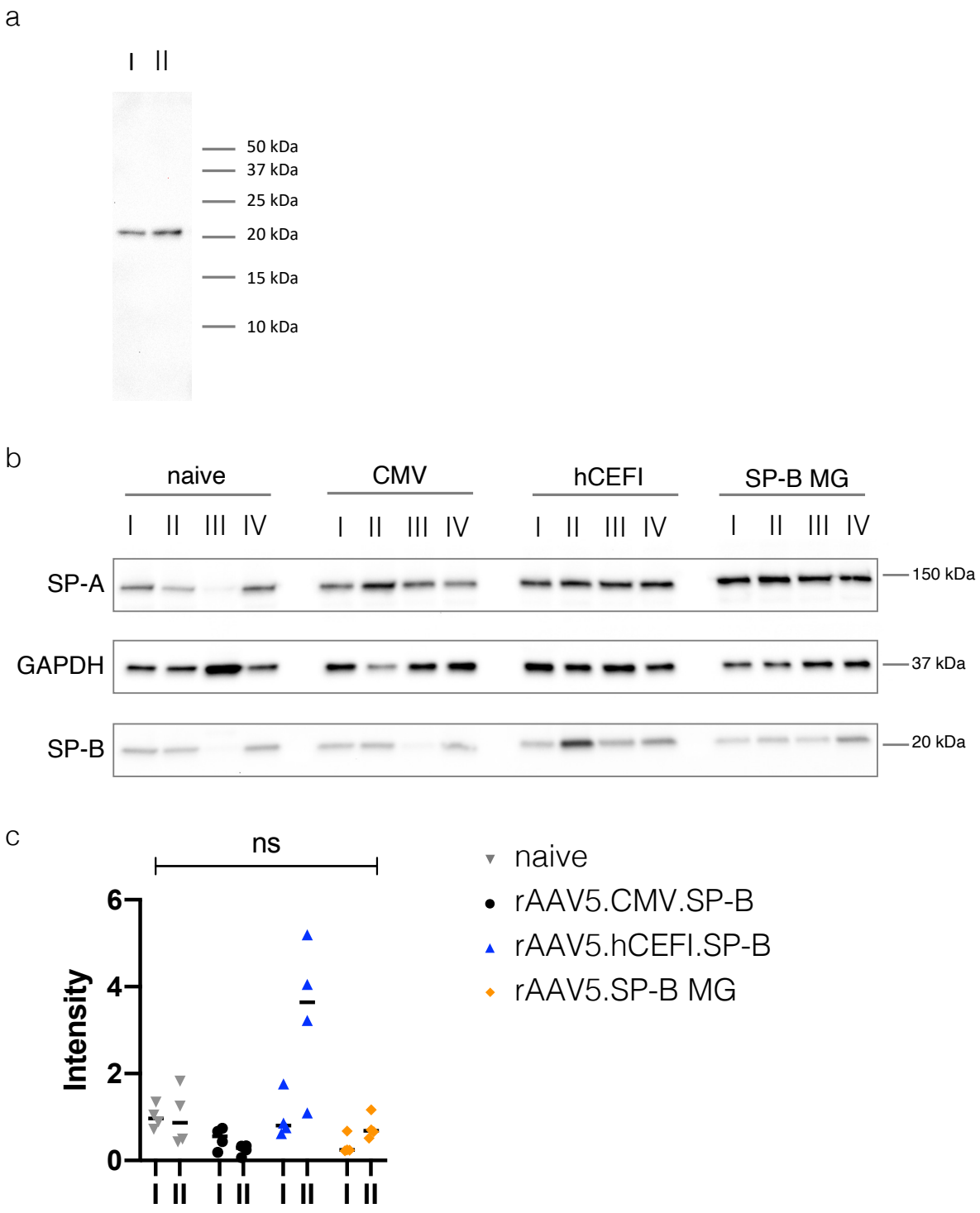


Figure 3.13 Western blot analysis of SP-B expression in murine BALF

Mice (n= 4) either remained naive or were dosed intranasally with 1E11 GC of rAAV5.CMV.SP-B, rAAV5.hCEFI.SP-B or rAAV5.SP-B MG. On day 14 after dosing, mice were culled and bronchoalveolar lavage fluid (BALF) sampled. BALF was analysed in Western blot analysis for SP-B using Anti-Pro + Mature Surfactant Protein B antibody, as well as SP-A and GAPDH as controls for even BALF sampling and total protein loading.

Western blot analysis for SP-B of BALF from two naive mice showing the entire SDS gel, including stacking gel (a). Western blot analysis for SP-A, GAPDH and SP-B of BALF from wildtype mice (n= 4) dosed with rAAV5 vectors expressing SP-B under various promoters (b). Quantification of SP-B and SP-A band intensities observed in two technical Western blot replicates plotted as SP-B band intensities normalised to SP-A and mean normalised SP-B band intensity of the naive group (c). Kruskal-Wallis test with Dunn's post hoc test for multiple comparisons.

3.3 Discussion

3.3.1 Selection of rAAV5 as a candidate serotype for transduction of murine ATII cells *in vivo*

In the studies described in this Chapter, the AAV serotypes 5, 6.2 and 9 were preselected based on a study published by Limberis *et al*. investigating multiple serotypes in murine airways and parenchyma where the results were reported by attributing scores based on transduction efficacy to each vector, with the highest performing serotype being 6.2 (6.2 > 6 = 5 > 9, [283]). Perhaps unexpectedly, all best performing serotypes in the lung parenchyma (with exception of 6/6.2) stem from different clades [283], with rAAV5 most distantly related (Figure 1.6). This indicates that it is difficult to predict AAV serotype function on the basis of the capsid amino acid sequence alone as these vectors are likely to transduce due to different receptors (Table 1.1) and emphasises the importance of experimental investigation.

The selected serotypes 5, 6.2 and 9 were compared after intranasal delivery for overall lung transduction rate (Figure 3.1) and the transduction rate in ATII cells (Figure 3.3). Although transduction in ATII cells was the ultimate aim, the overall lung transduction was a useful comparator with the published literature and provided a second measure for transduction rate using a large sample size, namely two whole lung cross-sections. For determination of transduction rate in ATII cells, confocal microscopy was used because of its improved signal-to-noise ratio which allowed for identification of transduced cells without digital autofluorescence correction in the highly autofluorescent lung tissue.

Limberis *et al.* identified rAAV6.2 as the serotype with the highest transduction in the lung parenchyma as well as for overall murine lung transduction, but did not study the transduction rate in ATII cells specifically [283]. In this chapter rAAV5 outperformed rAAV6.2 and it is conceivable that these different findings stem from quantification of

overall parenchymal transduction by Limberis *et al.* versus ATII cells only in this chapter. Additionally, parenchymal transduction was not quantified as percentage transduction, but transduced cells per field. Lung inflation is variable between animals, which means that the number of cells captured in one field is variable, adding a confounding factor; therefore, in the studies described in this chapter, transduced cells were directly quantified as either a proportion of ATII cells, or all cells (Section 3.2.1).

Limberis *et al*. provided an extensive overview of various naturally occurring and engineered serotypes in the murine lung but showed no direct comparison or statistical significance for serotypes 5, 6 or 6.2 in the murine lung parenchyma, only in the murine nose and airways [283]. For specific transduction of murine ATII cells, rAAV5 resulted in the highest transduction rate of 19.4 % and the only transduction rate significantly different from naive (Figure 3.3). Statistical testing was performed using the non-parametric Kruskal-Wallis test because values for percentages are fixed between 0-100 and therefore normal distribution cannot be assumed. Further experiments with more animals are needed to confirm a statistically significant difference of transduction rate in ATII cells between AAV serotypes.

For overall lung transduction, automated counting of EGFP-positive cells in whole lung sections was performed (Section 3.2.1) whereas in the Limberis *et al.* study [283], human AAT secretion into the serum was measured. Therefore, it is possible that depending on the cell types preferentially transduced by a specific serotype, and the accompanying reporter secretory activity, measurement of EGFP versus hAAT secretion could lead to different outcomes. Limberis *et al.* also used the CBA promoter, whereas studies in this chapter used CMV, which again could influence outcome. All of these factors, as well as different vector preparations and dosing route (e.g. intratracheal versus intranasal) can influence transduction rate, therefore identifying the best performing vector in our model

was an important first step. Another study compared rAAV5, 6.2 and 9 for transduction of murine parenchyma and found minimal transduction with rAAV5, but vector was delivered *in utero* and confounded by delivery route and potential differences in entry receptor expression in foetal versus adult mice in comparison with the here presented study [341].

In this chapter, transduction of the murine airway was observed with rAAV6.2 and 9, but not with rAAV5 (Figure 3.2); and this specificity for transducing parenchyma makes rAAV5 particularly attractive for an SP-B gene therapy vector as it preemptively reduces expression in non-target cells and thus decreases pressure on manufacturing. Although rAAV5 transduction was observed in the murine airways in other studies [218, 283, 342], parenchymal transduction was always dominant over airway or reported as very rare (> 1 % of transduction in airways, [218]). Possibly, rAAV5 transduction in murine airways in the present study, as identified by imaging EGFP in whole lung sections, was not observed because it was rare and obscured by high levels of airway autofluorescence. Notably, the observed whole lung cell transduction rate for rAAV was in line with ATII cell transduction rate (Section 3.2.1): the parenchyma is the largest region of the lung and the ratio of ATII:ATI cell type is ca. 3:2 in rodents, albeit ATI cells with their elongated shape cover the majority of the alveolar surface [343].

3.3.2 Validation of rAAV SP-B expression cassettes for rAAV production

After serotype selection, a rAAV transgene cassette was required to generate high expression levels in murine ATII cells *in vivo*. Transgene cassettes with CMV, hCEFI and *SFTPB* promoters, combined with either of two versions of SP-B transgene (SP-B or soSP-B), all resulted in expression of mature SP-B protein as determined by Western blot analysis (Figure 3.5). Direct comparison of SP-B and soSP-B showed no consistent difference in relative expression, although subtle differences might be detected with more biological replicates, or in other models with a more similar transcription profile to human

ATII cells such as H441 cells grown in air-liquid interface [344]. As the soSP-B construct was optimised for human expression, only human models are likely to be informative regarding any expression advantage it may possess. However, based on the available data, this study was continued with non-optimised version 'SP-B' only, to avoid duplication of work and following the view point that synthetic optimisation, especially of cDNA of human endogenous proteins, has the potential to cause unwanted, negative effects on expression, although remarkable increases in expression are reported in some studies ([345, 346], Section 1.3.1.2).

In biological replicates of the expression test, shifting band intensities between SP-B protein forms were observed (Figure 3.5). The effect was observed across plasmids tested (data not shown). One possible explanation for variable intensity in protein bands between independent experiments are differences in seeding density, which have been shown to impact SP-B expression in ATII cell-like human cancer lines [347]. For pAAV2.SP-B MG and pAAV2.soSP-B MG, only the mature SP-B and no intermediates were observed, likely because mature SP-B expression was strongest in this expression test, and other intermediates were below the detection limit. As A549 cells do not express SP-B or *SFTPB* mRNA without stimulation [338], low expression from the minigene constructs was anticipated.

A ca. 10-fold difference in rAAV vector production yields was observed depending on the promoter in the transgene cassette (Figure 3.6), such that transgene expression in the producer cell line seemed to negatively affect production yield, possibly indicating toxicity of the transgene product. If CMV was to be selected as the best performing promoter, efficient vector manufacturing might require repression of transgene expression by translational block [348].

3.3.3 Identification of a promoter with high expression levels in murine ATII cells after rAAV transduction using a newly established *in situ* hybridisation method

RNAscope was chosen as a quantitative method with additional localisation information to compare promoter-dependent SP-B transgene mRNA expression in ATII cells. During development of the RNAscope as a method for rAAV transgene expression levels, DNase and RNase digests revealed staining of vector DNA after rAAV transduction, but not genomic DNA (Figure 3.8). Potentially, RNAscope can avoid staining genomic DNA because of the density of DNA in the chromosomes. The effect of DNase digestion on RNAscope staining of rAAV transgene expression was observed across several animal experiments and vectors (Figure 3.8 and Figure 3.9), confirming the development of a reproducible method for transgene mRNA expression after rAAV delivery.

During the quantification of SP-B transgene expression using the established RNAscope method with DNase pretreatment (Figure 3.11), the cut-off for 'transduced cells' was chosen to be ≥ 2 dots per cell, to avoid quantification of false positives, however, at 0-2 dots per image in negative staining and double-digest controls, a lower level of ≥ 1 would have been possible. Notably, one single dot in an ATII cell in DNase-digested slides was a very rare event and the more conservative approach was chosen. Although RNAscope is a quantitative method, the mRNA copy numbers determined should be seen as an underestimate due to the sectioning process curtailing cells and potentially overlapping and/or unfocused signal dots. Therefore, this method should be seen as minimum mRNA copies per cell which, however, provides a consistent quantification method to compare expression between groups. To allow for capturing of more cell planes (section depth in z-direction) in one image and higher throughput, a conventional fluorescence microscope was chosen over confocal microscopy. Confocal Z-stack imaging would have allowed for

imaging of all focal planes within a cell, but this is time-consuming and ultimately unnecessary to answer the key research question: which promoter resulted in the highest expression levels in ATII cells.

The most commonly used ATII cell marker is SP-C expression. However in this study, *Sftpb* mRNA was chosen which is expressed by ATII cells in the lung parenchyma and club cells in the airways. The *Sftpb* RNAscope signal together with morphology still allowed for identification of ATII cells and was chosen because the *Sftpb* probe, which does not cross react with human SP-B transgene expression, had the potential to quantify murine endogenous expression versus transgene expression and determine the percentage of endogenous *Sftpb* expression – a surrogate for 'disease correction' potential. While the WPRE signal was clear and punctate, *Sftpb* signal was 'blobby' which hampered quantification, although the *Sftpb* signal was intense and filled the entire cell indicating high levels of *Sftpb* mRNA expression. Quantification of WPRE mRNA copies per cell showed that the hCEFI promoter achieved the significantly highest expression levels in murine ATII cells compared to all other vectors ($p = 0.044$ versus rAAV5.SP-B MG and $p < 0.0001$ versus all other groups, Figure 3.11) and could be the best choice for therapeutic SP-B transgene delivery to murine ATII cells.

The *SFTPB* promoter was investigated as the endogenous promoter of SP-B with expression predominantly in ATII cells and low-level expression in club cells. Because of the extremely high hydrophobicity of SP-B, accumulation of denatured SP-B and resultant ER stress are theoretical concerns, however, no adverse effects of expression in cells other than ATII cells have been described. WPRE mRNA expression was observed rarely in airways cells with any vector, which confirmed that rAAV5 transduction of airways is a rare event in our hands as mentioned above (Section 3.3.1). As serotype 5 already strongly limited transduction to ATII cells, this study was not suitable to investigate the potential of

the *SFTPB* promoter to restrict expression to target cells. Repetition of experiments with a vector with a less cell-specific tropism, such as rAAV6.2, would be more suitable to answer this question. In such experiments the investigation of the *SFTPC* promoter, which is known to restrict expression to ATII cells only, and showed high levels of reporter gene expression after plasmid delivery *in vivo*, would also be of interest [349].

CMV was included as a commonly used promoter expected to mediate some expression in ATII cells [321], however, it led to the lowest level of ATII cell expression (Figure 3.11). As it is additionally prone to promoter silencing (Section 1.3.1.1), it is not recommended for therapeutic SP-B expression in murine ATII cells.

DNase digestion strongly reduced rAAV vector RNAscope signal compared with sections not treated with DNase, as observed by the reduced transduction rate in ATII cells (Figure 3.12). The proportion of ATII cells staining positive for WPRE RNAscope signal determined in DNase-treated cells matched the transduction rate determined for rAAV5.CMV.EGFP (Figure 3.3), implying that signal in DNase-digested slides was a more accurate representation of transduced cells. This indicates that large proportions of the signal detected in untreated sections can be attributed to rAAV vector DNA. The differences in proportion of ATII cells staining positive for WPRE RNAscope signal might stem from the method of imaging and image analysis leading to a higher probability of detecting WPRE expression in cells with high expression levels, but there could be biological effects involved, too, such as differential promoter expression levels in ATII cell subsets [350]. Alternatively, the vector DNA RNAscope signal could stem from damaged vector DNA whose proportion is variable between vector preparations.

To investigate whether transgene mRNA expression corresponds to elevated levels of transgene SP-B protein in mice, the BALF of rAAV-dosed mice was analysed by Western blot. No statistically significant elevation of mature SP-B protein levels secreted to the

alveolar space was observed (Figure 3.13). Surfactant is a highly complex mixture of lipids and proteins, where the correct composition is critical for surfactant function and the secretion and recycling of surfactant are tightly regulated ([351], Section 1.2.2.2). It is possible that secretion of SP-B is regulated at the proteolytic processing stage and/or SP-B packaging in lamellar bodies is limited to a maximum level. Possibly for these reasons, despite high levels of transgene mRNA, increased secretion of SP-B into the alveolar space was not observed in wildtype mice. In line with this, only in the case of the highest-expressing hCEFI promoter was there a slight, non-statistically significant elevation of SP-B observed. Thus, although it is expected that the promotor driving the highest mRNA expression level should also result in the highest protein expression level, this could not be unambiguously demonstrated using Western blot analysis in wildtype mice. Similar studies in the murine model of SP-B deficiency, where no endogenous SP-B is expressed [320], were planned as a better model in which to investigate the effect of SP-B transgene protein expression and ultimately survival of the treated mice, but the model was not established in time for use in this thesis.

In summary, experiments in wildtype mice led to the selection of rAAV5.hCEFI.SP-B as candidate vector for SP-B gene therapy; and testing the benefit of this vector on survival in the murine model of SP-B deficiency is the immediate next step. In addition, as the ultimate aim is treatment of human patients, the candidate AAV serotype and transgene cassette need to be verified in human models of the lung.

Chapter 4 – Examination of rAAV gene therapy vectors in human lung models

Results in this chapter were partly published in Meyer-Berg et. al, Stem Cell Res Ther, 2020 as indicated in the figure legends [332].

4.1 Introduction

The choice of model for viral gene transfer is crucial for the investigation of rAAV tropism. The investigation of AAV serotypes in the mouse model after intranasal delivery allows for investigation of gene transfer *in vivo*, but viral entry receptor distribution is likely to be highly species-specific and variable AAV serotype tropism between species has been reported, as e.g. in human versus porcine airways [316]. One of the lessons learnt in the trials using rAAV2 to treat cystic fibrosis [177, 178], was that rAAV2 is safe, but generates insufficient CFTR expression in the airways for correction to meet the studies endpoint and create patient benefit, emphasising the necessity of rigorous testing of the delivery strategy ([352], see Section 1.4.3). Further, sialic acid distribution, which is an entry receptor for several AAV serotypes, has been shown to differ between human and mouse [353]. Sialic acid and its variable expression amongst species is particularly relevant here, as α-2,3-liked sialic acid is the primary and required entry receptor for AAV serotype 5 [278], the top performing serotype in murine ATII cells in Chapter 3. Therefore, AAV serotype tropism identified in the murine lung requires validation in a human lung model to select a rAAV vector candidate for gene therapy in the human lung parenchyma.

Human lung parenchyma (containing the target ATII cells) and airways are reported to have different expression of AAV entry receptors [354] and rAAV vectors are widely

reported to exhibit different airway versus parenchyma tropisms in animal models, including mice [283] and pigs [317]. An important requirement is that the epithelial cells of the human model are polarised. Models exhibiting polarised epithelium, with apical and basolateral surfaces demonstrate distinct rAAV vector tropisms as shown in human airway epithelial cultures grown in air-liquid interface to induce epithelial polarisation [248].

Ideally, the human lung model is of primary human cell origin. Many immortalised cell lines are cancer-derived which influences their transcriptome and proteome and often remains poorly characterised, possibly influencing expression of AAV entry receptors. Therefore, immortalised cell lines grown in 2D cell culture are unlikely to be a suitable model. Primary human cells are often used as a representative model, but in the case of human primary ATII cells, these de-differentiate into ATI-like cell types under standard culture conditions [355]. In addition, human primary ATII cells do not fulfil the requirement for polarisation to display distinct apical and basolateral surfaces.

4.1.1 Selection of a lung organoid model to mimic rAAV dosing to the human lung

Self-organising 3D cell culture models, also known as organoids, are typically composed of more than one cell type and can reflect properties or functions of a specific organ [356]. They are generated from human pluripotent stem cells (induced or embryonic) or from adult stem cells such as ATII cells [356, 357]. Organoids are reductionist because they lack vasculature, nervous or immune systems and are therefore not suitable to answer questions involving these systems.

Lung models grown from ATII cells can result in alveolospheres, which are 3D spheres containing ATII cells and sometimes ATI cells as well. Barkauskas *et al*. showed that murine ATII cells can give rise to ATI cells and act as stem cells in the adult lung [358]. In the study, they generated murine alveolospheres from isolated murine ATII cells by the

addition of a stromal cell population, that supported ATII cell differentiation. The alveolospheres obtained were small, ca. 4-5 μm in diameter, and it was unclear how the ATI and ATII cells were polarised. Barkauskas *et al*. tried to translate their finding to human ATII cells by isolating human lung cells by FACS and placing them in a supportive growth matrix with a foetal human lung fibroblast cell line. Emerging spheres formed colonies expressing ATII cell markers SP-C and HTII-280 [359] in a polarised fashion, however, colonies formed only at low efficiency and there was no evidence of ATI cells [358]. Furthermore, the generation of alveolospheres requires the regular supply of human lung tissue as an additional challenge.

Lung organoids generated from embryonic or induced pluripotent stem cell lines are more reproducible, tend to have higher efficiency for colony formation and do not depend on regular tissue donations. Gotoh *et al*. first generated ATII cells from human iPSC in co-culture with foetal human lung fibroblasts, to form alveolar spheroids with characteristics of the foetal lung [360]. Jacob *et al*. generated ATII cells with a more mature phenotype from human PSCs, which were able to form alveolospheres without mesenchymal cell support in 3D culture [324]. Alveolosphere cultures in 3D did not contain cell types other than ATII cells.

Lung organoids grown from human pluripotent stem cells (induced or embryonic) in a sequential differentiation protocol to mimic embryonic development tend to result in premature lung tissue, but with more cell type diversity. Dye *et al*. created lung organoids from human ESCs with 3D spatial organisation including both proximal- and distal-like lung epithelial cells [361]. The resultant lung organoids mainly contained ATII- and ATI-like alveolar progenitor cells expressing SOX9/SP-C or SOX9/HOPX [361, 362]. Even in prolonged culture, expression of SP-C was low and cells expressing the late ATII cell marker SP-B were rare. This indicates that the majority of the epithelial cells were

immature progenitors, and few were differentiated ATII cells. The ATII cells were further confirmed by the identification of lamellar bodies using electron microscopy [361]. The morphology of the LBs as published was not distinctly different in comparison with images of SP-B$^{(-/-)}$ LBs [363, 364], likely due to lack of maturity. In a follow-up publication by the same group, maturity of airway-like cell types was improved by the subcutaneous insertion of organoids in immunocompromised mice [365]. However, after engraftment the organoids lost their ATI and ATII cells. This generation protocol is therefore unsuitable to generate a model of the distal lung and the *in vivo* engraftment vastly increases the level of complexity for organoid generation.

The most advanced and mature lung organoid model generated from human stem cells was generated by Chen *et al*. [333]. These organoids termed lung bud organoids (LBOs) grown in 3D culture did not require engraftment *in vivo* for maturation and showed branching, morphogenesis in alveolus-like buds with more mature cell types and recapitulated lung development [333]. LBOs are generated by sequential differentiation of human ESC, mimicking human lung development. Lung development is a complex, multistage process requiring tight spatiotemporal control. Briefly, lung development starts with the development of the trachea from the foregut during the embryonic stage of lung development (Figure 4.1). During this stage the transcription factor NKX2.1 is expressed which is required for lung fate. NKX2.1 expression requires FGF2 and FGF10 secretion as well as WNT ligand signalling from the surrounding mesoderm in embryonic development. Retinoic acid signalling further supports NKX2.1 expression [366]. During lung branching and differentiation, the lung progenitor cells pattern in proximal and distal airway epithelium, with the distal epithelium eventually giving rise to the alveoli. During this process BMP4 and Wnt signalling are essential in maintaining the distal epithelium and stimulating branching. Alike, FGF7, FGF10 and retinoic acid signalling promote epithelial growth and branching [366]. Lung bud outgrowth is driven by localised expression of

FGF10 which requires inhibition of TGF-β signalling by retinoic acid signalling [367]. There is no one master regulator of lung development and the mentioned pathways work together in an intertwined fashion [367].

The matured LBOs comprise distal and proximal lung cell types, with a bias to distal lung, potentially allowing for identification of rAAV cell type tropisms. The LBOs showed basolateral/luminal polarisation as shown by uniform luminal MUC1 expression. The ATII cell markers HTII-280, SP-B, SP-C and ABCA3 were all present in luminal cells indicating a ATII cell phenotype that was more mature compared with previous publications. Lamellar bodies were present with a more mature morphology, and also functional as indicated by uptake and secretion of fluorescent SP-B. For ATI cells, only the bipotent marker HOPX for an ATI/ATII progenitor cell was present with no mature ATI cell markers, such as AQP5 or PDPN, observed. This combination of predominant expression of ATII cell markers and the absence of mature ATI cells is indicative of the canalicular stage of lung development (up to 24th week; see Figure 4.1); and accordingly, RNAseq mapped the developmental stage to the second trimester, which the authors analysed after 170 days of culture. The authors also used the LBOs to model a genetic lung disease, Hermansky-Pudlak Syndrome, via gene editing using CRISPR/Cas9 showing the suitability of this model to recapitulate a disease phenotype [333]. Disadvantages of the LBO model include its prematurity, which is a confounding factor for interpretation of results for neonatal adult lung, as well as the high cost, time and skill needed to differentiate ESC into LBOs. In summary, the LBO model is a polarised primary cell model of the human lung containing the target ATII cells at a sufficiently mature stage to express all their common makers; LBOs were selected as a human model for rAAV lung gene therapy.

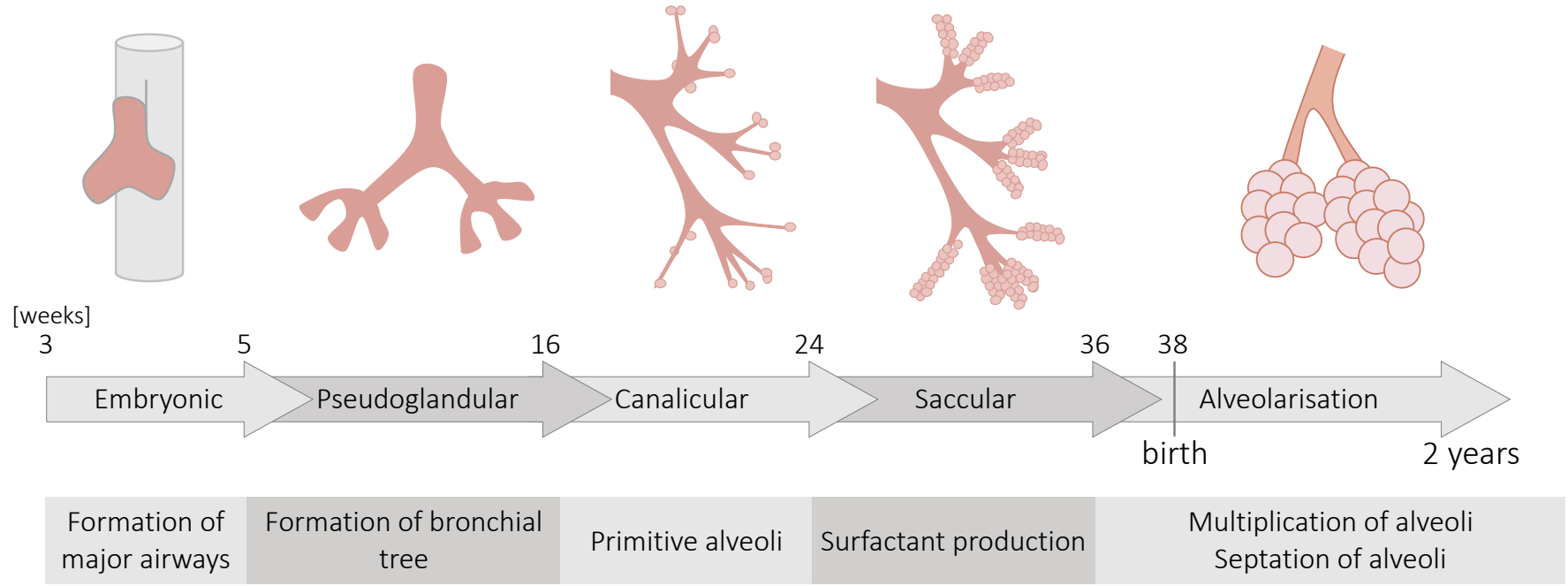


Figure 4.1 Stages of human lung development

Schematic to show stages of human lung development with duration in weeks. During the embryonic stage, the trachea forms from the endodermal tube, followed by formation of major bronchi. The bronchial tree forms during the pseudoglandular phase. The terminal and respiratory bronchioles emerge in the canalicular phase. Alveolar differentiation in ATII and ATI cells starts in the canalicular phase leading to appearance of primitive alveoli in the saccular phase, where fully functional lungs are formed. Air spaces expand, alveolar saccules form and surfactant production strongly increases during this phase. Alveoli continue to multiply, expand and mature during alveolarisation which lasts ca. until 2 years of age. Schematic based on [368-371].

4.1.2 Precision cut lung slices as a human lung *ex vivo* model to study viral infection

Precision cut lung slices (PCLS) are generated from donated resected adult human lung tissue. During PCLS generation, lung tissue from one donor is cut into multiple slices and subsequently cultured *ex vivo*, which allows for investigation of multiple conditions from a single donor [372]. PCLS retain many structural and cellular aspects of the lung in culture, making it an attractive model to study disease physiology, drug screening, infection and inflammation [372, 373]. Viability has been shown in PCLS for up to 14 days in human samples [374], and up to 6 months in animal samples, although viability declines after 4 weeks [331]. In contrast to LBOs, PCLS represent adult lung, but are often derived from aged and/or diseased patients, dependent on suitable lung donations, which introduces inter-donor variability. PCLS are generated from primary lung tissue such that all lung cell types are present, remaining in their native polarised tissue environment with the exception of the cut surface. This characteristic makes PCLS an interesting model to study viral infection, as was undertaken with e.g. bovine respiratory viruses, swine influenza A viruses and porcine respiratory coronavirus [375, 376]. PCLS have also been used to investigate gene therapy vector tropism such as that of oncolytic adenoviral vectors or pseudotyped rLV [331, 377]. In this chapter, human PCLS were used as a lung model to screen rAAV vectors.

The aim of the studies in this chapter is the screening of AAV serotypes in human lung models to assess if vector tropisms observed in the murine lung parenchyma extend to human lung parenchyma. Central to this question is if AAV serotype 5, a central component of rAAV5.hCEFI.SP-B gene therapy candidate vector identified in mice in Chapter 3, is suitable for transduction of human lung parenchyma. Little is known about AAV serotype tropism in the lung parenchyma, and in particular in human lung

parenchyma (Table 1.1). Therefore, a diverse panel of eight AAV serotypes was chosen for vector screening and two polarised, human lung models were employed; human LBOs to model the distal foetal lung and the human PCLS model representing adult human lung.

4.2 Results

4.2.1 Precision cut lung slices for rAAV vector screening

4.2.1.1 Use of murine PCLS as an *ex vivo* model for rAAV transduction

The AAV serotypes 5, 6.2 and 9 were investigated in the murine lung *in vivo* (Section 3.2.1). The AAV serotypes 5 and 9 were selected to determine if *ex vivo* transduction in murine PCLS was similar to transduction *in vivo*. Murine PCLS with ca. 3-4 mm diameter were prepared (Section 2.7.1) and transduced by topical administration of rAAV vectors carrying EGFP transgene (1E10 GC) at 10x less than the dose delivered intranasally *in vivo* (Section 2.7.2). Additionally, a high vector dose of 1E11 GC was used for rAAV5. On day 7 after transduction, PCLS were processed in fixed-frozen tissue sections (Section 2.7.2) and sections from the central region of the PCLS (to avoid the cut surface) were imaged for native fluorescence (Section 2.15). Cells were identified as transduced when punctate native EGFP fluorescence was observed together with a nucleus (Figure 4.2a-d). Transduction was observed with both AAV serotypes and vector doses. Transduced cells were found in the lung parenchyma, with the typical cuboidal morphology of ATII cells. PCLS sections contained very few airways, but when present no airway transduction was observed. Due to the overall low number of transduced cells (less than 15 transduced cells per section across rAAV vectors used, Figure 4.2e), the random selection of locations for imaging and quantification employed in *in vivo* experiments (Section 3.2.1) was not feasible. Instead, the tissue sections were systematically scanned and transduced cells were counted. Results were not significantly different between any of the experimental groups (Figure 4.2e). These results suggest that transduction of murine parenchyma *in vivo* could be replicated in murine PCLS *ex vivo*, but at much lower transduction rate, potentially due to poor PCLS tissue viability. Further, differences in transduction rate between rAAV5 and rAAV9 observed *in vivo* were not observed in PCLS *ex vivo* likely due

to the low numbers of transduced cells overall. This indicates that subtle differences in transduction rate may not be quantifiable in the PCLS model. As the 1E10 GC and 1E11 GC vector doses for rAAV5 resulted in similar transduction rates (mean number of transduced cells 2.3 versus 3.2 per section, respectively, Figure 4.2e), it is concluded that the PCLS is already saturated at the lower vector dose. Therefore, 1E10 GC was selected as the dose to take forward in PCLS to investigate vector transduction. However, the observed variability in viability in lung slices is a potential confounding factor that needs to be closely monitored in the following experiments.

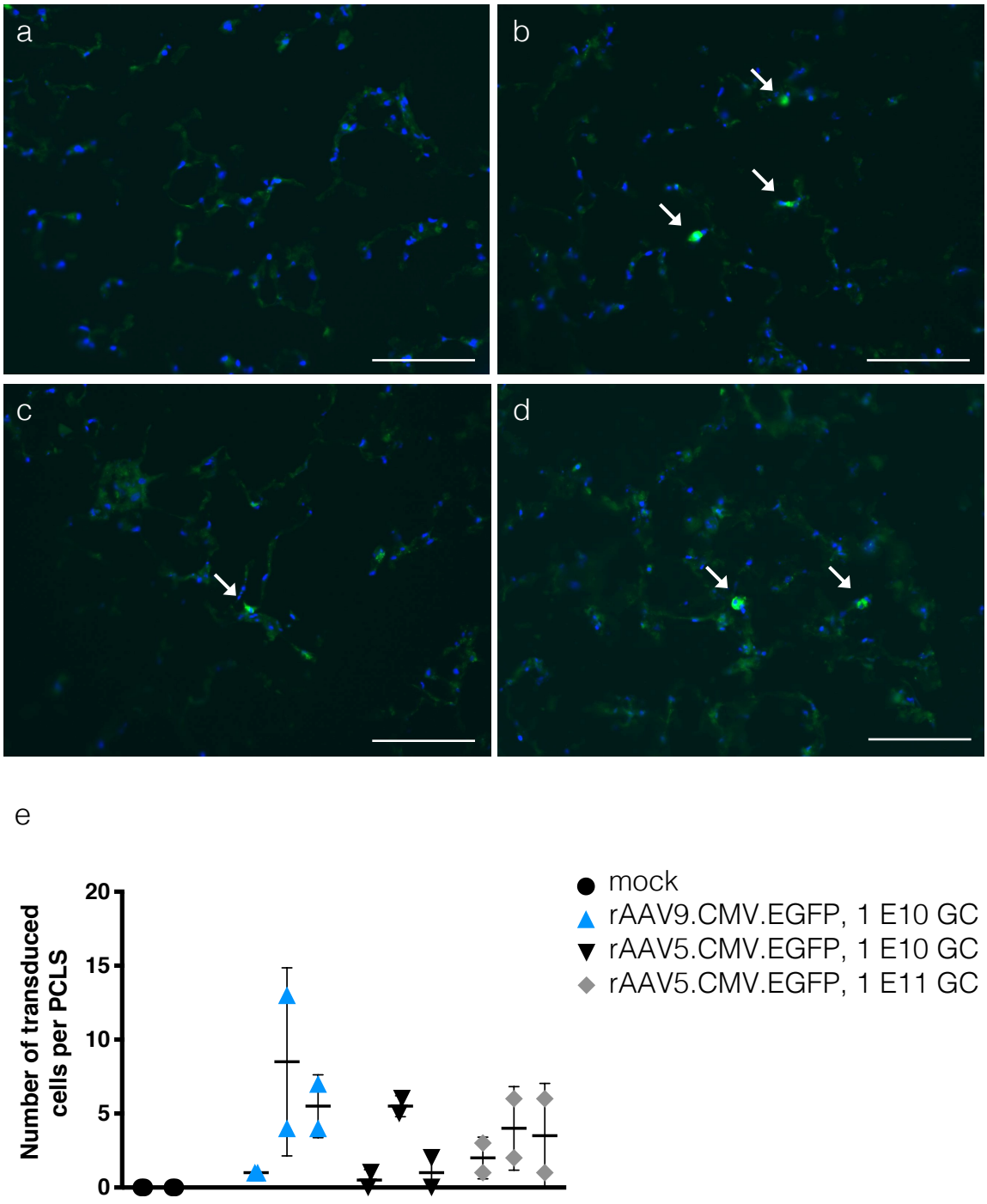


Figure 4.2 Transduction of murine PCLS with rAAV5 and rAAV9

Precision cut lung slices (PCLS) were generated from one BALB/c mouse, and 7 days later either mock transduced with vector diluent (PBS buffer, a) or transduced with 1E10 GC rAAV9.CMV.EGFP (b), 1E10 GC rAAV5.CMV.EGFP (c), 1E11 GC rAAV5.CMV.EGFP (d). On day 14 (day 7 after transduction), PCLS were processed in fixed-frozen sections, mounted with DAPI and imaged for EGFP native fluorescence (n= 2-3). Examples of cells counted as transduced are indicated with white arrows. Scale bar 100 µm. Mean number

of transduced cells per PCLS is shown (e) with each symbol representing the number of transduced cells counted in each of the two sections per PCLS scanned for EGFP fluorescence. Each mean represents one PCLS. Kruskal-Wallis test with Dunn's post-hoc test for multiple comparisons comparing means of each group was non-significant for all comparisons. PCLS were a kind gift of Dr Rosie Munday.

4.2.1.2 Screening of AAV serotypes in a human PCLS

The AAV serotypes 1, 2, 5, 6, 8, 9 and variants of 6, were screened in human PCLS using EGFP as reporter gene. PCLS of ca. 6-8 mm diameter were generated from resected human lung, dosed by topical administration 7 days later, and processed as fixed-frozen sections on day 8 after transduction (Section 2.7.2). To monitor tissue viability throughout the experiment (Figure 4.3), a live/dead tissue stain ('LD stain') was used, whereby green calcein stain becomes fluorescent by hydrolysis of intracellular esterases reporting live, metabolically active cells, and ethidium homodimer stains DNA if the cell membrane is compromised, as a measurement of dead cells (Section 2.7.1). PCLS viability was compared with a lysed PCLS serving as negative control for live stain and as positive control for dead stain (Figure 4.3a, e, i). On day 3 (Figure 4.3b-d) and day 7 (Figure 4.3f-h), PCLS showed high levels of live stain and low levels of dead stain, therefore, PCLS were used for transduction of day 7. On day 14 after generation (1 day before processing to fixed-frozen sections) PCLS viability was assessed again, with one of five PCLS analysed exhibiting LD staining similar to the staining observed on days 3 and 7 (Figure 4.3l). Three PCLS exhibited low levels for both live and dead stain (Figure 4.3j), with one PCLS almost completely negative for 'live' stain (Figure 4.3k), suggesting that metabolic activity was below the detection limit, but the cell membrane remained intact. Ultimately, it was concluded that viability was variable at the experimental endpoint.

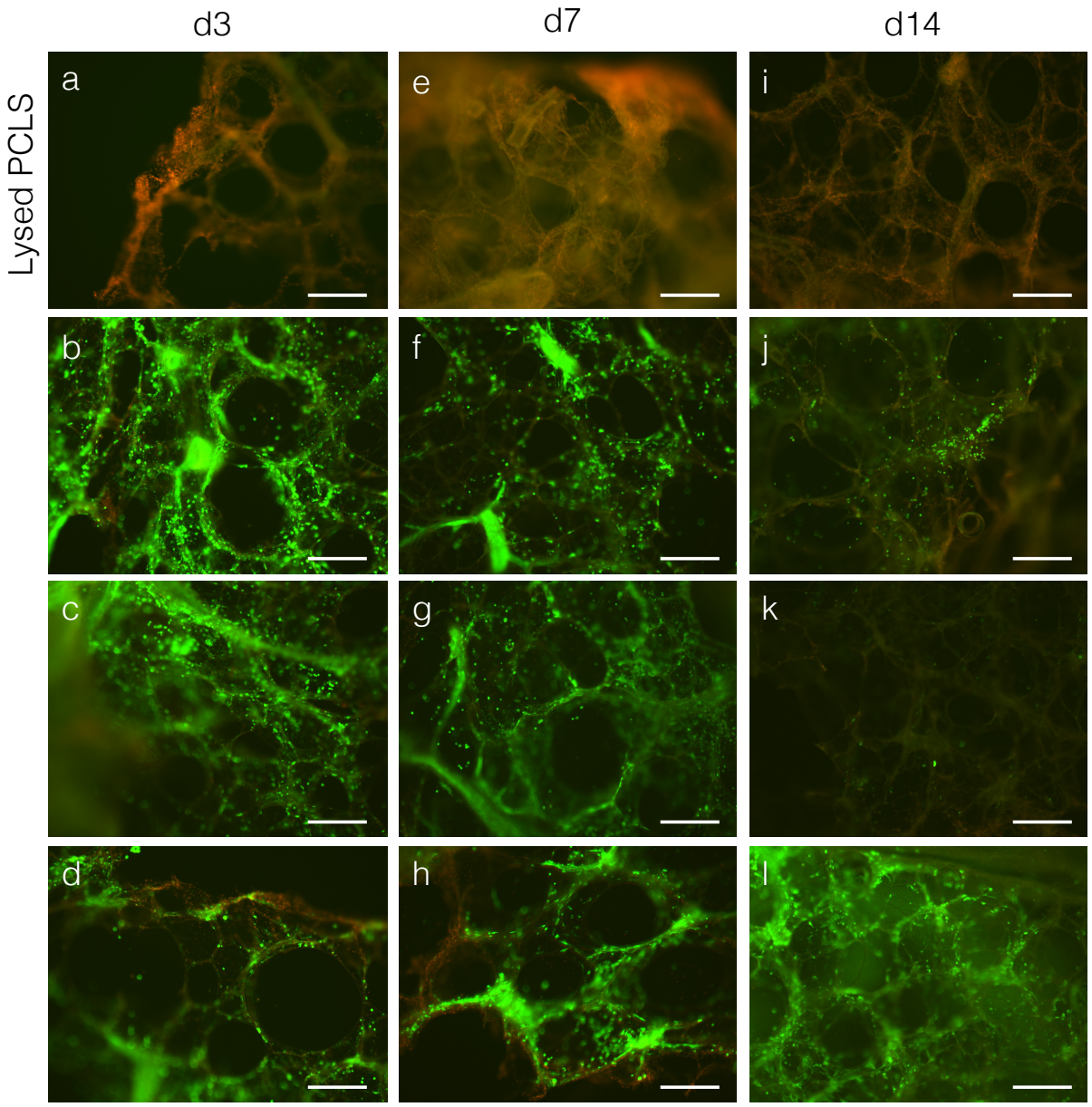


Figure 4.3 Viability of human PCLS from d3 to d14 after generation

Human precision cut lung slices (PCLS) were generated and 3, 7 and 14 days later were removed from culture and stained for viability using ethidium homodimer-1 to detect cells with compromised cell membrane (dead, red) and calcein for metabolically active cells (live, green). At each time point (d3, a-d; d5, e-h; d14, i-l), 3-5 PCLS were analysed with the live/dead stain. Lysed PCLS (a, e, i) served as negative control for the live stain and positive control for the dead stain. Human tissue was obtained under Oxford bio bank license OCHRE18A104, donor ID sORB 0112 183. Scale bar 500 µm.

On day 7, PCLS were either transduced with 4E10 GC of rAAV vectors or rSIV.F/HN vector expressing EGFP transgene. On day 11 (4 days after transduction), human PCLS were imaged *en face* for native EGFP fluorescence (examples shown in Figure 1.4). Punctate, high intensity EGFP fluorescence was observed in some images of PCLS transduced with rAAV6.CMV.EGFP vector, (e.g. Figure 4.4c, white arrows), which was clearly different from mock transduced PCLS (Figure 4.4a, b). Further, punctate, fluorescence was also observed (e.g. with rAAV6.2FF.CMV.EGFP, Figure 4.4d, white arrowheads) which was of a lower intensity than rAAV6.CMV.EGFP. It was concluded that this was suggestive of transduction and PCLS were processed in fixed-frozen tissue sections for further analysis.

Two non-consecutive sections from the central region of each PCLS were stained with DAPI and systematically scanned for native EGFP fluorescence. Structures that could be confused with transduced cells were identified in mock transduced sections (Figure 4.5a, b, white arrowheads). In contrast with these structures, punctate, high EGFP fluorescence with associated DAPI nucleus stain were identified in sections from PCLS dosed with the positive control vector rSIV.F/HN.CMV.EGFP (Figure 4.5c, d, white arrows). Structures identified with these characteristics were defined as transduced cells. For AAV serotypes 2, 6, 6.2, 6.2FF, 8 and 9 transduced cells were identified (Figure 4.5g-k, white arrows). The morphology of the (cuboidal-shaped) transduced cells in the lung parenchyma was indicative of ATII cells, although the morphology quality in the PCLS sections was a limitation.

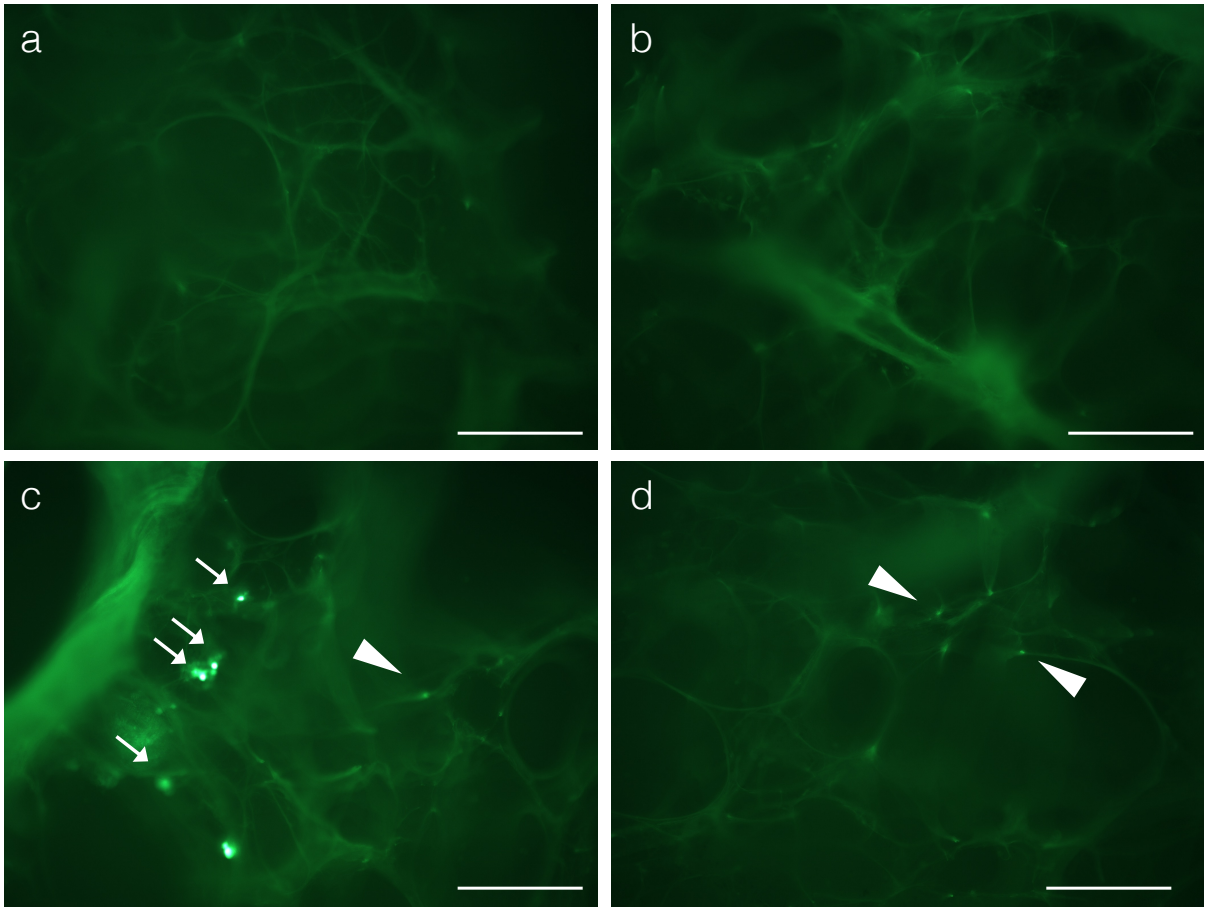
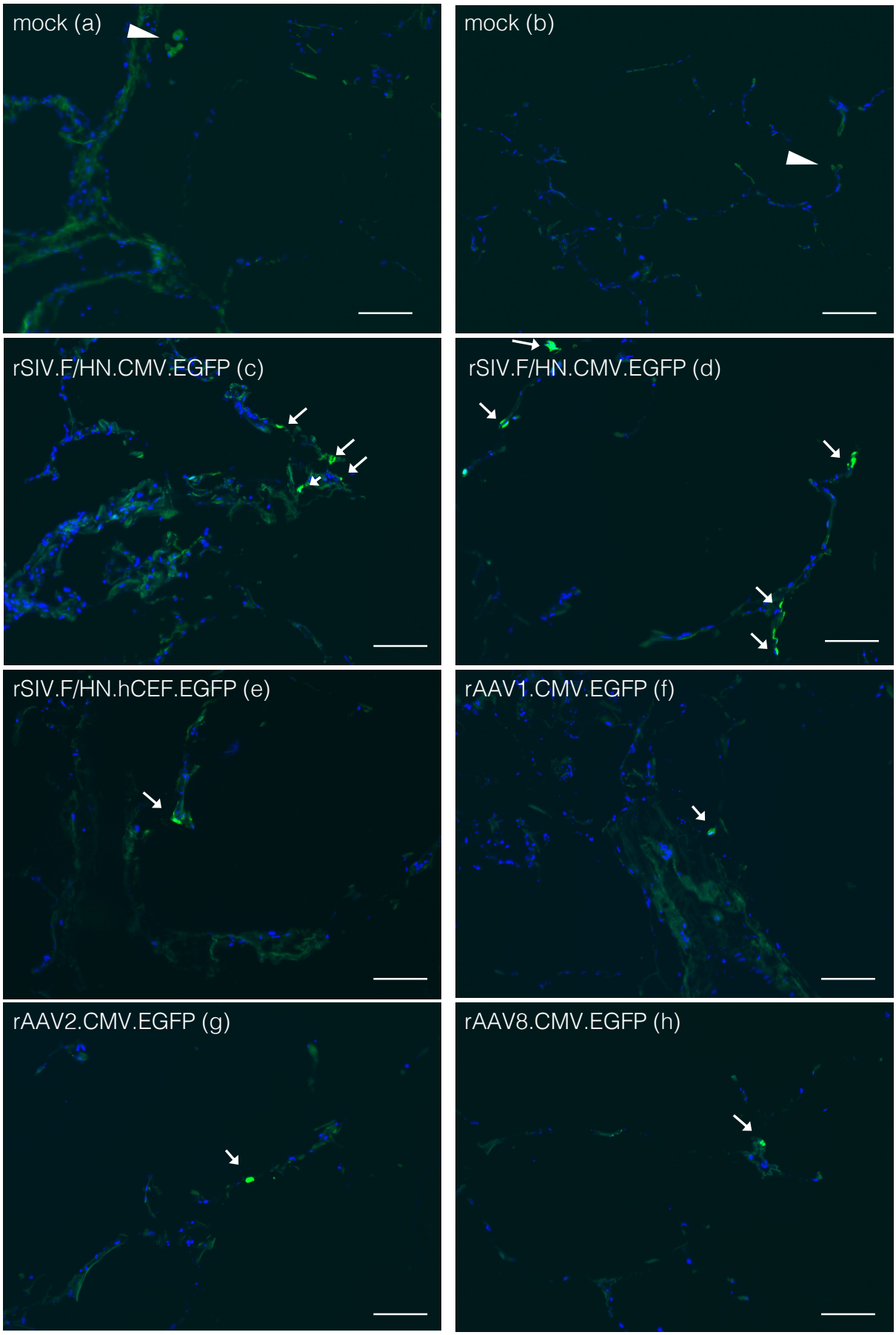


Figure 4.4 Images of human PCLS imaged *en face* for native EGFP fluorescence on day 4 after transduction with AAV serotypes

Human precision cut lung slices (PCLS) were either mock transduced or transduced with viral vectors on day 7 after generation and imaged *en face* on day 11 (day 4 after transduction) for EGFP fluorescence. Representative images of mock transduced PCLS (a+b). Images of PCLS transduced with 1E10 GC of rAAV6.CMV.EGFP (c) or rAAV6.2FF.CMV.EGFP (d) vector. Examples of transduced cells identified by native EGFP fluorescence are marked with white arrows. Examples of structures that with lower EGFP fluorescence, that are potentially transduced cells are marked with arrow heads. Human tissue was obtained under Oxford Radcliffe Biobank license OCHRE18A104, donor ID sORB0112183. Scale bar 200 µm.


mock (a)
mock (b)
rSIV.F/HN.CMV.EGFP (c)
rSIV.F/HN.CMV.EGFP (d)
rSIV.F/HN.hCEF.EGFP (e)
rAAV1.CMV.EGFP (f)
rAAV2.CMV.EGFP (g)
rAAV8.CMV.EGFP (h)

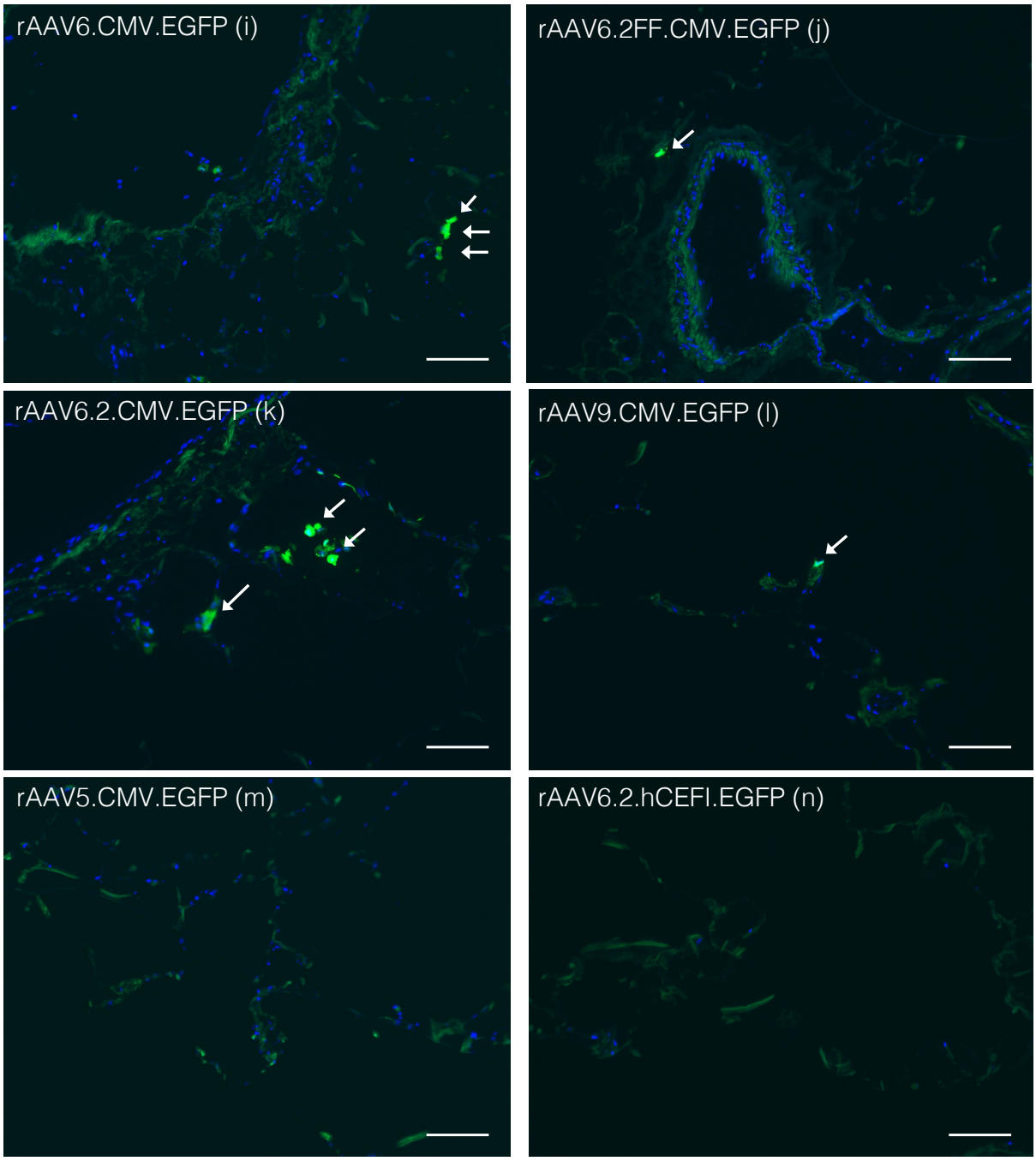


Figure 4.5 Images of human PCLS transduced with various AAV serotypes or mock transduced

On day 7 after generation, human precision cut lung slices (PCLS) were either mock transduced (a+b) or transduced with various viral vectors (c-n). On day 15 (day 8 after transduction), PCLS were processed in fixed-frozen sections and two sections per PCLS were systematically scanned for native EGFP fluorescence (n= 3-4). Merged images of GFP and DAPI channel are shown. Examples of cells possibly confused as transduced cells are indicated in representative images of mock transduced PCLS with white

arrowheads. Transduction with 3.2 E7 TTU rSIV.F/HN.CMV.EGFP (c+d) and 1 E8 TTU rSIV.F/HN.hCEF.EGFP (e) served as positive controls for transduction. Human PCLS were transduced with 4E10 GC of vectors rAAV1.CMV.EGFP (f), rAAV2.CMV.EGFP (g), rAAV8.CMV.EGFP (h), rAAV6.CMV.EGFP (i), rAAV6.2FF.CMV.EGFP (j), rAAV6.2.CMV.EGFP (k), rAAV9.CMV.EGFP (l) rAAV5.CMV.EGFP (m) and rAAV6.2.hCEFI.EGFP (n). Cells counted as transduced after imaging native EGFP fluorescence are marked with white arrows. For the vectors rAAV5.CMV.EGFP (m) and rAAV6.2.hCEFI.EGFP (n) no transduced cells were found. Human tissue was obtained under Oxford Radcliffe Biobank license OCHRE18A104, donor ID sORB0112183. Scale bar 100 µm.

For each AAV serotype tested, cells were counted in the two sections analysed per PCLS. Very few transduced cells were found overall and there was variability between PCLS transduced with the same vector. On average, 1-2 PCLS were completely negative per condition. Results are summarised in Table 4.1. Only structures with readily identifiable nuclei were counted as cells. For rAAV6.2.CMV.EGFP and rSIV.F/HN.CMV.EGFP the greatest number of transduced cells (high EGFP fluorescence + nucleus) were identified, followed by rAAV6.CMV.EGFP and rAAV6.2FF.CMV.EGFP (+++ and ++ respectively, Table 4.1). For rAAV2 and 8, only 1-2 transduced cells were found (+, Table 4.1). For rAAV1 and rAAV9, only potentially transduced cells at lower EGFP fluorescence were identified (Figure 4.5f, l) which were counted as ambiguous for transduction (+/-, Table 4.1). No transduced cells were found with the rAAV5 vector (Figure 4.5m; -, Table 4.1). Using the hCEFI or hCEF (an intron free derivative of the hCEFI element) promoter in rAAV6.2.hCEFI.EGFP and rSIV.F/HN.hCEF.EGFP, (Figure 4.5e, n), rather than the CMV promoter, resulted in lower numbers of transduced cells for both vectors (+++ → - and +++ → + respectively, Table 4.1).

To verify that the transduced cells in the lung parenchyma are ATII cells, PCLS were stained for the ATII cell marker pro-SP-C (Figure 4.6.). Similar to the viability stain, staining for ATII cells was variable between individual PCLS and overall few pro-SP-C-positive cells were found. Because the pro-SP-C stain required antigen retrieval, imaging for colocalisation of native EGFP fluorescence was not feasible. Given the overall very low numbers of transduced and pro-SP-C-positive cells, as well as reduced detection due to staining for EGFP instead of detecting native fluorescence, it was concluded that no convincing case for ATII cell identity of transduced cells could be made by immunohistochemistry.

Amongst the rAAV vectors tested, rAAV6.2 followed by rAAV6 and rAAV6.2FF transduced the most cells in the human parenchyma indicating that they are suitable candidates for human lung gene therapy for the parenchyma. To minimise the possibility that gene transfer was PCLS tissue donor-specific, and in hope of higher overall transduction levels and viability in another tissue donation, transduction of the human parenchyma repeated for serotypes 2, 5, 6, 6.2 and 6.2FF (n= 4-6) in PCLS. Viability declined over the course of this second experiment and was in general, as variable as in the previous donor. Transduction by imaging *en face* for native EGFP fluorescence (as shown in Figure 4.4 for the previous donor) was confirmed for serotypes 2, 6, 6.2 and 6.2FF. Similar to the PCLS from the previous donor, transduced cells were rare and transduction was not observed in every PCLS. Therefore, quantification of transduced cells or ATII cell stain was not reattempted. Serotype 5 (n= 4) was confirmed as a true negative for transduction of the human lung parenchyma in this second donor and subsequently a third donor by the systematic scanning PCLS sections for native EGFP fluorescence.

Importantly, AAV5, which was identified as a candidate serotype for gene transfer to ATII cells in mice, did not transduce human PCLS suggesting that it is not suitable for gene transfer to the human lung parenchyma. Other serotypes, such as rAAV6 and its engineered variants were found more suitable. However, because of overall low tissue viability and transduction level, as well as the limitations of this model described in the introduction (Section 4.1.2), these results prompted further investigation of rAAV tropism in human lung models.

Table 4.1 Transduction level of viral vectors in human PCLS

Viral vector	Cells identified in GFP channel/vector	Transduction level summary
rAAV1.CMV.EGFP	cells with low brightness	+/-
rAAV2.CMV.EGFP	1-2 cells with high brightness	+
rAAV5.CMV.EGFP	no transduced cells	-
rAAV6.CMV.EGFP	3-10 cells per vector with high brightness	++
rAAV6.2.CMV.EGFP	> 20 cells per vector with high brightness	+++
rAAV6.2.hCEFI.EGFP	no transduced cells	-
rAAV6.2FF.CMV.EGFP	3-10 cells per vector with high brightness	++
rAAV8.CMV.EGFP	1-2 cells/vector with high brightness	+
rAAV9.CMV.EGFP	cells with low brightness cells	+/-
rSIV.F/HN.CMV.EGFP	> 20 cells per vector with high brightness	+++
rSIV.F/HN.hCEF.EGFP	1-2 cells per vector with high brightness	+

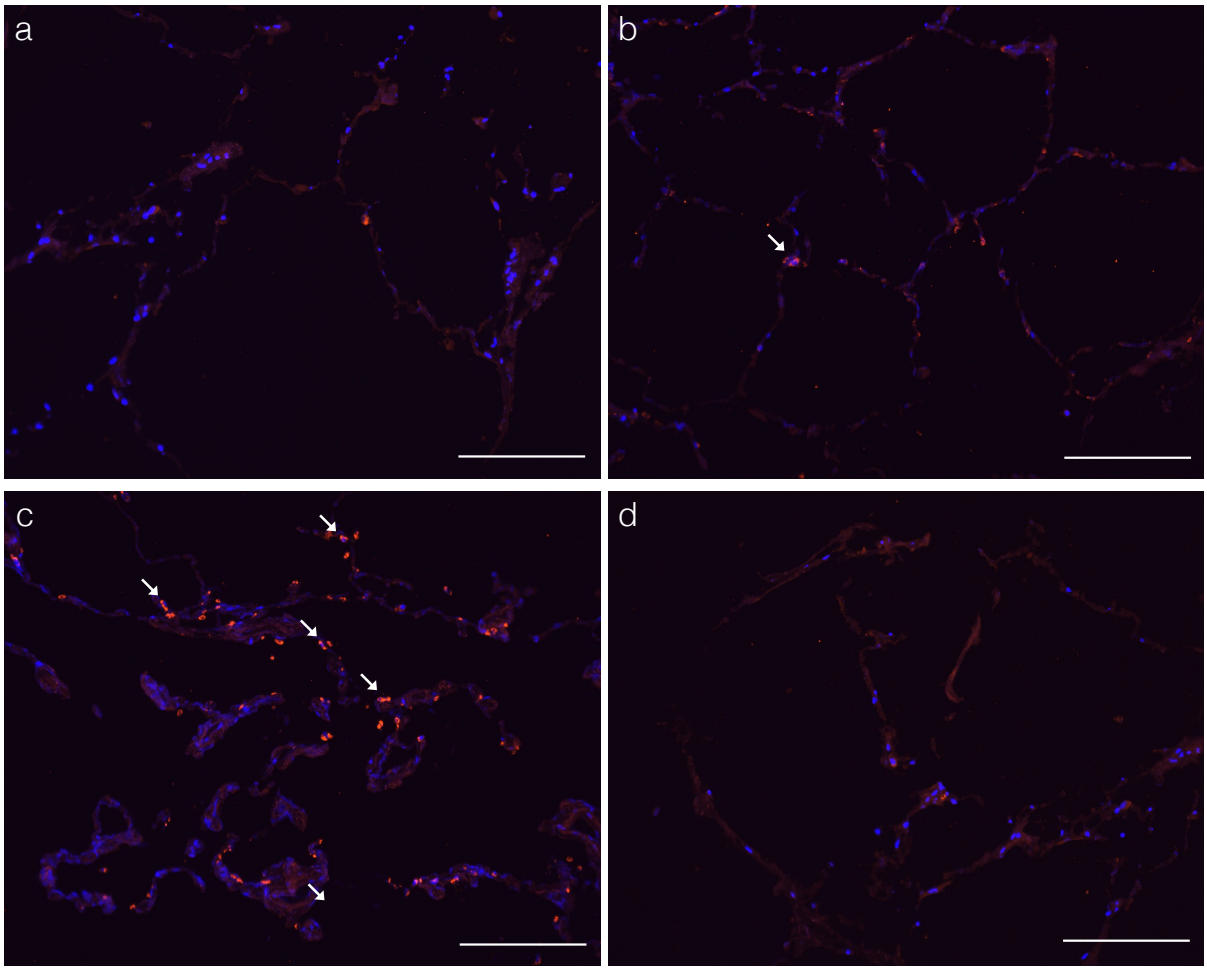


Figure 4.6 Representative images of sections from mock transduced human PCLS stained for the ATII cell marker pro-SP-C

On day 15 after generation (day 8 after mock transduction), mock-transduced precision cut lung slices (PCLS) were processed in fixed-frozen sections. After antigen retrieval, sections were stained for the ATII cell marker pro-SP-C and mounted with DAPI mountant. Representative merged fluorescence images of pro-SP-C (red) and nuclei (blue) are shown for PCLS mock_0002 (a), mock_0003 (b+c) and mock_0004 (d). Human tissue was obtained under Oxford Radcliffe Biobank license OCHRE18A104, donor ID sORB0112183. Scale bar 200 µm.

4.2.2 Human lung bud organoids as a model for rAAV vector screening

4.2.2.1 LBO model generation

Human LBOs were chosen as a reproducible primary human cell culture model with bias to distal (parenchymal) lung cell types, including the target ATII cells. Screening of AAV serotypes in this polarised 3D model allows for vector administration to the LBO lumen and apical cell surface, mimicking *in vivo* vector administration by inhalation. LBOs were generated by sequential differentiation as first described by Chen *et al*. [333], described in detail in [332], and outlined in Figure 4.7.

The ESC line AND-2 was obtained from the “Biobanco de células madre de Granada” at passages 29-35, karyotyped and confirmed as free of mycoplasma contamination. Human ESC were used in this study because of their more robust differentiation efficiency compared to iPSC lines as shown in direct comparison for similar lung organoid differentiation protocols [378].

Firstly, as part of endoderm induction in ESC, primitive streak formation was induced in embryonic bodies using WNT3a and BMP-4 on day 0 (Figure 4.7). The primitive streak formation in embryoid bodies gives rise to endoderm and mesoderm. Definitive endoderm was induced in embryoid bodies during day 1-4 using Activin A in presence of BMP-4 and FGF-2. High concentrations of Activin A acting on the TGF-β signalling pathway favour endoderm differentiation over mesoderm [379] and LBOs only contain minimal mesenchymal cell populations [333]. Endoderm cultures were directed to anterior foregut endoderm by sequential inhibition of first TGF-β signalling using SB-431542 and of BMP signalling using Noggin, followed by continued inhibition of TGF-β signalling together with inhibition of Wnt signalling using IWP2 [378].

Modelled on lung development *in vivo*, lung cell lineages were induced from anterior foregut endoderm cultures in suspension culture using CHIR99021, BMP4, FGF10,

KGF/FGF7 and retinoic acid [333]. Nascent organoids were placed in Matrigel on ca. d28 and treated with the same growth factors to induce maturation and branching (Figure 4.7). Cultures embedded in Matrigel for ca. 2 weeks, started to develop buds. The resulting LBOs can be mapped to the canalicular stage of lung development in the late second trimester according to Chen *et al.* (Figure 4.1) as characterised by the expression of the distal lung progenitor marker Sox9, formation of saccular structures reminiscent of the saccular phase expressing ATII cell markers, absence of ATI cells and by cross-referencing of LBO RNA sequencing data to second trimester foetal lung [333].

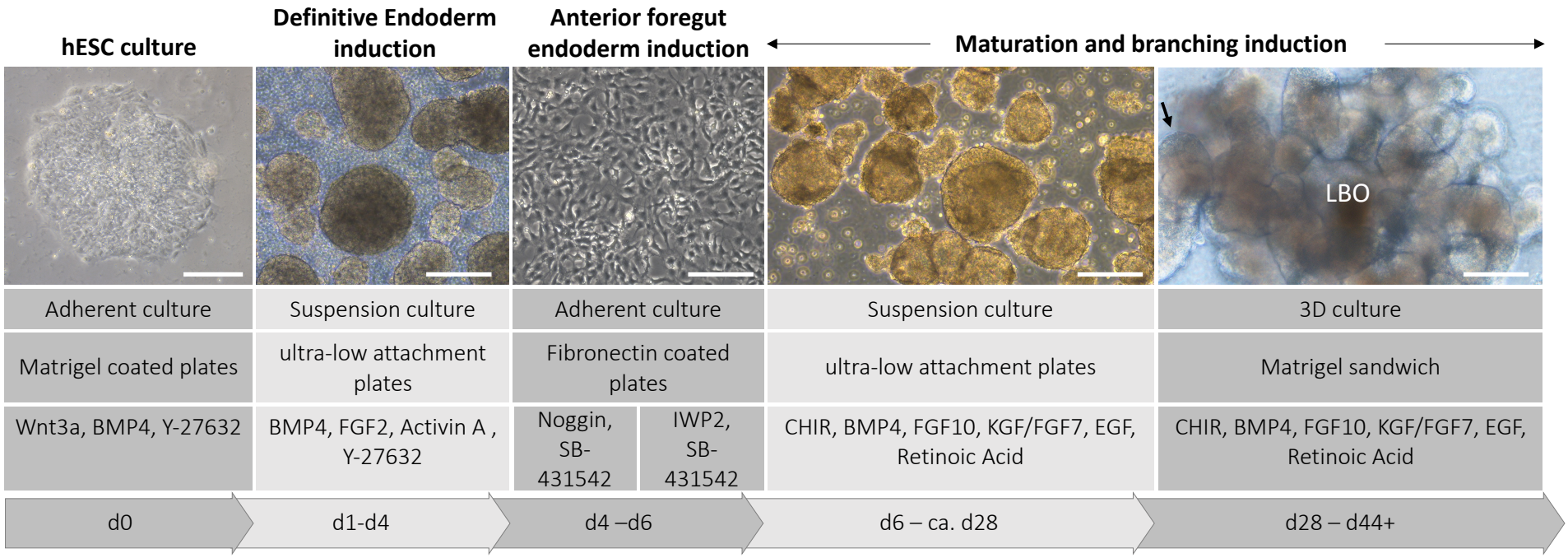


Figure 4.7 Key stages of LBO generation

Representative images of the key stages in lung bud organoid (LBO) generation are shown, together with their respective differences in culture, plate type/coating and key cellular factors, respectively. d0: Human embryonic stem cells (hESC) are routinely cultured in adherent mode and feeder cells (mouse embryonic fibroblasts) depleted on Matrigel coated plates 1 day before protocol start. d0-d4: Transition to suspension culture allowed for differentiation of embryoid bodies to definite endoderm. d4-d6: Transition to adherent culture on fibronectin coated plates and a two-step treatment with cellular factors induced anterior foregut endoderm. d6-d28: Transition to suspension culture allowed generation of nascent organoids. d28-d45+: mature LBOs are generated after nascent organoids are placed in a Matrigel sandwich,

with ‘buds’ starting to emerge after a few days as highlighted by the arrow. Mature LBOs are typically used for experiments from d45 onwards. Scale bars 200 µm.

Figure and figure legend modified from Meyer-Berg *et al.*, Stem Cell Res Ther, 2020. Collaborators at the ISCIII Madrid recorded images of steps 2-4 and performed model generation during d4 – ca. d40.

4.2.2.2 Establishment of a method suitable for dosing to the LBO lumen

To select an AAV serotype suitable for the transduction from the apical surface, the aim was to administer the vector to the LBO lumen. Because the average LBO size is 1-2 mm diameter, a microinjector delivery platform was chosen for vector administration. Firstly, a method of capillary pulling for microinjector needles suitable for injection to the LBO lumen was developed as microinjection to the LBO lumen had not yet been described in the literature. The key microinjector needle requirement was that it should provide sufficient penetration from the exterior of the LBO to reach the lumen with efficient delivery of relatively large (for a microinjector) volumes of up to 0.6 μL. Needles with a long continuous taper and tip diameter of ca. 20 μm were identified as fulfilling these requirements (for details see Section 2.8.2). LBOs were injected at 2-3 injection sites to allow for vector exposure throughout the organoid, especially in case the lumen was discontinuous. Microinjections were counted as successful when the injection of liquid caused the organoid to visibly pulsate. This microinjection method allowed for delivery of rAAV vector to the LBO and successful transduction with EGFP vector detected by native fluorescence (data not shown, microinjection development aided by Dr Maria Pilar de Lucas and Dr Alberto Zambrano) and was consequently used for rAAV vector administration and serotype screening. To improve comparability, vector titres were largely determined side-by-side.

4.2.2.3 Identification of AAV serotypes suitable for transduction of LBOs

LBOs aged day 59 or 79 were microinjected with AAV serotypes 1, 2, 5, 6, 6.2, 6.2FF, 8 and 9 expressing EGFP, or mock-injected with PBS buffer (Section 2.8.2). Alternatively, LBOs were injected with rSIV.F/HN expressing EGFP. After injection, LBOs were imaged *en face* for native EGFP fluorescence. On day 5 after injection, transduction with rAAV2, rAAV6 and variants rAAV6.2 and rAAV6.2FF yielded in high levels of EGFP fluorescence compared to mock-injected organoids (Figure 4.8a, f-i). Native fluorescence was lower,

but distinguishable from mock injection, in cultures injected with rAAV1 and rAAV8 (Figure 4.8d, e). Injection with AAV serotypes 5 and 9 resulted in EGFP fluorescence levels indistinguishable from mock injection (Figure 4.8a-c). In cultures injected with rAAV6.2.CMV.EGFP, native fluorescence was considerably more distinct compared with rAAV6.2.hCEFI.EGFP, which resulted only in low levels of fluorescence (Figure 4.8h, j). Injection with rSIV.F/HN did not result in EGFP expression distinguishable from naive regardless of the promoter (Figure 4.8k, l).

To summarise fluorescence levels observed from rAAV vectors, images were analysed for EGFP brightness intensity (Section 2.15), which was determined in the LBO surface area of each image as sum of pixel brightness values (Figure 4.9). The determined brightness intensity was normalised to LBO surface area and additionally to background fluorescence in the mock-injected group for improved comparability across experiments. The analysis confirmed the visual assessment that expression was highest for rAAV2, rAAV6 and variants rAAV6.2 and 6.2FF which was significantly different from mock injection (rAAV2 $p< 0.05$; rAAV6, rAAV6.2 and rAAV6.2FF $p< 0.005$) , while rAAV5 and rAAV9 resulted in the lowest levels of brightness. Further, analysis of EGFP brightness intensity quantified the decrease of native EGFP fluorescence by using rAAV6.2.hCEFI.EGFP instead of rAAV6.2.CMV.EGFP as ca. 5-fold (Figure 4.9).

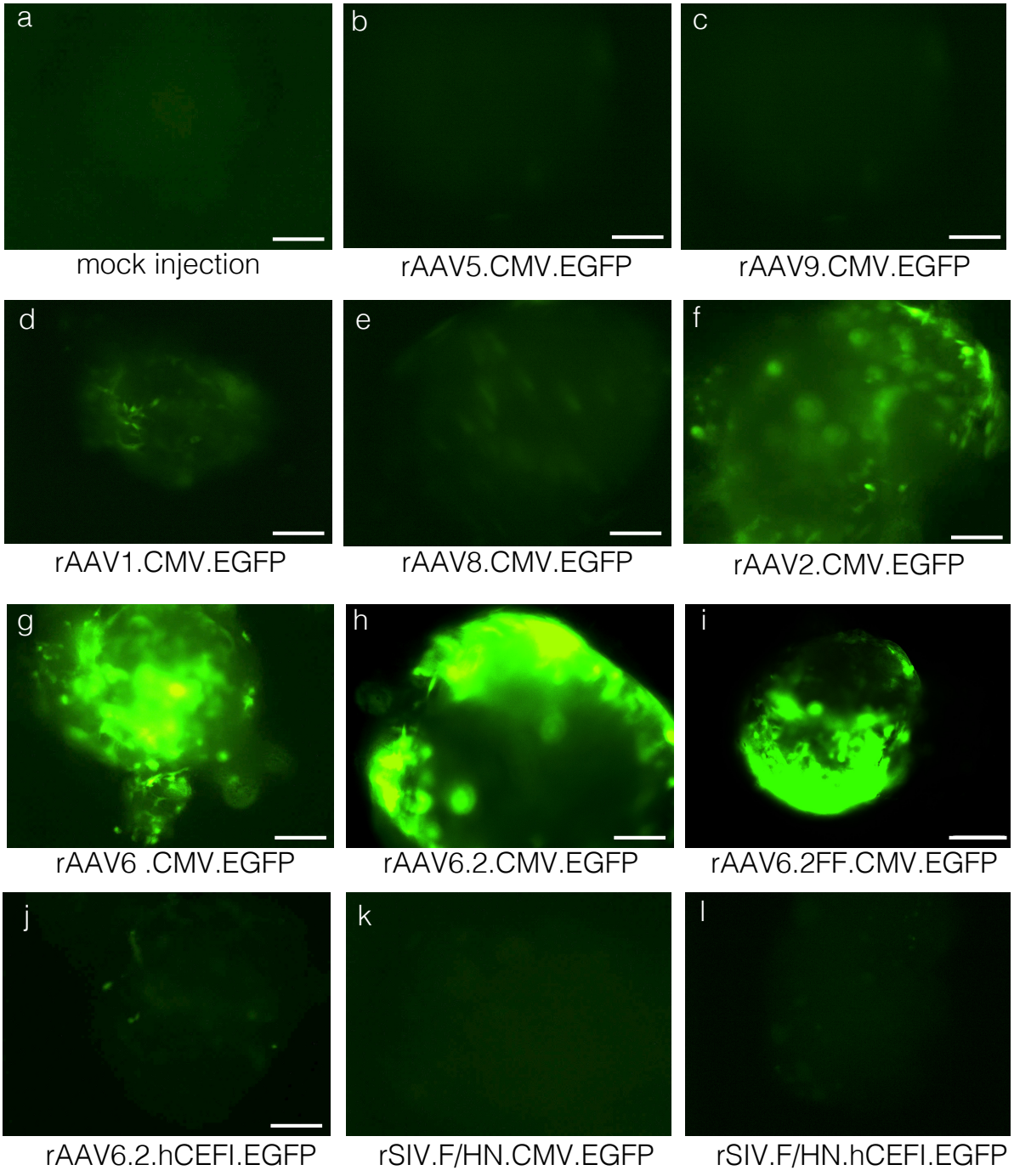


Figure 4.8 Native EGFP fluorescence in LBOs on day 5 after viral vector transduction

Lung bud organoids (LBOs) aged d59-79 were microinjected with 3.5E8 to 1E9 GC of rAAV vectors (b-j) or 2.6E5 to 3.8E5 TU of rSIV.F/HN (k, l) expressing EGFP from either the CMV promoter or hCEFI promoter as indicated. Mock injection with PBS buffer served as a negative control (a). Five days after injection, LBOs were imaged *en face* for EGFP fluorescence. One representative image from n= 3-4 injected organoids injected is shown; all LBOs displayed were injected d59. For vectors rAAV6.2, rAAV5, rAAV9 and

rLV vectors, the image is representative of two independent experiments, for rAAV6.2 and rAAV5 independent vector preparations were used between experiments. Scale bar 200 µm.

Figure and figure legend modified from Meyer-Berg *et al.*, Stem Cell Res Ther, 2020.

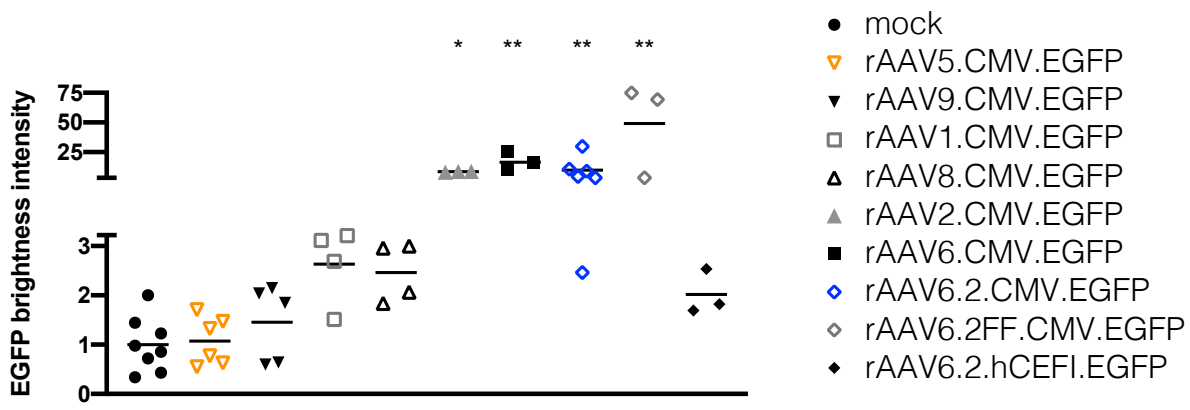


Figure 4.9 Quantification of EGFP brightness intensity in *en face* images of rAAV transduced or mock-injected LBOs

Lung bud organoids (LBOs, n= 3-4, aged d59-79) were microinjected with 3.5E8 to 1E9 GC of rAAV vectors or mock injected with PBS buffer. Five days after injection, LBOs were imaged *en face* for EGFP fluorescence. For mock injection and serotypes rAAV6.2, rAAV5, and rAAV9 LBOs (n= 3-4) from a second independent experiment were additionally analysed (LBOs aged d44-79). EGFP brightness intensity of LBOs in each image was determined as sum of the values of the pixels in the LBO area in each image and was normalised to LBO size and mean background fluorescence of the mock injected group. Each symbol represents one analysed LBO. Kruskal-Wallis test comparing mean rank of each group to the mock injected group. (*) $p < 0.05$; (**) $p < 0.005$.

Figure and figure legend modified from Meyer-Berg *et al.*, Stem Cell Res Ther, 2020.

4.2.2.4 Long-term rAAV expression in the LBO model

The LBO model was shown to have long-term potential for differentiation and proliferation up to day 170 [333], therefore, this model was used to study long-term CMV or hCEFI promoter directed EGFP expression after rAAV delivery. The hCEFI promoter was investigated as an alternative to the CMV promoter which is known to be prone to promoter silencing.

On day 59 of LBO differentiation, LBOs (n= 3) were microinjected with rAAV6.2.CMV.EGFP or rAAV6.2.hCEFI.EGFP vector. Mock-injected LBOs (n= 2) served as negative control. EGFP native fluorescence was observed as early as day 3 after injection with rAAV6.2.CMV.EGFP and was less robust in cultures injected with rAAV6.2.hCEFI.EGFP. After imaging on day 5, LBOs were imaged weekly. Native EGFP fluorescence was distinctly different in rAAV6.2.CMV.EGFP injected cultures compared with mock-injected cultures throughout the experiment until the experimental endpoint on day 63. Expression with the hCEFI promoter was also observed throughout the experiment, but at lower levels. After day 21, EGFP expression from both promoters was hardly detectable at 18.2 ms exposure. Therefore, exposure was increased to 100 ms for images shown for days 35 and 63 post-injection to remain within detection limits. In conclusion, both promoters lead to continued EGFP expression until the end of the experiment on day 63, but at decreasing levels. On day 63, EGFP expression from CMV and hCEFI vectors were more similar to each other than they were on day 5.

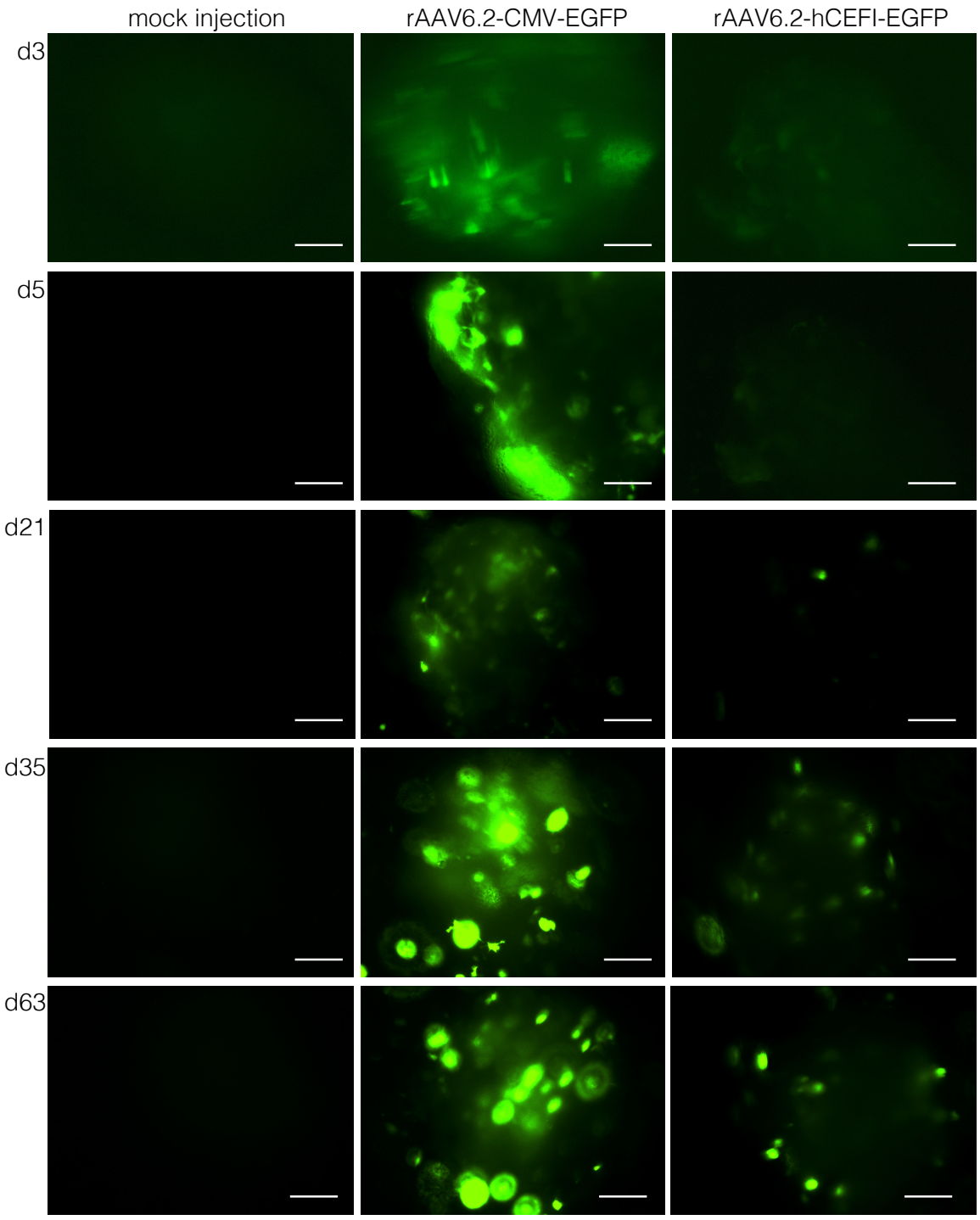


Figure 4.10 Long-term EGFP expression facilitated by CMV or hCEFI promoter after rAAV6.2 delivery to LBOs

Lung bud organoids (LBOs, aged d59) were microinjected with 3.5E8 to 1E9 GC of rAAV6.2.CMV.EGFP or rAAV6.2.hCEFI.EGFP as indicated or mock-injected with PBS buffer. Images were recorded *en face* with 18.2 ms exposure during d3-21 which was

increased to 100 ms for images taken d35-63 to remain within the detection range of the microscope. Scale bar 200 µm.

Images d21-63 were recorded by collaborators at ISCIII, Madrid.

4.2.2.5 Verification of the LBO's distal lung identity and characterisation of AAV entry receptors

LBOs were processed in fixed-frozen sections and stained for the ATII cell markers SP-C and SP-B (Section 2.10) to confirm distal lung identity (Figure 4.11). The SP-B and SP-C markers were both ubiquitously expressed in the honeycomb-like outer cell layers of LBOs, but not in the LBO centre, where the DAPI staining is indicative of the accumulation of cellular debris (Figure 4.11a, b). Colocalisation of native EGFP fluorescence with positive SP-B immunostaining was observed in LBOs transduced with serotypes 2, 6, 6.2 and 6.2FF as shown for serotype 6.2 (Figure 4.11c, d) indicating the suitability of these vectors to transduce the human lung parenchyma. The immunostaining for SP-C was more variable across LBOs, however, colocalisation for transduction with rAAV6.2 was observed (Figure 4.11e, f).

Recombinant AAV transduction depends both on primary glycan receptors and protein co-receptors (Table 1.1). To understand potential serotype entry receptors in LBOs, staining for known AAV primary glycan receptors and protein co-receptors was performed (Section 2.10). Immunostaining detected the characteristic subcellular staining of the universal AAV co-receptor (AAVR*,* Figure 4.12a, b). The lectin staining for α-2,3-linked sialic acid was negative using *Maackia amurensis* I (MAA I), but faintly positive for *Maackia amurensis* II (MAA II) with scattered cells staining positive in a more pronounced fashion (Figure 4.12c, d). While both MAA I and MAA II bind glycans with α-2,3-sialic acid linkage, MAA I favours binding glycans containing N-acetylglucosamine, while MAA II favours glycans containing N-acetylgalactosamine [380]. Lectin staining with *Sambucus nigra* agglutinin (SNA) was robust, indicative of pronounced α-2,6-linked sialic acid expression in LBOs (Figure 4.12g, h). Heparan sulphate immunostaining revealed positive staining in a 'spotted' staining pattern characteristic for this receptor ([381], Figure 4.12i, j).

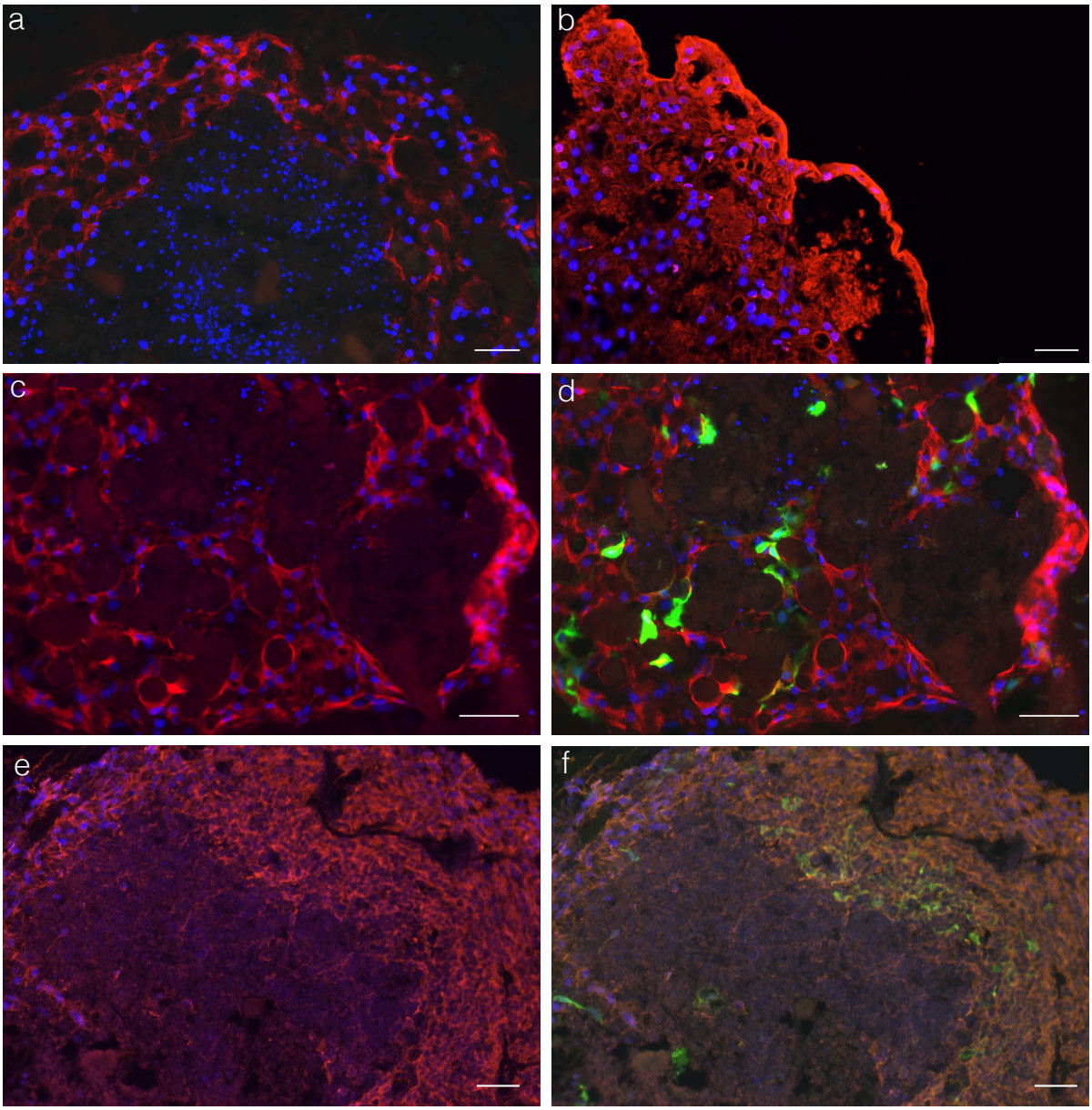


Figure 4.11 Immunohistochemistry for markers of alveolar type II cells in LBOs

Lung bud organoids (LBOs) aged d59-79 were mock injected (a, b) or microinjected with 3.5E8 to 1E9 GC of rAAV6.2.CMV.EGFP (c-f). On day 5 after injection, LBOs were processed in fixed-frozen sections and stained via immunocytochemistry. Representative images of LBO sections (n= 2-3) stained for ATII cell markers (red), including SP-B (a) and SP-C (b) with nuclei stained with DAPI (blue). Sections of LBOs microinjected with rAAV6.2.CMV.EGFP (c-f), stained for SP-B (c, d) and SP-C (e, f). Merged images of ATII cell marker staining (red) and DAPI (blue) are shown (c, e), as wells as merge of ATII cell markers (red), DAPI and EGFP (d, f). EGFP transgene expression was observed by native

fluorescence (d) or immunohistochemistry for EGFP after heat antigen retrieval (f). Scale bar 50 µm.

Figure and figure legend partially modified from Meyer-Berg *et al.*, Stem Cell Res Ther, 2020.

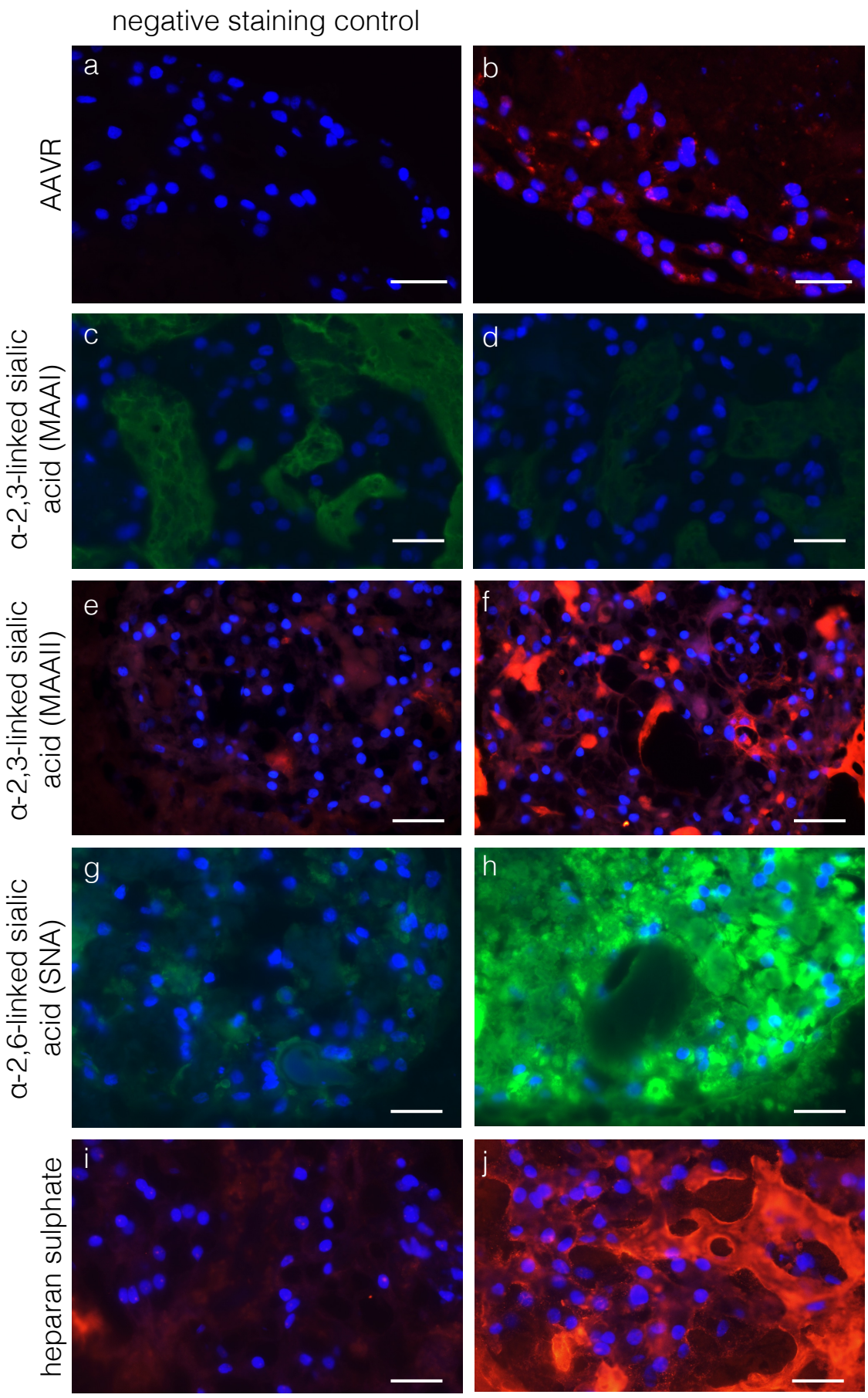

negative staining control
AAVR
α-2,3-linked sialic acid (MAAI)
α-2,3-linked sialic acid (MAAII)
α-2,6-linked sialic acid (SNA)
heparan sulphate
a
b
c
d
e
f
g
h
i
j

Figure 4.12 Immunohistochemistry and lectin stain for AAV entry receptors

Lung bud organoids (LBOs) aged d64-83 were processed in fixed-frozen sections and stained for various markers. Representative images of LBO sections (n= 2-3), with nuclei stained with DAPI (blue) and various markers (red or green) are shown, including: universal AAV co-receptor AAVR (a, b), glycan receptor α-2,3-linked sialic acid stained with MAA I (c, d), glycan receptor α-2,3-linked sialic acid stained with MAA II (e, f), glycan receptor α-2,6-linked sialic acid stained with SNA (g, h) and glycan receptor heparan sulphate (i, j). Staining with secondary fluorophore antibody only (a) and sections digested with sialidase A (c, e, g) and heparinase III (i) to remove glycans are included as negative staining controls. Scale bar 50 μm.

Figure and figure legend partially modified from Meyer-Berg *et al.*, Stem Cell Res Ther, 2020.

4.2.3 Identification of LBOs as a model of SARS-CoV-2 entry

As LBOs are a human model of the lung, we thought that they might be a good model for investigating viral entry into lung cells. The COVID-19 pandemic, which interrupted these studies, was caused by the worldwide spread of SARS-CoV-2 from the family of *Coronaviridae*, which uses the Angiotensin-converting enzyme 2 (ACE2) and transmembrane protease, serine 2 (TMPRSS2) as entry receptors in various tissues [382]. We were curious as to the utility of LBOs for SARS-CoV-2 research. To this end, LBO sections were stained for the SARS-CoV-2 entry receptors (Section 2.10). The presence of both human ACE2, for SARS-CoV-2 S-protein attachment, as well as TMPRSS2, involved in priming of the S-protein for entry, was confirmed by immunostaining (Figure 4.13). ACE2 staining was observed on the cell surface of scattered cells, whereas TMPRSS2 expression appeared more ubiquitous and also reminiscent of membrane localisation. Interestingly, LBOs were readily transduced with a SARS-CoV-2 pseudovirus (a lentivirus pseudotyped with SARS-CoV-2 spike protein) developed by co-workers (data not shown, studies performed in collaboration with colleagues: Dr Yue Du and Dr Kamran Miah).

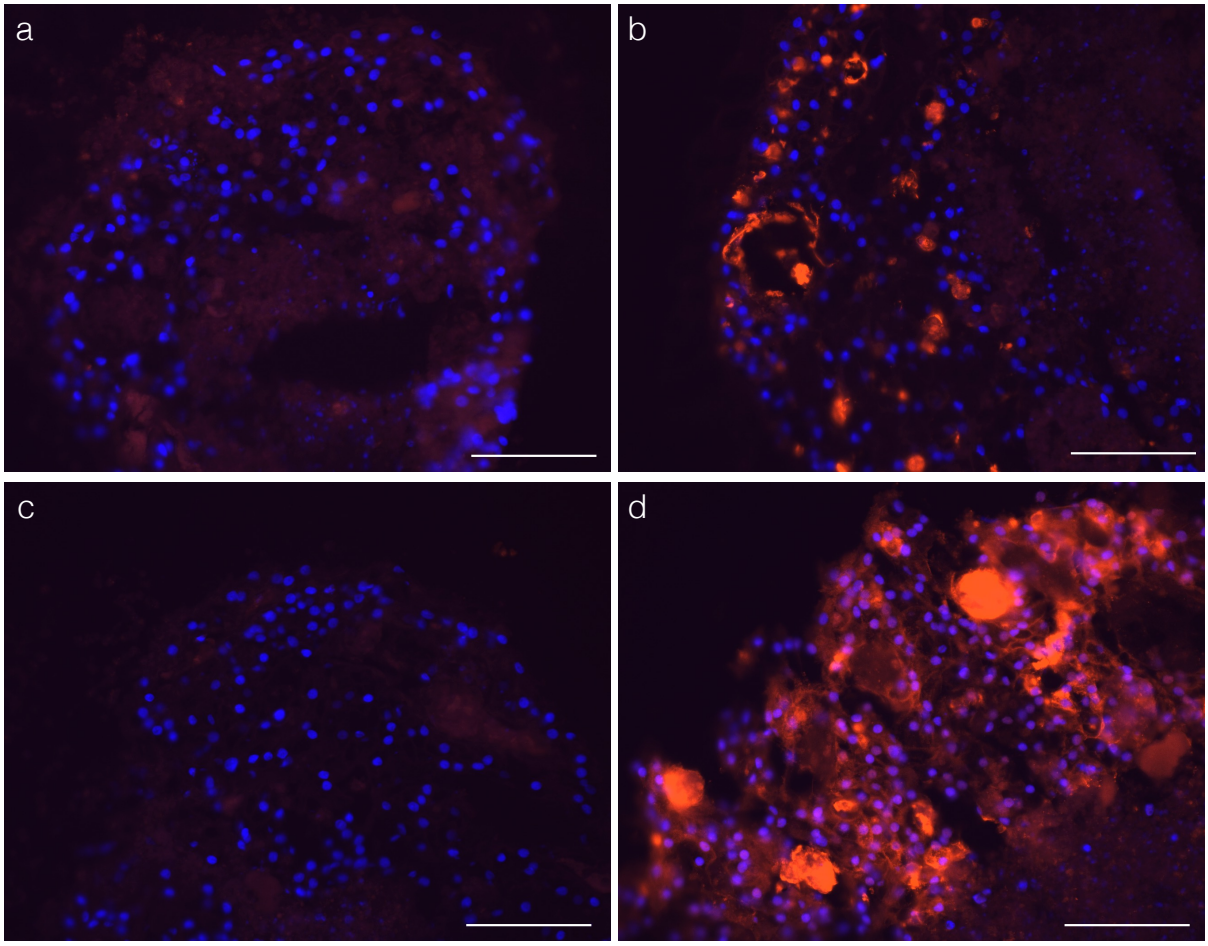


Figure 4.13 Immunohistochemistry for entry receptors of SARS-CoV-2

Lung bud organoids (LBOs) aged d64-83 were processed in fixed-frozen sections and stained for SARS-CoV-2 entry receptors. Representative images of LBO sections are shown, with nuclei stained with DAPI (blue) and entry receptors stained with immunohistochemistry (red), including: ACE2 (b), TMPRSS2 (d). No primary antibody control served as negative staining control for ACE2 an TMPRSS2 (a and c, respectively). Scale bar 100 µm. Representative of n= 2 stained organoids.

Figure as published under a CC-BY4.0 licence in the supplementary section of Meyer-Berg *et al.*, Stem Cell Res Ther, 2020.

4.3 Discussion

In this chapter, eight AAV serotypes were screened in the PCLS and LBO models of the human lung parenchyma. AAV serotype 5, which was found to be the best performing serotype in the murine lung parenchyma, did not transduce either of these two human models. Instead, serotype 6, its engineered variants and AAV2 transduced both human lung models in a robust manner (Table 4.2).

4.3.1 Vector screening in a model of the human adult lung

Mouse *in vivo* transduction was shown to be qualitatively replicated in murine PCLS *ex vivo* (Figure 4.2). Next, a panel of eight AAV serotypes was screened in the PCLS model, with transduction observed *en face* and PCLS sectioned to systematically search for transduced cells. Whereas *en face* imaging is quicker, sectioning and imaging is more labour intensive, but more sensitive because it reduces autofluorescence and signal-to-noise ratio by reducing section thickness from 500 μm to 7 μm. Amongst rAAV vectors, transduction was distinctly identified for serotypes 6.2, 6, 6.2FF, 2 and 8 (Table 4.2 and Figure 4.5). The vectors were listed here from highest to lowest number for transduced cells identified. However, murine PCLS showed that more subtle differences between vectors cannot be identified using in this model. The human PCLS viability was variable (Figure 4.3) and, with the low number of transduced cells observed overall, confirmation that rAAV6.2 is the quantitatively highest transducing vector identified in human PCLS remains speculative. Transduction of AAV serotypes 6, its engineered variants and AAV2 were repeated in a second donor and confirmed by imaging *en face*, together with a confirmed lack of transduction with rAAV5 in a second and third donor by scanning PCLS sections.

The low number of overall transduced cells in both donors could be due to an insufficiently optimised cell culture protocol for PCLS maintenance or poor tissue viability of the lung

tissue donation. A study assessing human PCLS viability using the LD stain amongst other viability read-outs and using an analogous media composition found that PCLS, after a recovery phase that lasted until ca. day 6, remained above 80 % viable with little variance until day 14 [374], correlating well with the time points for vector dosing and end point of the present study. Variable times for tissue viability of up to 3, and also up to 14 days, have been reported for human PCLS using similar media compositions [372]. It is conceivable, that the time between surgical lung resection and initiation of culture, as well as treatment in this time-frame (such as maintenance of continuous cold chain) could influence tissue viability and may have varied between reports thus negatively influencing viability in the present study. Another study investigating human PCLS viability found no significant difference in PCLS LD stain until the end of study on day 15, however, the number of pneumocytes as observed in haematoxylin/eosin stains decreased [383], providing another explanation for the low number of transduced cells as well as ATII cells observed overall (Table 4.1 and Figure 4.6). Assessment of clear exclusion criteria for tissue donations, such as the time since resection, the cold chain adherence, or optimisation of culture media, would likely be required to optimise use of the PCLS. In particular, to quantify vector tropism to human lung cells, the PCLS viability up to day 14 needs to be improved. A study in ovine PCLS included similar media composition, but additional supplements such as cAMP analogues to antagonise its effect on apoptosis and dexamethasone, showed higher levels of ‘live’ than ‘dead’ stain for up to 4 weeks [331]. Comparison of these supplements versus the media composition as used in the present study could be a promising starting point.

One of the advantages of PCLS is that they contain multiple cell types in a native tissue environment, however, in IHC stain for ATII cells in PCLS processed in fixed-frozen section on day 15 after generation, only few ATII cells with variability between sections were identified (Figure 4.6). The ATII cell stain required antigen retrieval, adding the

confounding factor of EGFP staining compared with already very few cells positive for native fluorescence. The occasional transduced ATII cell was observed, but no convincing case for ATII cell transduction could be made with such low numbers. Therefore, only the morphology of the transduced cells found in the parenchyma supported ATII cell identification. The scarcity of cells staining positive for the ATII cell marker pro-SP-C can be explained with tissue viability and reduced number of ATII cells on day 15, however, the possibility of ATII cell dedifferentiation after 7 days in culture needs to be considered. Overall, ATII cell dedifferentiation is less likely due to the 3D culture in a native tissue environment, the short culture time, and a lack of FBS in culture medium [355, 384], but this should be monitored when optimising PCLS culturing conditions for improved viability.

In summary, the PCLS model was generally unsuitable for quantitative analysis of vector transduction in target ATII cells, but did provide a human adult lung model with cells in a native cell environment to screen for vectors with tropism for human lung parenchyma. More importantly, the results also confirmed that AAV serotype 5 lacked efficacy in this model.

4.3.2 Human LBOs as a model for rAAV screening in context of entry receptor expression

LBOs were confirmed as a model of the human lung parenchyma by marker staining (Figure 4.11) and used to identify AAV serotypes 6, variants of 6 and 2 as vectors with tropism to the human lung parenchyma (Figure 4.9). Colocalisation with positive staining for ATII cell markers, suggested the suitability of these vectors to transduce this target cell type (Figure 4.11). Serotypes 5 and 9 did not transduce LBOs, discouraging their use as gene therapy vectors for the lung parenchyma (Figure 4.8).

A microinjection protocol was developed to model transduction from the luminal surface as during vector inhalation (section 4.2.2.2). The vast majority of vector was successfully

injected to the lumen; however, some leakage was observed, therefore a contribution to the observed transduction via basolateral transduction cannot be excluded. To estimate the effect of leakage, an approximately nine-times greater dose of rAAV6.2 was added to the culture medium (9E9 GC, n= 3, data not shown). No transduction was observed, however, indicating that basolateral transduction was not very prevalent.

In both the PCLS and LBO studies, EGFP was chosen as the reporter gene, primarily due to straightforward detection by non-destructive *en face* imaging. In addition, EGFP does not require indirect detection by antibodies, thereby eliminating potential false-positive IHC staining as a confounding factor during identification of transduced cells. Quantification of EGFP brightness summarised the transduction levels found in AAV serotypes (Figure 4.9). Although this assigned values to each image, EGFP brightness is not a quantitative method. Alternatively, the dissociation of LBOs with subsequent flow cytometry for EGFP positive cells or the use of the reporter gene luciferase would be more suitable in this regard to identify the AAV serotype with the quantitatively and significantly highest transduction levels in human models amongst the top performing AAV serotypes 2, 6, 6.2, and 6.2FF.

Amongst the limitations of organoids, their relative prematurity is the most likely to confound rAAV vector selection. Concerning known entry receptors, heparan sulphate expression is expected in lungs of healthy individuals [385, 386] and was also found in the LBO model. Sialic acid expression was investigated by Nicholls and colleagues in sections of foetal lungs from 20 weeks of gestation (Figure 4.1) by lectin staining [354]. The authors quantified staining intensities as (-) for MAA I, (+) for MAA II and (+) for SNA which was similar to lectin staining levels observed in LBOs in the present study (MAA I (-), MAA II (+), SNA (++), Figure 4.12). For paediatric and adult lung parenchyma, Nicholls *et al.* reported the same overall pattern as in foetal lung, but with increased intensity of MAA II

staining (MAA I (-), MAA II (++), SNA (+), [354]), indicating that there is little change in sialic acid primary entry receptor expression between foetal lung from 20 weeks of gestation to adult. In summary, prematurity is a limitation, but relatively well characterised limitation regarding sialic acid expression.

Another study investigating sialic acid expression of foetal lung by lectin staining found weak or negative staining by SNA in the canalicular stage of lung development, which increased in later lung developmental stages [387]. This staining profile of the canalicular stage of lung development, together with the staining profile identified by Nicholls and colleagues [354], would match the sialic acid expression profile observed in LBOs rather to post canalicular phases, such as the saccular phase, due to high levels of SNA staining observed (Figure 4.12). Because LBOs clearly represent foetal lung [333], they would be a useful model for *in utero* gene therapy, which has been previously investigated for SP-B deficiency using rAdV *in utero* in sheep [388].

Considering the increase in MAA II staining levels in later developmental stages of human lung compared to foetal lung (+ → ++) as described by Nicholls *et al*. [354], and only moderate levels of MAA II staining in LBOs (Figure 4.12), one could potentially underestimate transduction in LBOs with vectors that use α-2,3-linked sialic acid as an entry receptor, such as rAAV5 or rSIV.F/HN (Table 1.1 and [389]). If prematurity had a profound effect on selection of the vectors tested, then inconsistency between the models could be expected. This was only the case for the rSIV.F/HN vector which transduced PCLS at high levels but not LBOs. It is conceivable, therefore, that rSIV.F/HN transduction was negatively impacted by the prematurity of the LBO model and if α-2,3-linked sialic acid expression was rate limiting for rSIV.F/HN transduction in LBOs, then transduction by this vector should increase in more mature lung organoid models and in adult human lung.

More experiments are needed to understand the different outcomes for rSIV.F/HN transduction in human lung models.

Regarding characterisation of sialic acid receptors for rAAV entry (Table 1.1), the studies involved [256, 275, 278] did not differentiate between MAA I and MAA II, as it was common previously [380]. Therefore, it remains unclear if rAAV5 requires glycan receptors containing N-acetylglucosamine (bound by MAA I) rather than glycans containing N-acetylgalactosamine (bound by MAA II). If the former is the case, it is likely from MAA I staining in the present study (Figure 4.12) and studies by Nicholls *et al.* [354] that this receptor necessary for rAAV entry is not present in human lung parenchyma. If rAAV5 requires glycan receptors preferentially bound by MAA II, it was possible that rAAV5 transduction was underestimated in LBOs due to their prematurity, however, prematurity should not be the case in PCLS from adults (rAAV was negative in both models, Table 4.2), therefore indicating that rAAV5 was lacking a key entry factor other than α-2,3-linked sialic acid in this case.

Receptor staining for other entry receptors was also in line with transduction observed. AAVR is required for efficient transduction of all naturally occurring AAV serotypes tested [249], and was present in LBO sections (Figure 4.12). Robust heparan sulphate staining was observed in LBOs, which correlated with high transduction by serotypes binding heparan sulphate, namely AAV6 and 2. High levels of α-2,6-linked sialic acid were observed (Figure 4.12) which matched high levels of rAAV6 transduction. However, this is a vector with multiple primary glycan receptors (Table 1.1), and α-2,6-linked sialic acid expression is not necessary for rAAV6 transduction [238, 256]. The high levels of α-2,6-linked sialic acid glycan expression observed in LBOs (Figure 4.12) did not seem sufficient to facilitate high levels of rSIV.F/HN transduction.

In terms of duration of expression in LBOs, native EGFP fluorescence from rAAV6.2 using both the CMV.EGFP and hCEFI.EGFP transgene cassettes was observed for up to 63 days (Figure 4.10). Although CMV expression appeared more robust over the course of the experiment, the difference in EGFP expression compared with hCEFI seems to narrow as the study proceeded, which could be early signs of CMV promoter silencing. One of the central questions for rAAV gene therapy for SP-B deficiency is the question of how long will the episomal transgene expression last in a developing/growing lung with proliferating cell populations. In the context of the LBO model, additional characterisation of ATII cell proliferation is needed to investigate this further.

4.3.3 Complementarity of LBO and PCLS model

The experiments in this chapter identified AAV serotypes 6, variants of 6 and 2 as those with tropism for the human lung parenchyma and are therefore candidate vectors for gene therapy targeting the human lung parenchyma. Serotype 2 has previously been used in gene therapy clinical trials for cystic fibrosis, where it proved safe, but ultimately ineffective [179]. However, this is not contradictory to the results observed in the LBO and PCLS model, as airways contain different cell types, with viral entry receptors different from parenchymal ATII cells. Serotype 6.2 demonstrated one of the highest transduction rates in LBOs (Figure 4.9) as confirmed in multiple replicates and in PCLS, the most transduced cells were found for this vector (Table 4.1). Therefore, rAAV6.2 was selected as the lead vector for gene delivery to ATII cells in the human lung parenchyma going forward.

Both PCLS and LBOs were 3D primary models of the lung parenchyma, but differed in their origin (*ex vivo* culture versus *in vitro* generation), maturity (foetal versus adult lung tissue) and characterisation status (fully characterised human ESC line versus tissue from individuals affected by cancer). This strongly decreased the likelihood of the same

confounding factor in both models, building confidence in the rAAV tropisms observed to the lung parenchyma. Regarding gene therapy to the neonatal lung for SP-B deficiency, however, neither model represented the neonatal lung directly, thus other factors important to rAAV transduction remain as potential confounding factors. A highly desirable experiment to address this concern is the investigation of gene therapy vectors in PCLS generated from neonatal lung, although the availability of such tissue is very scarce.

With regards to selection of a suitable promoter, although expression from the CMV promoter was robust (Table 4.2), its use as a gene therapy promoter should be met with caution due to potential silencing *in vivo*. The hCEFI promoter could be suitable, especially as it demonstrated efficacy in the lungs of cystic fibrosis patients in clinical trials [174], but other promoters, such as the UBC promoter used successfully in *in vivo* studies to treat SP-B deficiency in the mouse model of SP-B deficiency could be an attractive choice [321]. In clinical trials, tissue specific promoters are frequently used currently, as well the constitutive CAG and CBA promoters ([103], Section 1.3.1.1). Tissue-specific promoters for the lung expression in ATII cells are the promoters in *SFTPB* and *SFTPC* [349, 390], driving SP-B and SP-C expression, which would be interesting to test in comparison before selecting a lead candidate vector in human. Further studies for promoter selection should ideally occur in a model with well characterised transcriptomics close to ATII cells, such as the ATII-like H441 cell line grown in air-liquid interface [344].

No model is perfect (or, all models are wrong, but some are useful [391]) and although some factors can be controlled regarding rAAV entry (Figure 4.12), there may also be unknown factors. Therefore, it is reassuring that rAAV tropism identified in two different human lung parenchyma models was well aligned. The experiments in this chapter suggest that the rAAV5.hCEFI.SP-B vector highlighted from *in vivo* mouse studies is most likely to be unsuitable for use in human, further suggesting that mice are a poorly suited

model for screening and selection of AAV serotypes for gene therapy in the lung parenchyma.

Table 4.2 Comparison of AAV serotype dependent transduction level in human PCLS and LBOs

Viral vector	Transduction level in PCLS	Transduction level in LBOs
rAAV1.CMV.EGFP	+/-	+/-
rAAV2.CMV.EGFP	+	++
rAAV5.CMV.EGFP	-	-
rAAV6.CMV.EGFP	++	+++
rAAV6.2.CMV.EGFP	+++	+++
rAAV6.2.hCEFI.EGFP	-	+/-
rAAV6.2FF.CMV.EGFP	++	+++
rAAV8.CMV.EGFP	+	+/-
rAAV9.CMV.EGFP	+/-	-

Chapter 5 – Gene therapy in a human disease model of SP-B deficiency

5.1 Introduction

In chapter 4, studies were performed using wildtype LBOs for rAAV vector screening. To create a human model of SP-B deficiency, in which to test SP-B gene therapy strategies, we wished to create SP-B deficient LBOs. Moreover, a human model of SP-B deficiency would also allow for screening of therapeutic gene editing.

To create SP-B deficient LBOs, the ESC line used for their generation requires a mutation clinically associated with SP-B deficiency. The RUES2 human ESC line was chosen for this purpose, as it was also used for the original LBO model creation [333]. In the same study, RUES2 cells were successfully gene edited to model the autosomal recessive genetic disease Hermansky-Pudlak syndrome which affects lysosome-related organelles and can involve pulmonary fibrosis.

Lung organoids were previously shown to accurately model many aspects of the pathophysiology of SP-B deficiency. Alveolospheres generated by Jacob *et al*. from iPSC from an individual homozygous for the 121ins2 mutation in *SFTPB* exhibited an absence of SP-B, the typical 'bubbly' LB morphology and an accumulation of misprocessed SP-C [324]. Similarly, in a study using iPSC generated from an individual homozygous for the 121ins2 mutation, lung organoids generated with a similar protocol as that used for LBO generation, resulted in strongly diminished *SFTPB* mRNA expression, the absence of SP-B protein, the typical LB morphology, and lower secretion levels of the phospholipid DPPC into the media [325]. Both studies showed phenotype correction either by gene

editing [324] or viral vector gene delivery [325] to the iPSC line, but not to the lung organoids or alveolospheres. These studies were limited to highlighting the modelling capabilities, rather than demonstrating correction of expression in these differentiated cultures, which would model gene therapy much more closely.

The aim of the studies in this chapter was to create a RUES2 human ESC line with mutations causing SP-B deficiency, which could be used to generate SP-B deficient LBOs as a human disease model. Such a model would be useful for screening gene addition or gene editing approaches, and could be used to further investigate rAAV6.2 as a candidate gene delivery vector.

5.2 Results

5.2.1 Establishment of a method for generation of a human ESC line edited in *SFTPB*

Several mutations in the *SFTPB* gene have been clinically described in association with the pathophysiology of SP-B deficiency. Here, two mutations were selected for creation of a disease model from gene edited human ESC cells. The 121ins2 mutation in exon 4 was selected because it is the most prevalent mutation in *SFTPB* accounting for ca. 66 % of all cases ([9, 73], Figure 5.1). Insertion of this mutation requires precise edits via the low activity HDR pathway and the likely success of this editing method in stem cells is dependent on an optimised delivery of gene editing tools and screening of many edited clones. In addition, therefore, an alternative, clinically-described mutation, involving a large deletion comprising two exons (exon 7 and 8) which requires less precise editing, was included with the expectation that this mutation would be easier to achieve via the more active NHEJ pathway ([78, 79], Figure 5.1).

After choosing editing strategies, sgRNAs were designed and tested in HEK293T cells (Figure 5.2). For the 121ins2 mutation, a sgRNA target site was already described which was previously used to correct this mutation to wildtype in iPSC derived from a patient homozygous for the 121ins2 mutation [324]. Here, the reported sgRNA was altered to target the wildtype allele (Figure 5.2B). Advantageously, the Cas9 cut site for this sgRNA was at the precise site of intended editing, disrupting the sgRNA target site and reducing the likelihood of rebinding after successful editing [138].

To create a large deletion comprising exons 7 and 8, two sgRNAs were designed and selected in both the intronic region upstream of exon 7 and downstream of exon 8 using CRISPOR online design tool (crispor.tefor.net, [392]). Single gRNAs were selected on the basis of predicted efficiency, cutting frequency determination (CFD) specificity score [393],

the number of predicted off-targets, their location (intron favoured over exon) and their number of mismatches required for binding, especially if mismatches were outside the 12 bp adjacent to the PAM. Selected sgRNAs for targeting exon 7/exon 8 and testing in HEK293T cells are shown in Figure 5.2C.

The efficiency of the five selected sgRNAs for *SFTPB* targeting was validated *in vitro* by transfection of sgRNA/Cas9 RNP complexes (Section 2.4.2). Cas9 protein without sgRNA was transfected as negative control and RNP containing a sgRNA targeting the *RELA* gene encoding a transcription factor provided by the manufacturer was included as positive control. On day 3 post transfection, genomic DNA was extracted, the region of interest amplified and sequenced by Sanger sequencing. The ICE v2 analysis tool (Section 2.4.3.2) was used to compare sequencing traces of edited and naive samples, shown here for sgRNA exon 4_121ins2 (Figure 5.2). The positive control sgRNA targeting *RELA* resulted in 54 % indel formation (Figure 5.2A) and the negative control resulted in 0 % editing (R^2= 1, data not shown). The indel formation rate was 77 % using the exon 4_121ins2 sgRNA (Figure 5.2B), therefore it was concluded that the sgRNA can mediate efficient levels of gene editing. All sgRNAs tested for the exon 7/exon 8 deletion resulted in indel formation (Figure 5.2C). The sgRNAs exon 7_126rev and exon 8_40fw resulted in the highest levels of indel formation and were therefore selected for the exon 7/exon 8 gene editing strategy in RUES2 cells.

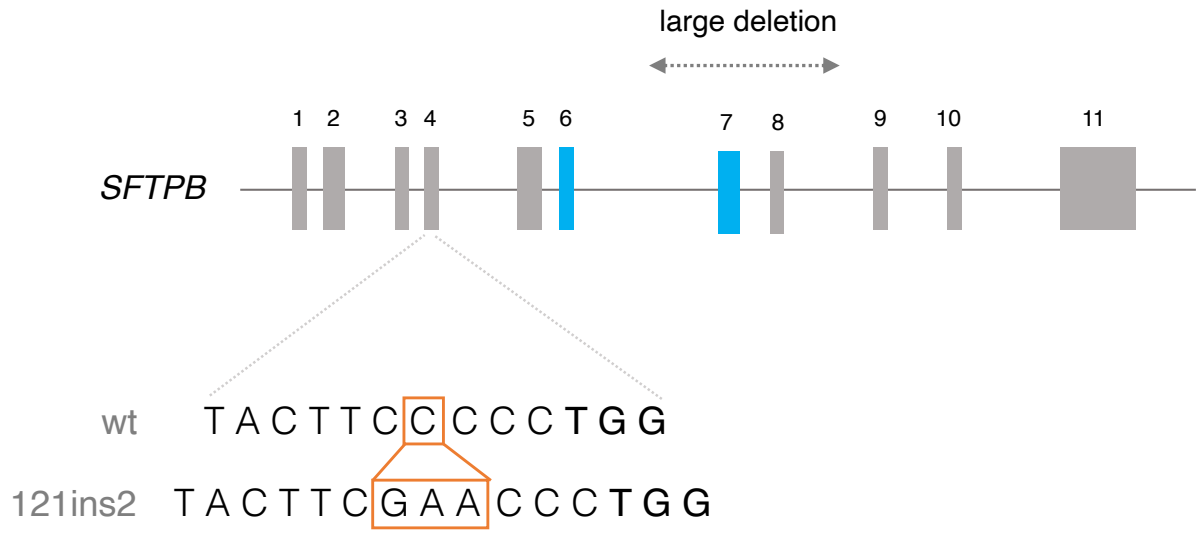


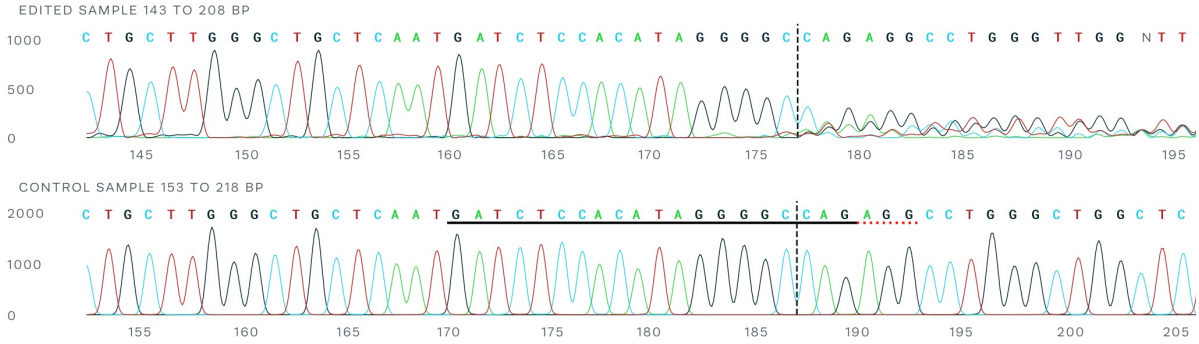
Figure 5.1 Schematic of editing strategies to create a human ESC line for generation of SP-B deficient LBOs

The SFTPB gene comprises 11 exons, the first 10 of which are transcribed, with exons 6 and 7 (blue) encoding the mature SP-B protein. Two clinically described mutations were selected for generation of gene edited human ESC lines: the 121ins2 mutation (C>GAA) in exon 4, which is the most prevalent mutation in SFTPB [9, 73], and a large deletion comprising exons 7 and 8 [78, 79].

A

Verification of editing protocol using RELA control sgRNA

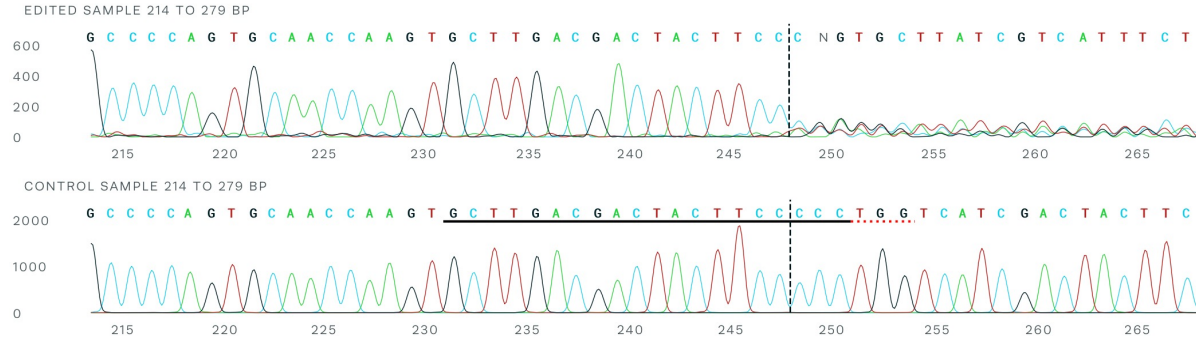
sgRNA	sgRNA target	PAM	Indel [%]	Model fit [R^2]
RELA control	GATCTCCACATAGGGGCCAG	AGG	54	0.96



B

Verification of a sgRNA for the insertion of 121ins2 mutation

sgRNA	sgRNA target	PAM	Indel [%]	Model fit [R^2]
exon4_121ins2	TTGACGACTACTTCCCCC	TGG	77	0.91



C

Selection of sgRNAs for generation of a large deletion in exon 7/exon8

sgRNA	sgRNA target	PAM	Indel [%]	Model fit [R^2]
exon 7_126rev	GGCAAGGCTATTCACAAATA	TGG	15	0.97
exon 7_41rev	GCCAGCTTCAGACCTACGTC	TGG	2	1
exon 8_40fw	TCATGGTCAGGGCCCGTTCA	AGG	68	0.94
exon 8_91fw	CAGCCAGAGACACCTAGGAT	GGG	59	0.92

Figure 5.2 Identification of sgRNAs for efficient *SFTPB* targeting in HEK293T cells

Cas9 protein was complexed with individual sgRNAs (1:1.3 ratio) and transfected in HEK293T cells. After ca. 72 h genomic DNA was extracted, and the region of interest analysed by Sanger sequencing and ICE v2 (Synthego) analysis. Results of ICE analysis including traces of edited and control sample with highlighted sgRNA target site (underlined) and PAM (red dashed line) are shown for targeting RELA control gene (A)

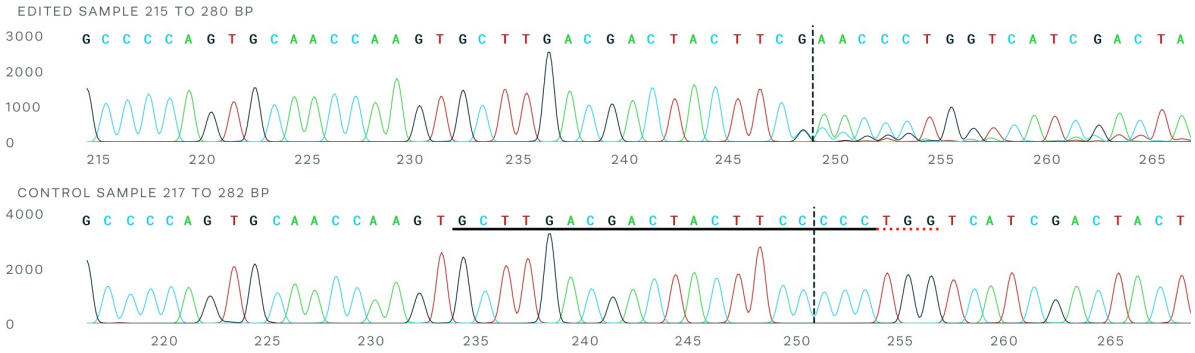
and exon 4 – 121ins2 (B). Results of ICE analysis for sgRNAs targeting upstream of exon 7 and downstream of exon 8 (C).

Advice on sgRNA design was obtained from Dr Philip Hublitz and experiments performed in collaboration with Dr Esme Magro-Lopez.

For delivery of RNP to the ESC line RUES2, cells were electroporated using the 4D nucleofector (Lonza). The electroporation of RUES2 cells was optimised in preliminary experiments using a plasmid expressing GFP (supplied with P3 Primary Cell 4D-Nucleofector™ X Kit S, Lonza), including method of dissociation for single cell suspension (timing, dissociation reagent), seeding density and survival (in collaboration with Dr Esme Magro-Lopez, data not shown, Section 2.4.3.1). Two 4D nucleofector electroporation programmes proprietary to Lonza were compared for RNP delivery: the CB150 programme which is generally recommended by the manufacturer for human ESC, and the CA137 programme which was found to lead to higher levels of indel formation than CB150 in the human ESC line H9 [394]. RUES2 cells were electroporated with RNP containing the exon 4_121ins2 sgRNA and matching ssDNA donor using either the CB150 or CA137 electroporation programme (Figure 5.3) and analysed for editing efficiency (Section 2.4.3.2). The CA137 programme led to 89 % indel formation according to ICE analysis (Figure 5.3A) which was higher than the 57 % found for the CB150 programme (Figure 5.3B). The ICE v2 algorithm also found higher levels of specific editing for the 121ins2 mutation with the CA137 programme (knock-in score 59 % versus 34 %). Encouragingly, the signal for the desired nucleic acids for the 121ins2 mutation were already clearly visible in the Sanger sequencing traces of RUES2 cells electroporated using the CA137 programme (Figure 5.3A); this programme also resulted in higher levels of indel formation in a second electroporation experiment (data not shown). Therefore, CA137 led consistently to higher levels of gene editing than the CB150 programme. In summary, the optimised method using the CA137 programme (described in 2.4.3.1) for RNP and donor delivery in RUES2 cells led to high levels of HDR gene editing for the 121ins2 mutation, at levels suitable to proceed to single cell cloning and clone screening to isolate a clonal 121ins2 edited RUES2 ESC line.

A

Electroporation programme CA137				
sgRNA target	PAM	Indel [%]	Model fit [R^2]	Knock-in score [%]
GCTTGACGACTACTTCCCCC	TGG	89	0.97	59



B

Electroporation programme CB150				
sgRNA target	PAM	Indel [%]	Model fit [R^2]	Knock-in score [%]
GCTTGACGACTACTTCCCCC	TGG	57	0.96	34

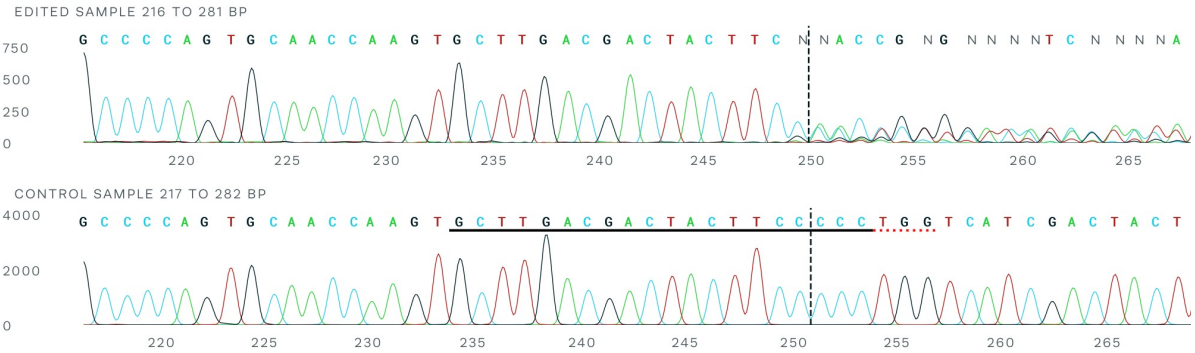


Figure 5.3 Comparison of 4D nucleofector electroporation programmes for efficient RNP delivery in RUES2 cells

RUES2 cells were electroporated with 300 pmol exon4_121ins sgRNA, 300 pmol donor DNA and 40 pmol Cas9 protein using the 4D nucleofector programme CA137 (A) or CB150 (B). At 24 h post electroporation, genomic DNA was extracted and editing in exon 4 evaluated by ICE v2 analysis. Results of ICE analysis including traces of edited and control sample with highlighted sgRNA target site (underlined) and PAM (red dashed line) are shown. Experiments performed in collaboration with Dr Esme Magro-Lopez.

5.2.2 Generation of a 121ins2 gene edited human ESC line

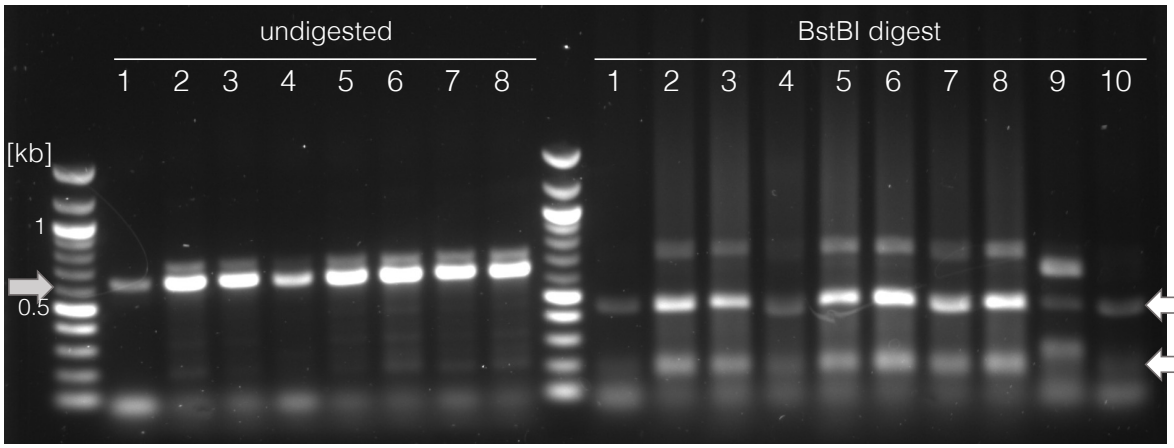
RUES2 cells were electroporated as described above using either the 121ins2 targeting strategy or using the exon 7/exon 8 large deletion strategy, and single cell clones were generated in two subsequent single cell cloning experiments by limiting dilution (Section 2.4.3.3). The first single cell cloning did not result in the isolation of homozygous 121ins2 edited clones, therefore a culture isolated in this first single cell cloning experiment was used to repeat the study (= second single cell cloning experiment). In the second single cell cloning experiment, 42 clones were screened for the 121ins2 editing by restriction fragment length polymorphism. Results for clones 1-10 are shown in Figure 5.4a. A 638 bp PCR product was anticipated which could be cut by restriction enzyme BstBI into a 457 bp and 181 bp fragment if editing were successful. A PCR product with an endogenous BstBI restriction site was included as positive control for digestion (data not shown). The PCR product for clones 1-8 included a dominant band of ca. 640 bp (in Figure 5.4a, grey arrow). For some clones an additional, fainter band was observed which appeared slightly higher (ca. 700 bp) and was observed only in clones with higher levels of genomic template DNA. This band was deemed non-specific. After BstBI digest, the 640 bp band was cleaved in all clones but clone 9. For clones 1-8 and 10, fragments at ca. 450 bp and 200 bp appeared, indicating successful 121ins2 editing. For clone 9, the banding pattern did not match the expected pattern for homozygous 121ins2 editing, in particular the retention of the 640 bp band was indicative of wildtype sequence. Of the 42 clones sequenced in this single cell cloning experiment, the 67 % were identified as homozygous for 121ins2 editing and for 26 % of clones the pattern matched neither wildtype nor 121ins2 editing which were likely either heterozygotes or not true single cell clones (Figure 5.4b).

Of the identified 121ins2 clones, 14 were expanded and confirmed by four sequencing reactions, twice after isolation (example for clone 10 shown, Figure 5.4c) and twice during

expansion before banking (example for clone 10 shown, Figure 5.4d). Clones 3 and 10 were selected for further characterisation. None of the wildtype clones survived expansion.

For creation of a large deletion in exons 7 and 8 in RUES2 cells, the ICE v2 algorithm predicted 41 % indel formation at the cut site upstream of exon 7 and 43 % at the cut site downstream of exon 8 after electroporation of RNPs with sgRNAs exon 7_126rev and exon 8_40fw. However, the ICE v2 model fit was poor for both (R^2 = 51 %, data not shown), because the algorithm compares the edited with the control trace and is likely not designed to take large deletions into account. During clone screening, PCR amplicons were directly analysed by gel electrophoresis, because the deletion comprised ca. 820 bp. Since the isolation of 121ins edited clones was achieved first, single clone screening for the large deletion in exon 7/exon 8 was ceased at this point.

a

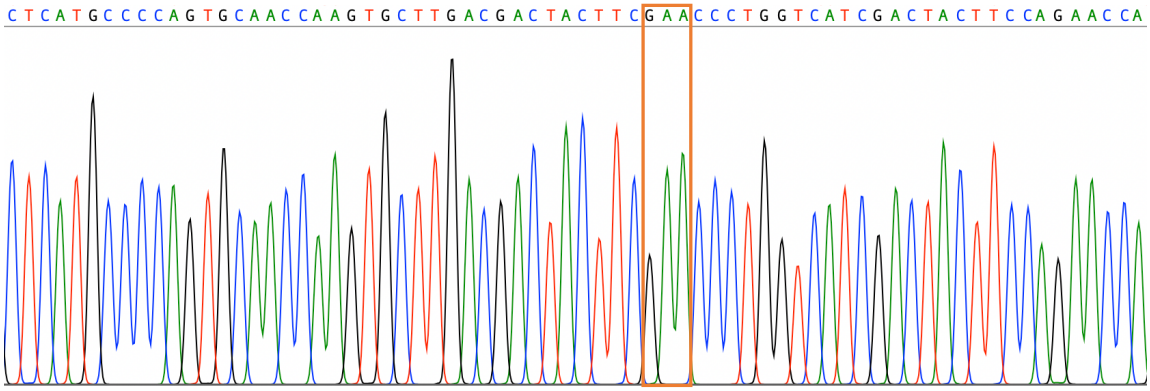


b

Results of single cell cloning experiment as identified by RFLP

Percentage	Outcome
67 %	homozygous for 121ins2
26 %	heterozygous clones or mixed population
7 %	wildtype

c



d

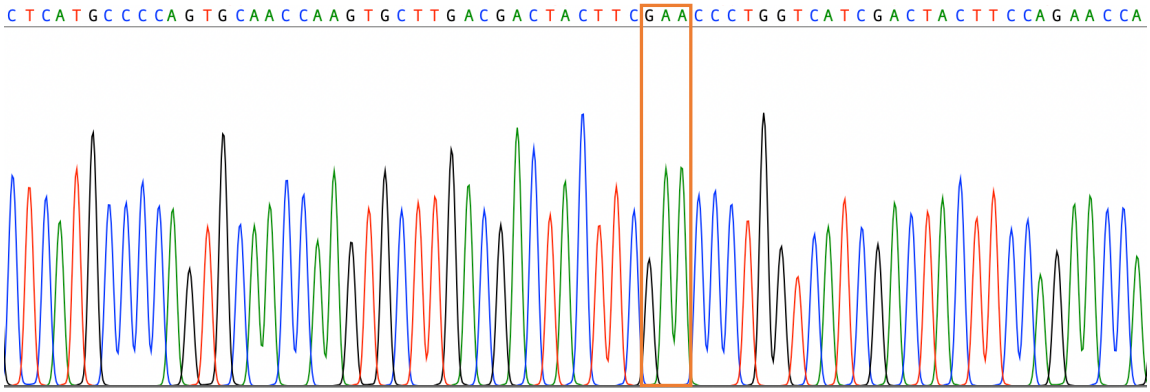


Figure 5.4 Clone screening via restriction fragment length analysis and subsequent confirmation of gene editing by Sanger sequencing

After CRISPR/Cas9 gene editing to create the 121ins2 mutation in the *SFTPB* gene and subsequent limiting dilution, emerging clones were analysed by restriction fragment length polymorphism (a). If the 121ins2 mutation was successfully generated, the ca. 638 bp PCR product (grey arrow) would be digested with BstBI into 457 bp and 181 bp fragments (white arrow). Results of clone screening via this method are shown (b). Sequencing of clone 10 after isolation (c). Sequencing of clone 10 before freezing (d). Clones with restriction fragment length polymorphism indicative of the homozygous insertion of the 121ins2 mutation were confirmed after isolation and during expansion by Sanger sequencing, as shown here in c, d for clone 10.

5.2.3 Characterisation of 121ins2 edited human ESC clones

The karyotype of the isolated 121ins2 edited RUES2 clones 3 and 10 were characterised along with their parental cell line using Karyostat microarray by Thermo Fisher Scientific Characterisation Services. This approach allows for detection of > 2 Mb for chromosomal gains and > 1 Mb for chromosomal losses. Normal karyotypes of a female individual with no aberrations were found in the parental RUES2 line (Figure 5.5a) and isolated clones 3 and 10 (Figure 5.5b and Figure 5.5c, respectively).

Subsequently, clones were analysed alongside the parental line for retention of their stem cell status via immunocytochemistry for stem cell markers. Nuclei of the parental RUES2 line stained positively for the transcription factor OCT4 and for the cell surface marker SSEA4 (Figure 5.6A). The same staining pattern was observed for the 121ins2 edited clone 10 (Figure 5.6B). Similarly, SOX2 expression was found in the nuclei and TRA-1-60 staining in on the cell surface of RUES2 colonies (Figure 5.6C), which was indistinguishable from staining observed in clone 10 derived cells (Figure 5.6D). Staining observed for clone 3 was similar to that for clone 10 and RUES2, and staining was negative for the A549 cell line used as negative staining control (data not shown).

In summary, the staining indicated that the editing and isolation of single cell clones of RUES2 did not affect the stem cell status of the selected clones. Further, it was concluded that both stem cell clones 3 and 10 retained a normal karyotype and are therefore both suitable for disease model creation. Future experiments used clone 10 as a 121ins2 edited RUES2 ESC line.

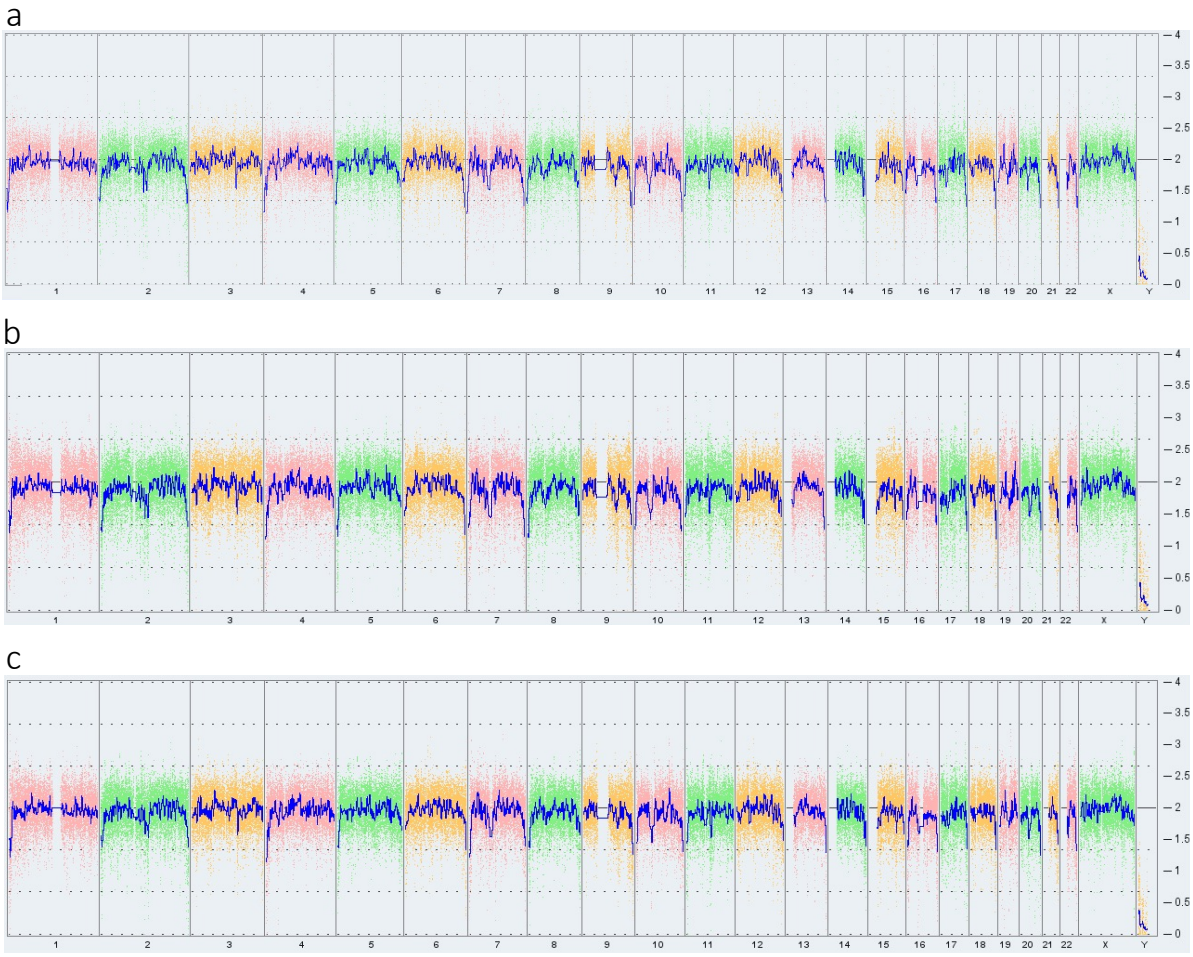


Figure 5.5 Karyostat microarray analysis to confirm normal karyotype

Whole genome DNA was extracted and analysed for chromosomal aberrations using the Karyostat microarray (Thermo Fisher Scientific Characterisation Services). Signal intensities from individual probes targeting somatic and sex chromosomes are shown as dots in green, red or orange and as average signal curve for the parental RUES2 cell line (a) and 121in2 edited RUES2 clones 3 (b) and 10 (c).

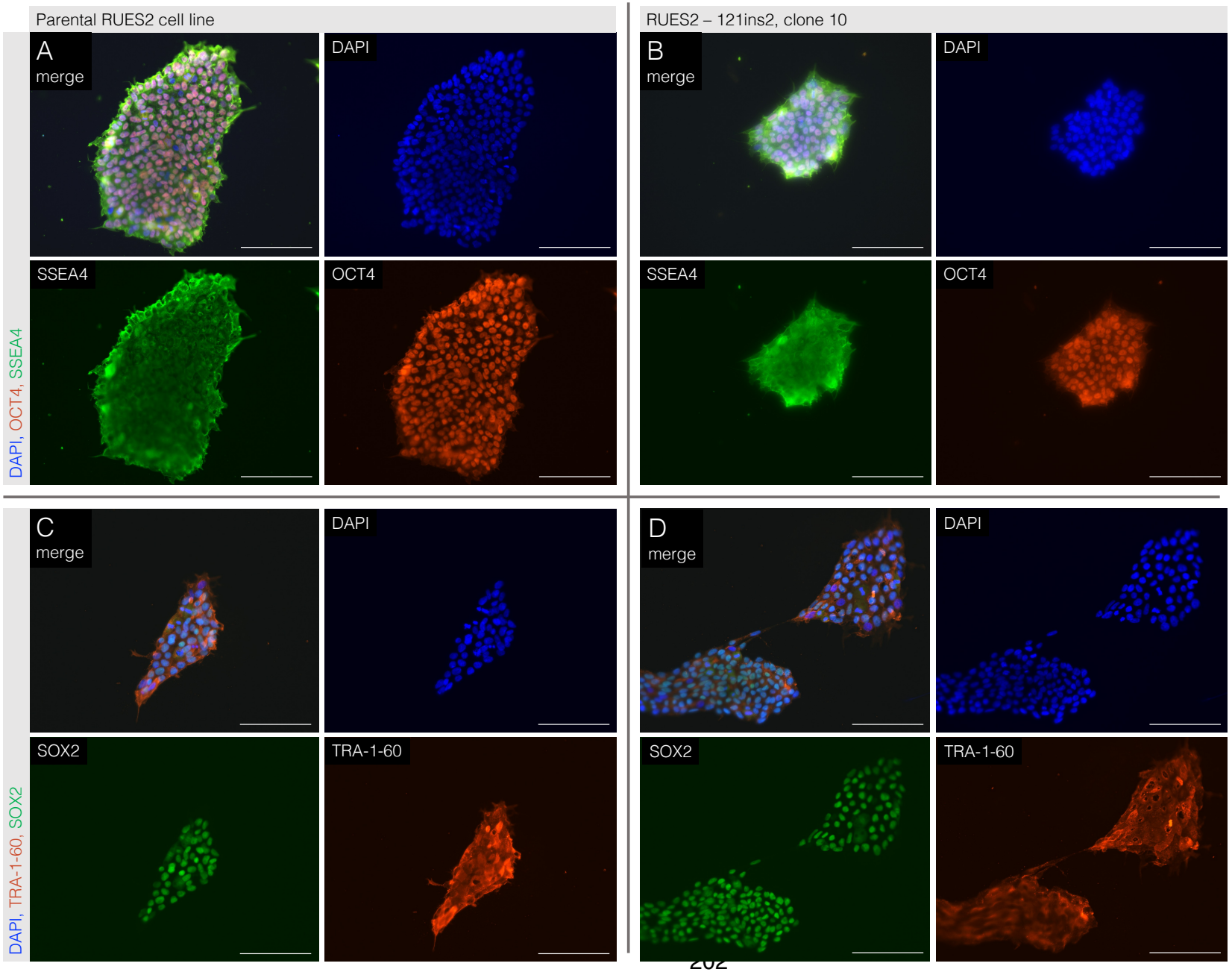

Parental RUES2 cell line
RUES2 – 121ins2, clone 10
A
merge
DAPI
B
merge
DAPI
SSEA4
OCT4
SSEA4
OCT4
DAPI, OCT4, SSEA4
C
merge
DAPI
D
merge
DAPI
SOX2
TRA-1-60
SOX2
TRA-1-60
DAPI, TRA-1-60, SOX2

Figure 5.6 Immunocytochemistry for stem cell markers of the RUES2 cell line and 121in2 edited RUES2 clone 10

One day after seeding, RUES2 cell colonies were fixed and subsequently the parental RUES2 cell line (A, C) and the 121in2 edited RUES2 clone 10 (B, D) were analysed for expression of stem cell markers SSEA4 and OCT4 (A, B) as well as SOX2 and TRA-1-60 (C, D) by immunocytochemistry. Separate channels for stem cell markers (red, green) and DAPI as well as merged images are shown. Scale bar 100 µm.

5.2.4 Generation of LBOs from the 121ins2 edited stem cell line and its parental line

LBOs were generated from the 121ins2 edited RUES2 ESC line and its parental cell line in parallel via the key stages as described in Figure 4.7 (Section 2.8.1). No morphological differences between cultures were observed during differentiation from ESC cells to LBOs (Figure 5.7), and cultures looked very similar to those generated in Section 4.2.2.1. Miniature buds were observed on day 10 in LBO progenitors which expanded and branched further after embedding in Matrigel. Matured LBOs on day 75 and 79 in culture are shown in Figure 5.7.

At the end of definitive endoderm induction on day 4, definitive endoderm induction efficiency was determined by evaluating the proportion of c-KIT$^+$/CXCR4$^+$ cells using flow cytometry (Section 2.13). The parental RUES2 line yielded 90.2 % c-KIT$^+$/CXCR4$^+$ double-positive cells (Figure 5.8a, b) and the 121ins2 edited line achieved 90.5 % (Figure 5.8c, d), thus the levels achieved were comparable in both lines and well above recommended threshold of ≥ 80 % for LBO generation [333, 378].

At the end of anterior foregut endoderm induction on day 6, cultures were analysed for expression of the transcription factor and anterior foregut endoderm marker FOXA2 by immunocytochemistry (Section 2.11, Figure 5.9). High proportions of nuclei stained positive for FOXA2 in both cell lines, indicating appropriate anterior foregut endoderm induction.

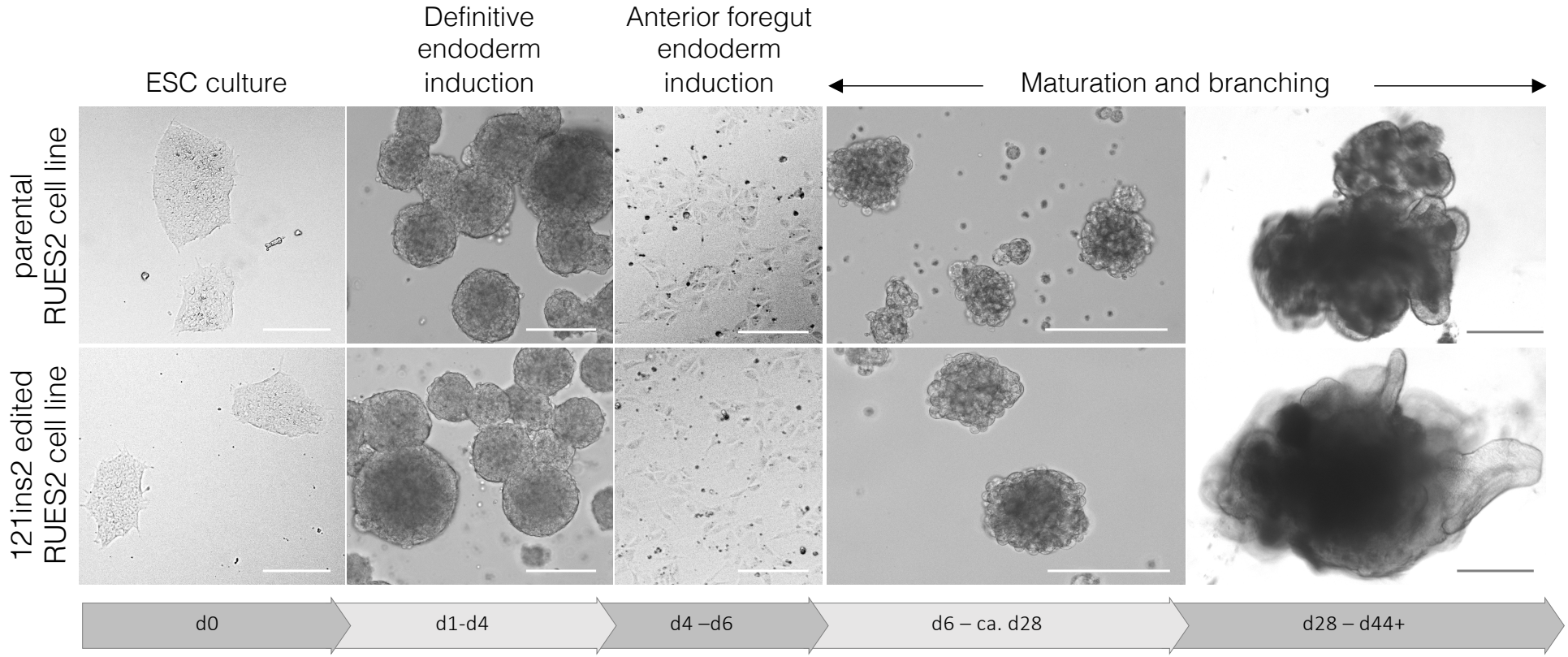


Figure 5.7 Brightfield images of LBO generation key stages from RUES2 cells and 121ins2 edited RUES2

Representative images of the key stages in lung bud organoid (LBO) generation are shown side by side for 121ins2 edited RUES2 cell line and their parental cell line. Images of ESC cultures were taken on day 0, on day 3 during definitive endoderm induction, on day 5 during anterior foregut endoderm induction and on day 12 for LBO progenitors. Mature LBOs were imaged on day 75 and 79. Scale bars 200 µm.

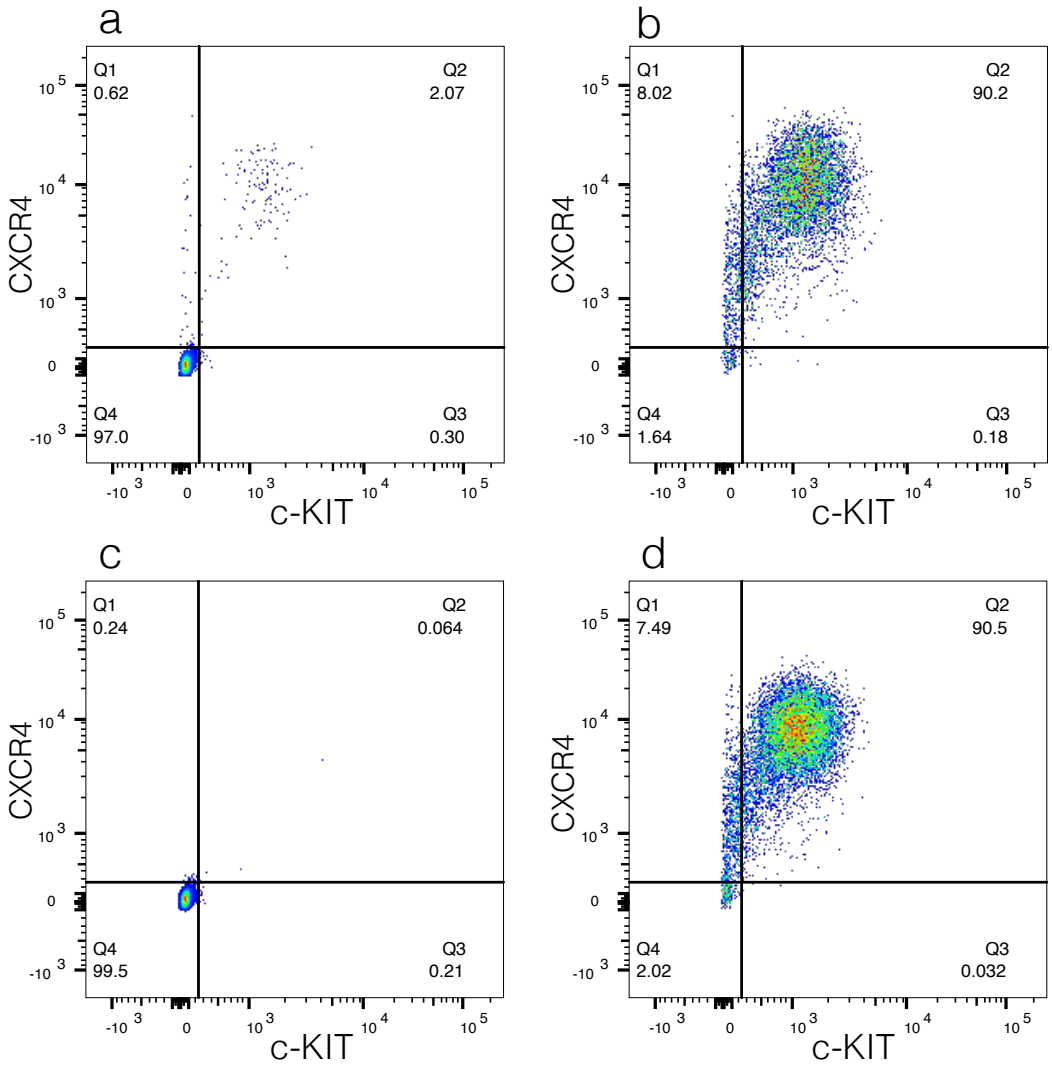


Figure 5.8 Examination of definitive endoderm induction efficiency by flow cytometry for c-KIT and CXCR4

On day 4 of the lung bud organoid (LBO) generation protocol, embryoid bodies were dissociated and fixed to determine definitive endoderm induction efficiency in cultures derived from the RUES2 cell line (a, b) or 121ins2 edited RUES2 cell line (c, d). Cells either remained unstained (a, c) or were stained for CXCR4 and c-KIT expression (b, d) and analysed by flow cytometry; ca. 6060 events were recorded for the RUES2 cell line and ca. 9420 events for 121ins2 edited cell line. Events were gated for cells (forward scatter/side scatter, FSC/SSC), single cells (forward scatter area/forward scatter height, FSC-A/FSC-H), and finally for definitive endoderm markers. Flow cytometry was performed at LSRII flow cytometer by Dr Kamran Miah.

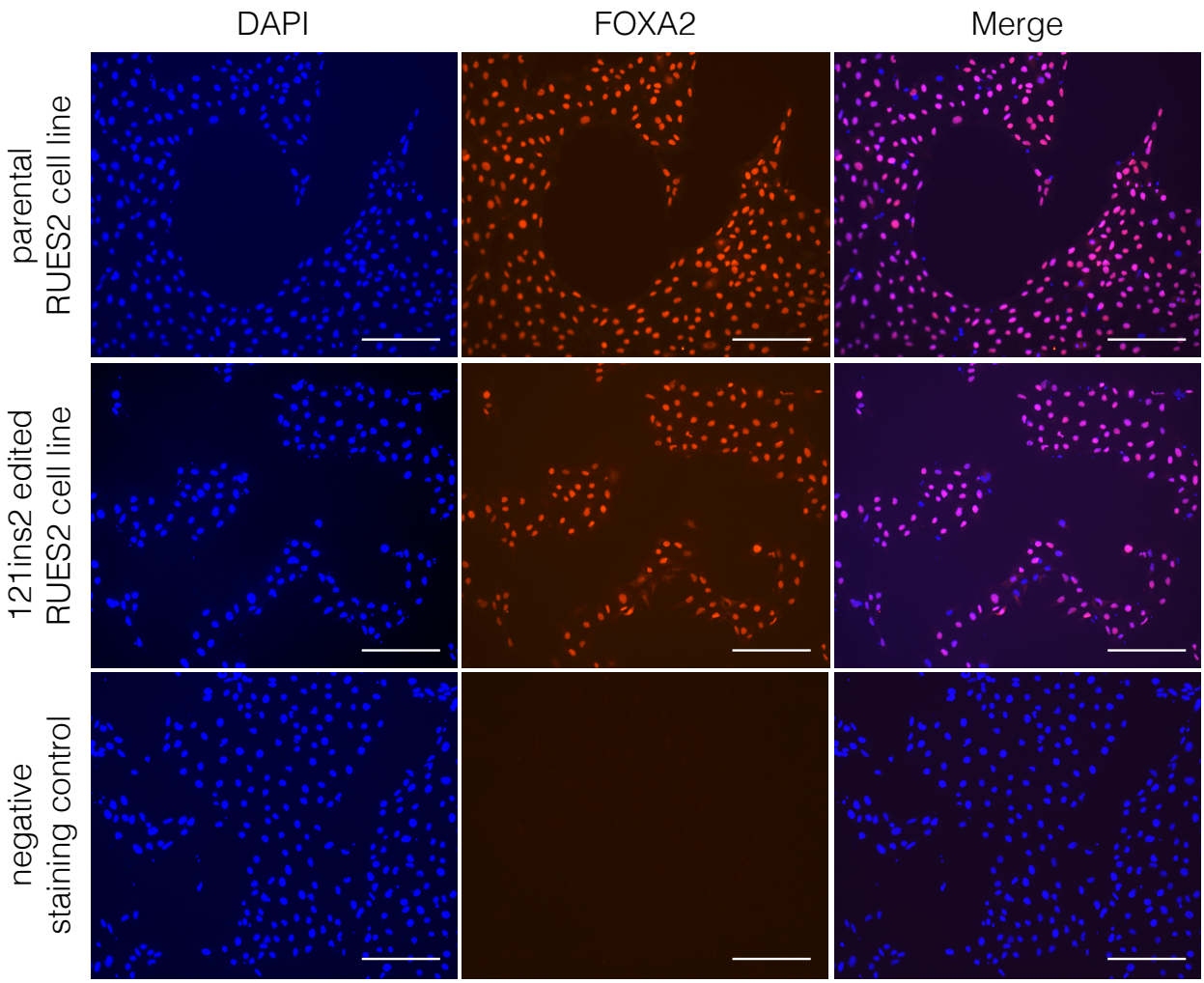


Figure 5.9 Verification of anterior-foregut endoderm induction by immunocytochemistry for FOXA2

On day 6 of the lung bud organoid (LBO) generation protocol, cultures derived from the RUES2 cell line or the 121ins2 edited RUES2 cell line were fixed and immunocytochemistry analysis of FOXA2 expression was performed. Staining omitting the primary antibody served as negative staining control. Images of nuclei (DAPI) and FOXA2 staining (red) are shown individually and merged. Scale bar 200 µm.

Mature LBOs generated from 121ins2 edited RUES2 cells and their parental line were analysed for SP-B expression by Western blot analysis (Figure 5.10). HEK293T cells transfected with pAAV2.CMV.SP-B (Section 2.4.2) served as a positive control for the immunoblotting of SP-B. Bands aligning with those reported in the literature for SP-B proprotein, processing intermediate, and mature SP-B homodimer (Figure 5.10a), were observed in transfected control cells at 42, 22, and 15 kDa, respectively (Figure 5.10b, lane 1, 2, grey arrows). No SP-B expression was observed in LBOs generated from either of the parental RUES2 or the 121ins2 edited line at this exposure level (Figure 5.10 b, lane 3, 4). At high exposure times (Figure 5.10c), two faint bands at ca. 42 kDa and 26 kDa were visible in samples from wildtype LBOs (Figure 5.10c, lane 3, white arrows) which correlates with the reported sizes of the SP-B proprotein and processing intermediate. This indicated low levels of endogenous SP-B expression in wildtype LBOs. The observed bands were attributed to the SP-B proprotein and processing intermediate, but no band attributable to the mature SP-B homodimer was observed, possibly due to overall low expression levels. SP-B expression is further indicative of distal lung identity in LBOs. The bands observed at 42 and 26 kDa were not present in samples from 121ins2 edited LBOs (Figure 5.10 c, lane 4). In comparison with the faint bands found in lysates from wildtype LBOs (Figure 5.10 c, lane 3), no bands < 50 kDa were found in lysates form 121ins2 edited LBOs (Figure 5.10 c, lane 4). This blot is indicative of the absence of SP-B expression, suggesting successful editing of LBOs resulting in SP-B deficiency.

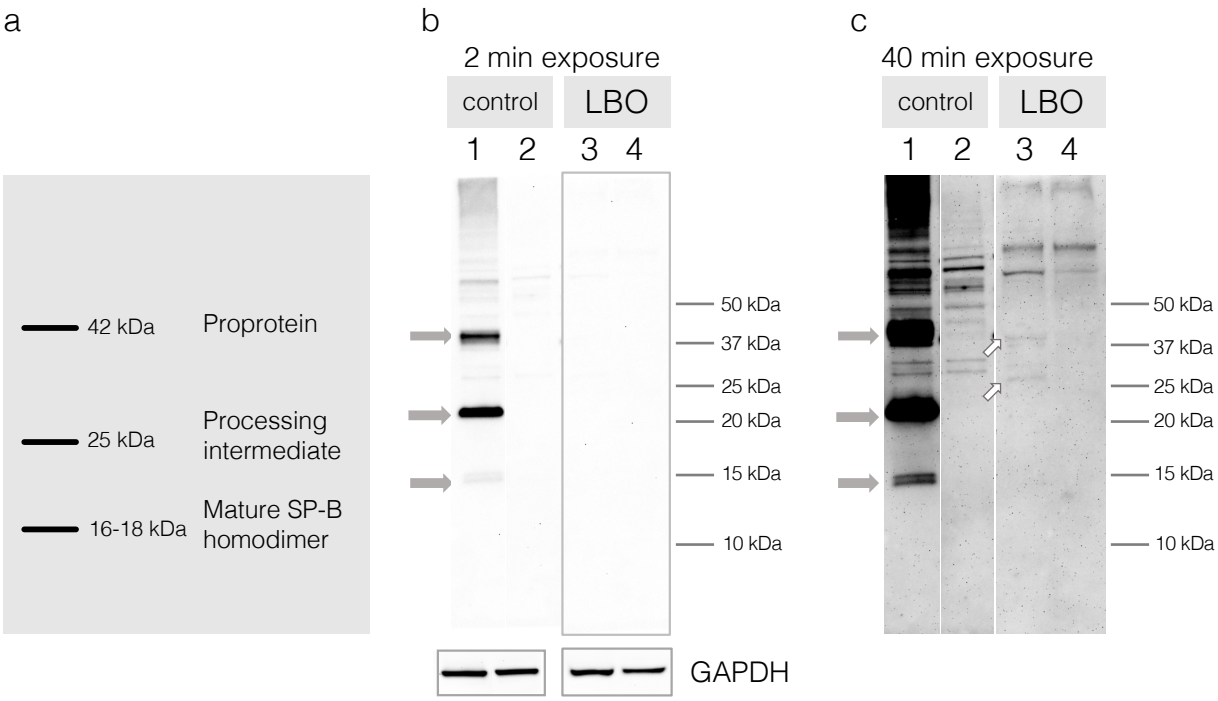


Figure 5.10 Western blot analysis for SP-B expression in wildtype and gene edited LBOs as well as transfected control cells

Schematic of Western blot to show the respective sizes expected of SP-B proprotein, processing intermediate and mature SP-B homodimer (a). After non-reducing SDS electrophoresis, membranes were immunoblotted for SP-B, followed by membrane stripping to blot for GAPDH as a loading control. For SP-B immunoblotting, 2 min exposure (b) and 40 min exposure (c) of the same membrane are shown. HEK293T cells were transfected with pGM705 (pAAV2.CMV.SP-B, lane 1) or remained naive (lane 2), lysed and 5 µg total protein analysed. An equivalent of three lung bud organoids (LBOs) aged day 77-81 generated from the parental, wildtype RUES2 cell line (lane 3) and from the 121ins2 edited cell line were lysed and analysed. Grey arrows indicate bands corresponding to the SP-B proprotein (42 kDa), processing intermediate (anticipated 25 kDa, here ca. 22 kDa) and mature SP-B homodimer (anticipated 16-18 kDa, here ca. 15 kDa). White arrows indicate bands specific to LBOs generated from the parental RUES2 cell line (ca. 42 and 26 kDa). Blot is representative of two technical repeats.

5.2.5 Gene therapy to restore SP-B expression in SP-B deficient LBOs

SP-B deficient LBOs generated from 121ins2 edited ESC were microinjected with rAAV6.2.CMV.SP-B (Table 2.6) to model rAAV gene therapy in a human disease model. Serotype 6.2 was shown to transduce LBOs at high levels and was identified as a suitable AAV serotype lead candidate, similar to results from studies in the human PCLS model (see Chapter 4). On day 5 after rAAV microinjection, LBOs were lysed and analysed for SP-B transgene expression by Western blot analysis (Figure 5.11). HEK293T cells transfected with pAAV2.CMV.SP-B (Table 2.1) served as a positive control for the SP-B Western blot. SP-B bands as reported in the literature (Figure 5.11a) aligned with the observed bands in lysates of transfected HEK293T cells for SP-B proprotein, processing intermediate, and mature SP-B homodimer at 42, 22, and 15 kDa respectively (Figure 5.11b, lane 1, grey arrows). In lysates of rAAV6.2.CMV.SP-B microinjected 121ins2 edited LBOs, but not naive 121ins2 edited LBOs, a strong band at 22 kDa and a band at 16 kDa were observed (Figure 5.11b, lane 3 and 4, white arrows). These were identified as the SP-B processing intermediate and SP-B homodimer. At longer exposure times, a faint additional band at ca. 42 kDa was observed in microinjected 121ins2 edited LBOs which can be attributed to the SP-B proprotein (Figure 5.11c, lane 4, white arrows). This blot shows successful SP-B gene transfer and therapeutic transgene expression in a human disease model with a clinically relevant rAAV vector.

The 121ins2 edited LBO model was generated with the potential to model therapeutic gene editing in mind. LBOs (n= 4) remained naive or were microinjected with Cas9 RNP comprising the sgRNA targeting the 121ins2 mutation as well as ssDNA donor to correct the 121ins2 mutation with HDR gene editing (Section 2.8.3). The sgRNA used the same PAM as in Section 5.2.2 that was published previously [324]. On day 3 post injection, genomic DNA was extracted and gene correction in exon 4 analysed by Sanger sequencing and ICE v2 (Figure 5.12). No indel formation was observed by ICE v2 analysis

and no interference was observed in the traces by visual assessment. Therefore, the therapeutic editing strategy as applied was unsuccessful and alternative gene editing and delivery strategies should be considered and optimised.

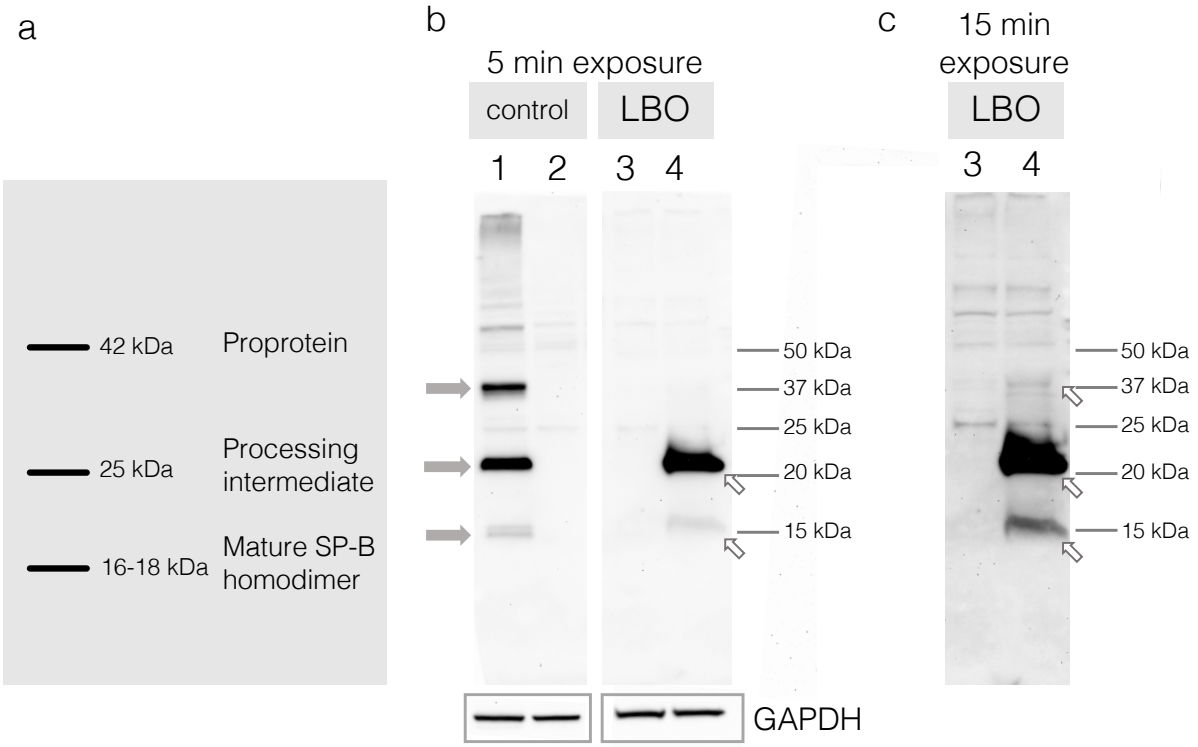


Figure 5.11 Western blot analysis for SP-B expression in 121ins2 edited LBO after rAAV6.2.CMV.SP-B microinjection

Schematic of Western blot to show the respective sizes expected of SP-B proprotein, processing intermediate and mature SP-B homodimer (a). Lung bud organoids (LBOs) generated from the 121ins2 edited RUES2 cell line were either microinjected with rAAV6.2.CMV.SP-B or remained naive. On day 5 after microinjection, non-reducing SDS electrophoresis was performed, membranes were immunoblotted for SP-B, followed by membrane stripping to blot for GAPDH as a loading control. For SP-B immunoblotting, 5 min exposure (b) and 15 min exposure (c) of the same membrane are shown. HEK293T cells transfected with pGM705 (pAAV2.CMV.SP-B, lane 1) or remained naive (lane 2), served as positive control; 5 µg total protein analysed. LBOs (n= 3, aged day 91-95) generated from the 121ins2 edited cell line remained naive (lane 3) or were microinjected with a total of ca. 2.75E8 GC of rAAV6.2.CMV.SP-B (lane 4). On day 5 post dosing, LBOs were lysed, lysates pooled and 72 µg total protein analysed for SP-B expression. Grey arrows indicate bands corresponding to the SP-B proprotein (42 kDa), processing intermediate (25 kDa, here ca. 22 kDa) and mature SP-B homodimer (16-18 kDa, here ca. 15 kDa) in transfected control cells. White arrows indicate bands specific to

microinjected LBOs; ca. 22 and 16 kDa in 5 min exposure and ca. 42 kDa, 22 kDa and 16 kDa in 15 min exposure blots. Microinjections were aided by Dr Yue Du.

Microinjection of Cas9 RNP and donor DNA in LBOs for 121ins2 correction

sgRNA target	PAM	Indel [%]	Model fit [R^2]	Knockin-Score [%]
TTGACGACTACTTCGAACCC	TGG	0	1	0

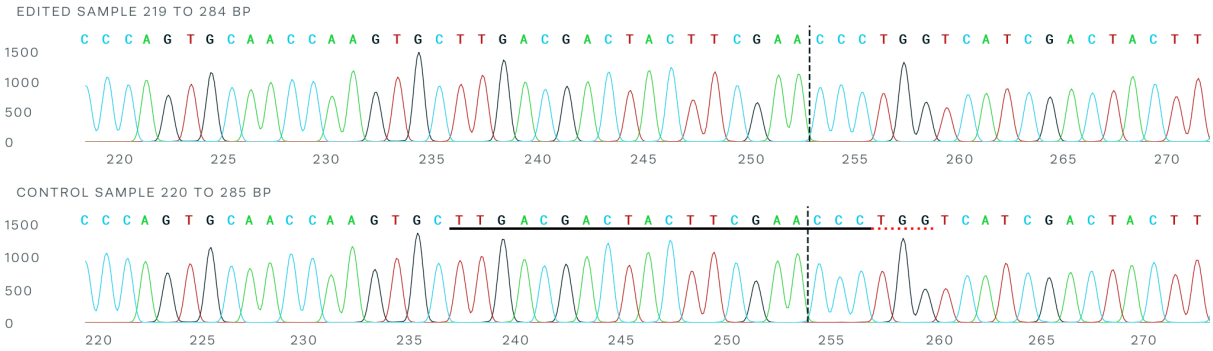


Figure 5.12 ICE analysis for therapeutic gene editing in 121ins2 edited LBOs

Lung bud organoids (LBOs, n= 4) generated from the 121ins2 edited cell line and aged day 99-103 remained naive or were microinjected with a total of 40 pmol Cas9 complexed with 300 pmol sgRNA well as 300 pmol ssDNA donor complexed in Lipofectamine™ CRISPRMAX™. Genomic DNA was isolated on day 3 post dosing and exon 4 analysed by Sanger sequencing. Results of ICE v2 analysis including traces of edited and control sample with highlighted sgRNA target site (underlined) and PAM (red dashed line) are shown. Sequence trace and result representative of n= 4. Microinjections were aided by Dr Yue Du.

5.3 Discussion

In this chapter, a strategy for editing *SFTPB* in human ESC was established and a 121ins2 edited human ESC line was generated (Figure 5.4). The edited cell line was characterised for retention of a normal karyotype and stem cell status (Section 5.2.3) and used to generate SP-B deficient LBOs (Figure 5.7). During all cell line characterisation studies and LBO generation at key stages, the RUES2 cell line and 121ins2 edited line, as well as their derived cultures, were indistinguishable. This indicates the absence of any detectable negative effects by the complex cell biological processes involved, resulting in e.g. loss of pluripotency or off-target gene editing.

5.3.1 SP-B deficient LBOs as model for SP-B gene transfer

The 121ins2 edited line were sequenced multiple times including after isolation, before cell banking, and at defrosting at the start of LBO generation to ensure that true clonal resources were created, and that the 121ins2 mutation was not lost. Interestingly, differences in SP-B expression observed between the wildtype and 121ins2 edited LBOs was modest (Figure 5.10). This could be due to low expression levels of SP-B in the wildtype LBO cultures. Unfortunately, LBO material availability was very limited in the studies presented. This was especially the case for wildtype LBOs, which restricted biological study replicates and ultimately prevented experimental repetition of the Western blot studies characterising SP-B expression, where the use of larger pools of organoids may have been helpful.

After rAAV6.2 gene transfer and expression of SP-B from the CMV promoter, the SP-B proprotein, processing intermediate and mature SP-B homodimer were observed in 121ins2 edited LBOs via Western blot analysis (Figure 5.11). Thus, microinjection of rAAV vector into the LBO lumen provided a useful model of SP-B gene transfer. Interestingly, after rAAV6.2 delivery of the SP-B transgene, higher levels of the SP-B processing

intermediate were observed compared with the mature SP-B homodimer (Figure 5.11), likely a result of limited proteolytic processing capacity, either due to the cellular prematurity of the LBO model, or true processing limits of the organelles involved in SP-B biosynthesis. The processing limits are a topic of interest for future SP-B gene therapy research, as accumulation of unprocessed SP-B processing intermediate could cause ER-stress and ultimately apoptosis of cells expressing the therapeutic SP-B. When comparing endogenous and exogenous SP-B expression in LBOs, the SP-B processing intermediate observed from endogenous expression (26 kDa, Figure 5.10) migrated in SDS-PAGE with a slightly higher molecular weight than the exogenous SP-B processing intermediate expression (22 kDa, Figure 5.11). One potential explanation for this is the glycosylation status of the SP-B intermediate (Figure 1.3), which lead to the observation of ca. 23-25 kDa intermediates previously [344, 395], and here, protein overexpression might have led to a differential glycosylation rate in wildtype versus rAAV6.2 transduced LBOs protein.

Due to limited availability of LBOs, and the need to maximise the opportunity to achieve detectable gene transfer, the CMV promoter was selected to drive high levels of transgene expression in LBOs as it mediated robust EGFP expression levels in LBOs in Section 4.2.2.3. Future SP-B gene transfer studies in SP-B deficient LBOs could involve experiments using a wider range of relevant promoters not prone to silencing long-term to drive high SP-B expression levels after rAAV6.2 delivery. Possible options include the CBA promoter used in the Zolgensma gene therapy product (EMA Procedure No. EMEA/H/C/004750/0000), or more cell- or tissue-specific promoters such as the *SFTPC* or *SFTPB* promoter investigated in this thesis in murine ATII cells (Section 3.2.4).

Due to limited LBO availability towards the end of this study, no further analysis of SP-B expression correction was conducted, however, experiments to show restoration of surfactant homeostasis were planned, such as assessment of potential increased levels

of mature SP-C. Additionally, a functional readout of rescue SP-B expression could be performed by measuring surfactant secretion levels as reported previously, using mass spectrometry for DPPC content [325]. Alternatively, a functional read-out could involve assessment of changes in the morphology of LB observed using transmission electron microscopy [321]. Although characterisation of changes in LBs has been widely reported, it has proved difficult to observe statistically significant differences [321]. Staining of acidic organelles (including LBs) using the lysotracker dye has previously been used successfully to differentiate the phenotype of Hermansky-Pudlak syndrome from wildtype in lung organoids; this might also be helpful for staining and characterisation of wildtype LBs compared with 'bubbly' LBs observed in SP-B deficiency, if they differ in acidity [396].

5.3.2 Gene editing in SP-B deficient LBOs carrying the 121ins2 mutation

The transfection of RNP in 121ins2 edited LBOs for 121ins2 correction did not result in indel formation or 121ins2 correction (Figure 5.12) and the limited availability of LBOs did not allow for further experiments. The absence of indel formation was most likely due to insufficient delivery of gene editing tools. To improve efficiency, the transfection protocol could be optimised. Additionally, the DNA donor could, if advantageous, be delivered by rAAV6.2, an approach which has been shown to strongly improve HDR gene editing and gene insertion of up to 2.2 kb in human PSC compared with the use of a conventional plasmid DNA donor [394]. Cas9 delivery could also be improved by rAAV delivery, however, this is less desirable from a translational point of view as prolonged nuclease expression increases the likelihood of off-target edits [397].

Crucially, an HDR-dependent editing strategy could be unsuitable for the LBO model at the chosen time point. HDR gene editing has relatively low activity, due at least in part to its dependence on the S and G2 phase of the cell cycle making actively proliferating cells a requirement for efficient editing [398]. Potentially, an earlier time point of LBO maturation,

where proliferation rates are higher, would have allowed for higher HDR gene editing rates. This is also a reminder that the level of cell proliferation in the developing lung also needs to be considered when designing a therapeutic gene editing strategy. Importantly, the neonatal lung, the target for SP-B gene therapy, continues to grow, expand and mature after birth, a property that could be helpful for gene editing involving the HDR pathway. Moreover, since the neonatal lung is still developing, successfully edited ATII cells with their stem/progenitor properties could serve as a founder cell population, that could support long-term expansion and disease correction.

In summary, a 121ins2 edited human ESC line was isolated and used to generate a human LBO model of SP-B deficiency. Microinjection of rAAV6.2 to the LBO lumen allowed for expression of mature SP-B protein, showing successful SP-B gene transfer to a human disease model of SP-B deficiency with a viral vector suitable for clinical development.

Chapter 6 – Discussion

6.1 Translational potential of rAAV vectors for gene therapy for SP-B deficiency

Surfactant protein B deficiency is a genetic disease with urgent patient need, very well understood pathophysiology researched for over 25 years (Section 1.2.4), and a mouse disease model [320]. Usually, understanding the disease and availability of disease models are the first steps to treatment development, but for SP-B deficiency, there is still no treatment option apart from rarely attempted lung transplantation.

Treatment of SP-B deficiency likely requires normalisation of SP-B production in ATII cells, making gene therapy an attractive therapeutic approach, especially utilising rAAV as a vector previously found safe in human lung [178, 399]. As of May 2020, Zolgensma is an EMA approved rAAV gene therapy for SMA, a disease that causes muscle wastage by degeneration of motor neurons and is often fatal within the first year of life [400]. Zolgensma showed tremendous therapeutic effect in clinical trials [401, 402], by enabling patients that usually cannot even sit without assistance to be able to dance, as showcased in a video by the trial participating hospital (https://bit.ly/3iIFrZ0). The efficacy and subsequent approval of Zolgensma is a strong affirmation for the field of rAAV gene therapy, as it shows that rAAV vectors can treat formerly fatal diseases and generate real patient benefit long-term (> 6 years [401]). Therefore, this should be seen as encouragement for rAAV vector development to treat rare diseases, especially as organs such as the lungs, which are more challenging to treat, are on the horizon. However, questions regarding the extremely high price of Zolgensma (ca. 1.8 million GBP) and other orphan drugs as well as their impact on health economics remain to be solved.

6.1.1 AAV serotype 6.2

Delivery is the first step in gene therapy development, requiring an understanding of viral vector tropism for target tissues. At the beginning of this thesis, very little was known about rAAV tropism for the lung parenchyma (Section 1.5.4), particularly the human lung. Previous difficulties in translation from mice to large animal models and humans have been observed ([179, 302], Section 1.5.5), with literature describing substantial variation in lung entry receptor expression between animal models and humans, as well as different receptor distribution within the lung [163, 183]. The hypothesis therefore emerged that mice are an inadequate model of viral vector tropism selection in the lung. In this thesis, vector tropism data from studies in mice, was collected to test this hypothesis and for future proof-of-principle experiments in the mouse model of SP-B deficiency (Table 6.1), in later phases of the project. The majority of the studies in this thesis, therefore, focused on human models of the lung parenchyma for vector screening, in the hope of identifying AAV serotypes suitable for swift translation for SP-B deficiency gene therapy and other diseases affecting distal lung.

In the search for a suitable AAV serotype for gene therapy for SP-B deficiency, three AAV serotypes were first tested for *in vivo* transduction of ATII cells in the murine lung parenchyma (Figure 3.3, Table 6.1). This was followed by screening of six naturally occurring AAV serotypes and two engineered variants, selected based on reported references for lung (Table 1.1), in human LBOs and PCLS as models (Chapter 4, Table 6.1). Amongst these serotypes, rAAV6.2 produced the most promising results. In particular, rAAV6.2 generated one of the highest levels of transduction in LBOs created from the AND-2 human ESC line in two independent experiments (Figure 4.9). Serotype 6.2 also resulted in the highest transduction rate in the PCLS model (Figure 4.5, summarised in Table 4.2), which was confirmed in a second PCLS donor. Therefore, rAAV6.2 was chosen as the lead serotype and it successfully delivered the SP-B transgene

in the LBO model of SP-B deficiency (Figure 5.11). The LBO model of SP-B deficiency was generated from clones obtained from the RUES2 human ESC line, showing that rAAV6.2 was able to transduce LBOs generated from ESC lines from two different donors.

The use of two different lung parenchyma models for AAV serotype selection with different characteristics (Table 6.1), generated from a total of four donors, is reassuring for the selection of rAAV6.2 as a candidate vector for gene delivery to the human distal lung, suggesting its selection is independent of donor- or model-specific effects. The rAAV6.2 vector was suitable for gene delivery to both models of the adult lung as well as developing lung. In addition, there is minimal change in the primary (glycan) entry receptor expression between foetal (> 20 weeks gestation), paediatric and adult lung parenchyma [354]. Therefore, one can speculate that rAAV6.2 may be suitable for lung gene delivery from foetal developmental stages through to the adult lung. This may allow for gene delivery to the developing lung *in utero*, to the neonatal lung for treatment of surfactant-associated genetic diseases, and to the paediatric or adult lung for later onset surfactant-associated genetic diseases.

Although serotype 6.2 might be a suitable choice for immediate translational development, in future it could readily be substituted by more efficient and/or more cell type-specific vectors. Serotype 6.2 could be replaced by either naturally occurring serotypes not yet tested for lung tropism, such as rAAV4 recently indicated for porcine alveolar tropism [403], or, more likely, engineered AAV capsids with higher specificity and efficacy for human ATII cell transduction. Although capsids can be engineered by rational approaches, the development of directed evolution for AAV capsids can greatly accelerate capsid engineering. Together with appropriate human models [302], such as lung organoids or human organ explants, capsid evolution approaches provide a powerful approach to identify the next generation of rAAV vectors for distal lung targeting. Random methods

such as error-prone PCR and capsid shuffling were first used to generate a heterogenous pool of AAV capsids ready for selection [404], however, this random approach creates many non-viable variants. Directed evolution can also be combined with rational design approaches to generate a more effective library with more viable capsids. Examples include the incorporation and subsequent selection of a peptide library with demonstrated affinity to the target cells in AAV capsids [405] and the targeted mutagenesis of the capsid variable regions (Figure 1.7) associated with viral entry [311]. In the future, computational library generation followed by generation of sequences on a gene chip could become more common, with removal of non-viable, non-assembling capsid variants from the libraries, as recently demonstrated by a machine learning algorithm trained on next-generation sequencing data from previous directed evolution experiments [406]. Of the fundamental challenges in the gene therapy field (manufacturing (cost of goods), delivery specificity and interaction with the immune system), specificity is central in reducing the pressure on the other two, as AAV capsids capable of more targeted delivery will require a lower dose to reach efficacy. Therefore, the field of capsid engineering has tremendous potential for advancing rAAV vector translation.

The studies described in Chapter 4, showed that the rAAV6.2 vector successfully transduced the PCLS and LBO models of the human lung parenchyma. The morphology of the transduced cells in human PCLS (Figure 4.5), as well as the ATII cell marker staining in LBOs (Figure 4.11), was indicative of transduction of ATII (or ATII-like) cells. However, transduction in these models did not allow for examination of specificity for ATII cells versus ATI or club cells, due to their absence in the LBO model [333]. LBOs show a distal lung bias [333] as also observed by the large areas staining positive for distal lung markers in Chapter 4 (Figure 4.11).

The LBO model was the first human PSC, self-organising lung organoid model to be mapped to the second trimester, therefore it was the most mature lung organoid model available; the ATII (or ATII-like) cells were well characterised, showing SP-B and SP-C expression as well as LBs with mature phenotype [333] making LBOs a well suited model to study diseases and their treatments involving ATII cells [407]. Regarding ATI cells, however, LBOs contain only ATI progenitor cells expressing the very early marker HOPX [333]. Lung organoid protocols published while this thesis was in preparation kept the general differentiation steps as outlined in Figure 4.7 for LBO generation [333, 408] and experimented with different media components and growth factors. This led to increased cell variety, including larger mesenchymal cell populations for orchestration of lung branching morphogenesis, club cells [407, 409, 410] and ATI progenitor cells, now expressing several (early) ATI markers in addition to HOPX [325, 396, 407, 409]. The greater cell type variety in more recent organoid models could allow for investigation of individual cell type tropisms and specificity. However, there appears to be a 'glass ceiling' in organoid development, since no protocol has been generated to allow development representative of distal human lung past the second trimester of gestation from PSC. Generation of organoids representing adult tissue from PSC, rather than adult stem cells, is a general challenge in the organoid field. Several authors also tried engraftment *in vivo* to enhance maturity of lung organoids [333, 365, 409], but with only modest success (not past the second trimester limit) and while interesting for studying maturation, from the viewpoint of model generation limiting to fewer research groups and omitting the 3R potential of organoids.

Human *in vitro* models of the lung parenchyma with mature ATII (or ATII-like) phenotypes can also be generated more quickly without tissue donations, by differentiation of human PSC directly into ATII cells [324, 411]. This approach, however, comes at a cost, either in cell type diversity (with only ATII cells organised in spheres [324]), or in terms of self-

organisation and polarisation of the cultures [411], and is therefore less suitable for vector selection. However, it was possible to generate these cultures and other organoids from iPSC derived from patient fibroblasts, opening the possibility for personalised drug screening for e.g. IPF [412]. Interestingly, the current focus on Covid-19 research might accelerate the identification of such models, as ATII cells are target cells of SARS-CoV-2 research and lung organoids were shown to accurately model viral entry [413].

As closely related species, NHP *in vivo* studies could form a bridge between results obtained in *in vitro/ex vivo* human lung models and clinical SP-B gene therapy trials. NHPs such as macaques, baboons and African green monkey are commonly used in biomedical research requiring a large animal model [414] and have been used for studying human immunology and pathophysiology of respiratory infections, including influenza viruses [415], coronaviruses [416], and human respiratory syncytial virus [414, 417], making NHPs an attractive model for viral gene therapy to the respiratory tract. Due to the field's interest in cystic fibrosis, most NHP studies focused on rAAV airway transduction (Table 1.1), however, a study comparing rAAV1 and rAAV5 in Rhesus macaques after intratracheal lung delivery *in vivo* got closest to investigating rAAV lung parenchyma tropism in NHP [255]. In this study, the authors investigated vector genomes and luciferase activity in various lung regions, including parenchymal, where apart from small airways and blood vessels the majority of the tissue should be ATI and ATII cells [255]. A ca. 10-fold higher transduction level was observed with rAAV1 compared with rAAV5 in vector genome copy number over all regions, including parenchymal, again discouraging the use of rAAV5 in the lung parenchyma. In Chapter 4, serotype 5 did not transduce human LBOs in two independent experiments (Figure 4.8, Figure 4.9), or human PCLS from three donors (Figure 4.5) and was therefore determined as negative for transduction of human lung parenchyma. This showed, in line with data from the NHP

model [255], that rAAV5 is likely a poor vector choice for gene therapy to the lung parenchyma.

AAV serotype 6 has not been investigated in lung NHP models so far, but just as AAV1 it uses glycans containing both α-2,3- and α-2,6-linked sialic acid as receptors and the AAVR co-receptor, with the additional flexibility to use heparan sulphate as glycan receptor (Table 1.1). This means that glycan receptors are less likely a hinderance for serotype 6 compared with AAV1. AAV6.2 is AAV6 with the F129L change; with 129L found in most other AAV serotypes (129F is only found in AAV5 and AAV6, see Figure 1.7). This change is not in a region associated with heparan sulphate binding, therefore the prediction is that rAAV6.2 will be able to use the same glycan entry receptors as rAAV6.

AAV6, but not AAV1, was additionally shown to transduce more efficiently in presence of its co-receptor EGFR, as highlighted by 37-58 % decrease in transduction after EGFR knock-down *in vitro* [284]. EGFR plays an important role in lung homeostasis and development [418], where it is physiologically expressed in ATII cells as a key regulator of ATII cell maturation [418, 419] and branching morphogenesis [420]. Therefore, EGFR expression is expected in normal developing, neonatal lung, but in order to benefit rAAV6 transduction, EGFR needs to be expressed on the apical surface. Instead, EGFR expression is thought to be located on the basolateral surface in adult polarised epithelial cells [418], as observed in e.g. gut epithelium [421], but it remains unclear how this would apply to the neonatal (still developing) lung.

An additional component necessary for efficient rAAV6.2 entry is AAVR (Table 1.1), the presence of which was confirmed in LBOs (Figure 4.12). In HAE cultures, AAVR expression was also observed, but only on the basolateral, and not the apical, surface of the cultures [180]. The authors also hint that lack of a co-receptor, such as AAVR, could help explain the lack of efficacy in clinical trials for cystic fibrosis using rAAV2, (which also

requires AAVR for efficient transduction, Table 1.1). In this thesis, AAVR was not studied for its expression in ATII cells on the apical versus basolateral surface, therefore the levels and location of AAVR expression in human ATII cells remain unclear. In the LBO model, however, transduction with the AAVR-dependent vectors rAAV2 and rAAV6 was observed, along with AAVR receptor staining in LBOs (Figure 4.12). Closer characterisation of AAVR localisation by costaining of apical and basolateral surface markers with AAVR as performed in HAE cultures [180], could determine if AAVR receptor expression is rather located on the apical surface or basolateral (such as in HAE) and pose a hindrance to transduction of AAVR-dependent serotypes, strongly diminishing the number of eligible serotypes (Table 1.1).

In summary, there are still multiple unknowns regarding the requirement for AAV entry receptors, and also co-receptors, in human lung and human neonatal lung, which can complicate the rational selection of AAV serotypes. The LBO and PCLS models used in this thesis were selected for their potential to model the effects of delivery of rAAV to the apical surface of the lung; and both models showed high levels of transduction with rAAV6.2. As there are likely still gaps in the knowledge of receptors and other host factors orchestrating rAAV vector entry, knowledge of the receptors is more suitable to predict lack of or diminished transduction due to the absence of an entry receptor, rather than predicting that a vector will transduce lung. With the current knowledge on entry receptor expression in the human distal lung, there are no data on the absence or otherwise of receptors necessary for rAAV6.2. Therefore, on the basis of LBO and PCLS transduction and consideration of the literature on human lung receptor expression rAAV6.2 can still be recommended for gene delivery to the human lung parenchyma.

Another potential hurdle in the translation of inhaled gene therapy vectors (Section 1.4.2) is trapping of the vector in the lung mucus. One of the possible reasons for the failure of

rAAV2 to demonstrate efficacy in cystic fibrosis in clinical trials [177], was the subsequently observed trapping of the vector in lung mucus [185]. This mucus trapping phenomenon was observed in rAAV particle tracking experiments, where adhesive binding and steric obstruction of rAAV2 was noted in sputum from cystic fibrosis patients [185]. Initially, the ability of rAAV2 to bind heparan sulphate was suspected to be at least partially responsible for this vector trapping. Later, however, rAAV6 was shown to be able to rapidly diffuse through sputum from cystic fibrosis patients, suggesting a serotype-dependent (as yet undefined) trapping mechanism [287], which is unlikely to involve adhesive binding of AAV entry glycans since rAAV6 binds all known AAV glycan receptors including heparan sulphate (Table 1.1). Assuming that rAAV6.2 behaves similarly to rAAV6 in this regard, rAAV6.2 vector particles would also be expected to rapidly diffuse thorough the mucus barrier. Interestingly, barriers to efficient airway delivery, such as mucus, may be bypassed when directly delivering to the lung parenchyma. Delivery to the airspaces can be achieved using a micro-spray catheter, guided by endoscope imaging, through the endotracheal tube and tracheobronchial tree to the alveolar region [422]. Therefore, the study of vector deposition close to the target alveolar space is an important next step in rAAV gene therapy development, and ideally should be investigated in large animal models such as piglets that contain the barriers to pulmonary gene therapy such as antibodies and alveolar macrophages in the lung parenchyma (Section 1.4.2) and have a similar lung development and anatomy as neonatal humans ([423], Table 6.1).

From case studies of individuals with partial SP-B deficiency, it may be possible to estimate how much SP-B expression is required for clinical benefit. As a starting point, it is likely that 50 % of SFTPB expression results in normal lung function as individuals heterozygous for the 121ins2 mutation are without symptoms (Section 1.2.4). In a patient with a splice mutation, 12-16 % of adult, physiologic, mature SP-B expression levels, as determined by Western blot of SP-B protein in lung homogenates [9, 77], enabled the patient to survive

on supplemental oxygen without intubation for 3 months before they underwent bilateral lung transplantation [77]. Although there are many case studies [70], details of the percentage of normal SP-B expression levels supporting spontaneous breathing are extremely rare. Data for residual SP-B expression, compared with normal neonatal levels, would be helpful to estimate target therapeutic levels. In the LBO model of SP-B deficiency, high levels of SP-B expression were achieved, several fold higher than the endogenous levels observed in wildtype LBOs (Figure 5.11), demonstrating that the rAAV6.2 vector is capable of mediating high levels of SP-B expression in a model of SP-B deficiency. Once a functional read-out is developed in this model, such as DPPC secretion levels and SP-C proteolytic processing, it could aid in proof-of-principle experiments for translation, while also delivering on the principle 'R' of the '3Rs' in animal research – Replacement (Animals (Scientific Procedures) Act 1986).

The potential for therapeutic benefit ultimately depends on how efficiently rAAV6.2 (or an improved vector) could transduce ATII cells in ventilated neonates, and the percentage and distribution of transduced ATII cells required to express SP-B. Once SP-B is expressed, correctly processed, and secreted from the cell as surfactant, the question is how well can that surfactant spread across the lung surface to prevent local atelectasis and lung injury. Quick absorption of surfactant to the air-liquid interface, as well as rapid lateral distribution during inhalation, are prerequisites for normal surfactant function (Section 1.2.1). Native porcine pulmonary surfactant showed very rapid (seconds or less) lateral distribution of pulmonary surfactant to equilibrate surface pressure at physiologic temperatures, while surfactant lipids without surfactant proteins did not reach equilibrium [424]. Overall, this indicates that functional pulmonary surfactant can rapidly spread across the lung surface, leading to the hypothesis that vector-transduced, surfactant-producing cells do not need to be completely evenly distributed throughout the lung for therapeutic benefit. Importantly, packaging of pulmonary surfactant in the ultra-

condensed, dehydrated state that can only be achieved with functional SP-B, is likely required for optimal, quick LB surface adsorption [56], therefore it is likely that only LBs from SP-B-expressing ATII cells can contribute to surfactant homeostasis.

Table 6.1 Comparison of lung models for rAAV gene therapy investigation

	Multi organ model	Model for human lung	Developing lung model	Model of SP-B deficiency
Wildtype mice	Yes	No	No	No
SP-B deficient mouse model	Yes	No	No	Yes
Wildtype pig	Yes	No	(Possible)	No
PCLS	No	Yes	No	No
LBO	No	Yes	Yes	No
SP-B deficient LBO	No	Yes	Yes	Yes
NHP	Yes	Yes	(Possible)	No

6.1.2 Promoters for therapeutic transgene expression in ATII cells

In gene therapy, promoters are designed to drive therapeutic levels of transgene expression in target cells. In this thesis the CMV, hCEFI and *SFTPB* promoters (as in Figure 3.4) were investigated in combination with SP-B and soSP-B transgenes. The soSP-B transgene was developed with the aim of boosting transgene protein levels, but was not taken further, as ultimately it did offer higher transgene expression levels than the SP-B transgene in human A549 cells (Figure 3.5). Human H441 cells when grown at the air-liquid interface, however, have been shown to model ATII cells more closely than H441 or A549 cells in conventional, submerged culture conditions in terms of the similarity of the transcriptome to human ATII cells *in vivo* [344]. Such a model might be more suitable for further investigation of human codon-optimised transgene cDNA.

The CMV, hCEFI and *SFTPB* promoters were firstly investigated in mice *in vivo* (Figure 3.10) using an adapted RNAscope method (Section 3.2.3, Section 2.14.2) to quantify transgene mRNA expression after rAAV transduction (Figure 3.9). The detection of high levels of vector DNA with conventional RNAscope probes was surprising and not previously reported although, in retrospect, not completely unexpected at a high (1E11 GC) rAAV dose to the lung. Recently, this phenomenon was also described by other authors who investigated rAAV DNA distribution in mice after IV dosing and circumvented the problem by designing sense/anti-sense probes while using nuclease digests as controls [425]. Thus, it was independently confirmed that the conventional RNAscope protocol (Section 2.14.1) was unable to differentiate between transgene mRNA and vector DNA. Compared with conventional RNAscope methods, the protocol developed in Chapter 3 provides the field with a more accurate quantification of localised transgene expression and does not require additional probe design.

The hCEFI promoter (CpG-free form of the **h**uman **C**MV enhancer element, **EF**1a promoter and a synthetic **I**ntron) was initially developed for long-term transgene expression in response to the observation that the CMV promoter was prone to transcriptional silencing after non-viral vector delivery to the murine lung [97]. The hCEFI promoter resulted in the highest levels of SP-B mRNA expression in the target ATII cells after rAAV delivery in mice *in vivo* (Figure 3.11), and analogously, in comparisons of hCEF (an intron free derivative of the hCEFI element), CMV, and EF-1alpha promoters in murine lung after rSIV.F/HN delivery, hCEF also mediated the highest and sustained transgene expression levels [7]. Interestingly, the hCEFI promoter was less robust than the CMV promoter in human PCLS and the LBO model (Table 4.2). The reason for this is unknown, but could be due to reduced availability of transcription factors required for high-level hCEFI promoter activity in these models. The LBO model is less developmentally mature and could therefore have a different transcription factor profile than adult lung. The rapid expression kinetics of the CMV promoter [97] might be favoured at the early endpoints used in PCLS studies. In addition, a biological effect limiting expression levels driven by the hCEFI promoter in human ATII cells (in the context of rAAV vectors) cannot be excluded. Due to the limited availability of the LBO model of SP-B deficiency, only the CMV promoter (vector rAAV6.2.CMV.SP-B) was tested for SP-B expression in LBOs (Figure 5.11). Overall, however, hCEFI remains an attractive candidate promoter for lung gene therapy, especially as it has already been used safely in a clinical trial of aerosolised plasmid DNA/cationic lipid (GL67A) formulation to cystic fibrosis patients [174]. Interestingly, EGFP expression was clearly detectable in human LBOs long-term after rAAV6.2.hCEFI.EGFP transduction (Figure 4.10) indicative of its suitability for long-term expression in the human lung (Figure 4.10); a feature particularly beneficial in gene therapy for SP-B deficiency.

Generally, there is a trend in more recent gene therapy clinical trials towards using tissue- and cell-specific promoters, because they can drive high levels of expression in target

cells, while also reducing off-target expression [103]. The endogenous human *SFTPB* promoter (*SFTPB* promoter sequence from -1040 of start codon as described in Pilot-Matias [336]) is a ATII- and club cell-specific promoter [426]. The human *SFTPB* promoter was previously shown to mediate ATII and club cell specific expression after intratracheal delivery of rAdV to murine lung [427] and resulted in the second highest expression levels in the murine ATII cells after intranasal delivery (Figure 3.11). From a therapeutic perspective, the use of human endogenous promoters for therapeutic transgene expression has two advantages: firstly, endogenous regulation of expression levels could potentially avoid excessive expression levels in the target cells causing ER-stress and ultimately apoptosis [428]. Secondly, endogenous promoters also allow for regulation of gene expression in response to physiologic changes, e.g. down-regulation of the *SFTPB* promoter by the TGF-β pathway in inflammation and lung injury [426]. There is also potential to design a minimal *SFTPB* promoter with possibly higher expression levels or to combine it with other enhancer elements as has been achieved with the *SFTPC* promoter [102]. However, a limitation of such designs is that they might lack cis-acting regulatory elements for endogenous regulation of expression. To avoid high levels of ER-stress which may ultimately lead to apoptosis of therapeutic-protein expressing cells, the promoter expression strength could be titrated in a human ATII cell model such as ATII cells generated from human PSC [411, 412] against markers of ER-stress involved in the unfolded-protein response pathway, e.g. proteolytic processing of the marker ATF6 [429]. Such future studies could also investigate if high expression levels of SP-B transgene could elicit ER-stress in ATII cells, as might be a potential risk due to the complex proteolytic processing and hydrophobicity of SP-B, but has not been shown.

Lastly, the CMV promoter mediated the lowest transgene expression levels in murine ATII cells *in vivo* (Figure 3.11) but was incorporated in the rAAV6.2.CMV.SP-B vector to model gene therapy in a human model of SP-B deficiency (Figure 5.11) due to its robust

performance in this model. The CMV promoter can drive robust expression in a variety of tissues and is still frequently used in current clinical trials [103]. However, there is evidence from non-viral vector systems [97, 430, 431] and rAdV delivery [432] that the CMV promoter is sensitive to transcriptional silencing in the lung and thus may be unsuitable for long-term transgene expression in the lung. Interestingly, CMV-dependent rAAV5-medaited lung transgene expression appears to be less sensitive to this phenomenon [218], indicating that CMV promoter silencing is not only tissue dependent [433-436], but also vector-dependent and highlights limits to the modularity of gene therapy building blocks (Section 1.3.1.1). When designing a translational vector for use in the human lung, the CMV promoter is a risky choice as long-term expression using CMV in rAAV vectors in the lung has not yet been sufficiently investigated and, with many other promoters to choose from (Section 1.3.1.1), this risk is unnecessary.

In summary, amongst the vectors investigated in this thesis, the hCEFI and *SFTPB* promoters are more attractive choices translationally, especially with the hCEFI promoter already used in clinical trials [174]. Promoters should be further investigated in models with a transcriptome characterised as similar to human ATII cell *in vivo*, such as primary human ATII cells, or H441 cells or bronchoalveolar donor cells grown in air-liquid interface [344, 413]. Alternative promoters could include the CBA promoter used in the Zolgensma gene therapy product enabling long-term expression in motor neurons (EMA Procedure No. EMEA/H/C/004750/0000) or the CASI promoter (CMV enhancer, chicken β-actin promoter, UBC enhancer with splice donor and acceptor sequences) which enabled long-term SP-B expression in ATII cells after rAAV6.2FF gene transfer in the Melton/Whitsett mouse model of SP-B deficiency [161], however, it remains unclear how well promoters selected in mice translate to human ATII cells.

6.1.3 Investigation of rAAV gene therapy in the mouse model of SP-B deficiency

The potential for gene therapy to treat SPB-deficiency has attracted much interest. Several non-viral formulations have shown proof-of-principle for correction in a murine disease model [321, 322], but in all cases the duration of the effects was only modest, measured in days to weeks (reviewed in Section 1.6.1). During the preparation of this thesis, a study was published investigating the potential for rAAV gene therapy in the mouse model of SP-B deficiency developed by Melton/Whitsett [320] using the serotype AAV6.2FF [161]. As determined by flow cytometry, ca. 79 % of ATII or ciliated cells (identified by $CD45^{-}$/$EpCAM^{++}$) were transduced after intratracheal administration of 1E11 GC rAAV6.2FF.CASI.SP-B vector (no other serotypes were tested). Delivery of rAAV to pups aged 8-10 days, with doxycycline withdrawal for SP-B deficiency phenotype induction 11-13 days after dosing, resulted in a median survival of 205 days, highlighting that rAAV can mediate long-term survival *in vivo*. Given that rAAV5 achieved the highest transduction rate in murine ATII cells in Chapter 3 (Figure 3.3), and transduction of ATII cells and ciliated cells was not differentiated by Kang *et al*. [161], the question remains whether rAAV5 might lead to even longer survival in this model. Despite expressing SP-B in ATII and ciliated cells using a constitutive promoter, no adverse effects by off-target expression, measured by body weight or inflammatory cytokines, were observed during the study [161]. This indicates that constitutive promoters, such as the hCEFI promoter used in this thesis (Figure 3.11), might be suitable for gene therapy for SP-B deficiency, despite expression in non-target cells.

Kang *et al.* further showed that rAAV6.2FF.CASI.SP-B could rescue pups and adult mice from respiratory distress by rAAV dosing 2 days after doxycycline withdrawal in adult mice, or at the same time as withdrawal in pups aged 13-15 days [161]. Both scenarios are more representative of the clinical situation in which rAAV could be delivered to neonates diagnosed with SP-B deficiency shortly after birth. Unfortunately, neither the proportion of

transduced cells, nor the degree of SP-B expression/correction achieved was reported in this study [161]; this would have been extremely helpful to benchmark the correction level required for survival in this model and for extrapolation to future dosing required in the clinic [161]. Also in this study, the authors investigated rAAV6.2FF transduction in human PCLS generated from foetal lung (16 weeks gestational age), and demonstrated successful reporter gene and SP-B gene transfer [161], in line with results presented in this thesis in human LBOs and human adult PCLS (Table 4.2). Although none of these models directly represent neonatal lung, there is increasing evidence from multiple models that rAAV6.2FF can transduce both adult and foetal developing lung.

6.2 Challenges for translation of rAAV gene therapy for SP-B deficiency

Despite the exciting advances in clinical trials leading to the approval of rAAV gene therapies, there are still translational challenges in the clinical development of rAAV gene therapy (Section 1.5.3), including toxicities observed at very high rAAV doses, declining expression over time and the interaction of rAAV vectors with an immune system primed by wildtype AAV exposure, or previous vector administration. Other than challenges in clinical development, challenges regarding high cost of gene therapies (high cost of vector production, high costs for payers) and its impact on health economics need reconsidering.

6.2.1 Immune system interaction

Understanding of the immune response to gene therapies is crucial to successful development. As argued previously (Section 1.4.2), the lung has evolved multiple barriers for host defence, including immunological barriers. Pre-existing immunity to AAV capsids due to exposure to wildtype AAV has been a major limitation to rAAV gene therapy. Currently, pre-existing immunity to rAAV vector capsids above a certain serum dilution

threshold tested in ELISA is an exclusion criterium in rAAV clinical trials [437]. Pre-existing immunity has not been researched for AAV6.2 capsid specifically, but can be approximated by AAV6 because the F129L substitution in AAV6.2 (which is the only difference between AAV6 and AAV6.2) is present in AAV1. It was shown that serum from of individuals seropositive for AAV6, 100 % cross-reacts with AAV1 (but only 73 % *vice versa*) [223]. Studies conducted in Europe reported 37 % seroprevalence for AAV6 amongst 56 individuals tested at 1:20 serum dilution in the AAV neutralisation assay [223] and another study reported 64 % seroprevalence amongst 11 individuals at 1:10 serum dilution [438]. Taken together, it is thus highly likely that a significant proportion of the adult population has some degree of pre-existing neutralising antibody to rAAV6.2, potentially rendering them ineligible for rAAV clinical gene therapy without further intervention.

The gene therapy field has developed several workaround strategies for pre-existing immunity. In the NHP model, cleavage of circulating IgG with the endopeptidase Imlifidase *in vivo* allowed for efficient vector administration to the liver (IV) despite moderate titres of AAV neutralising antibodies [225], an approach which could be adapted to the lung by direct lung administration. Further, plasmapheresis is an approach already used safely clinically to physically remove circulating antibodies with clinical benefit [439]. In the NHP model, plasmapheresis strongly (100-fold) decreased AAV neutralising antibody titres and allowed for efficient IV vector administration [440]. Analogously, in another NHP study, plasmapheresis allowed for transgene expression of rAAV delivery at similar levels compared to naive animals [441, 442]. Further, a plasmapheresis column allowing for specific depletion of AAV neutralising antibody is being investigated to decrease the risk of infection [440]. For lung administration, plasmapheresis could be either combined with Imlifidase lung treatment, or whole lung lavage to deplete antibodies in the BALF, as applied in a patient with autoimmune PAP (whole lung lavage and plasmapheresis [443]). Lastly, capsid engineering is a possible strategy to escape pre-existing immunity. The

AAV6.2FF serotype showed 10-fold increased resistance to pooled serum containing neutralising antibodies in comparison with AAV6 [310], indicating that AAV6.2FF could be an alternative capsid choice for AAV6.2 to partially circumvent pre-existing immunity as both AAV6.2FF and AAV6.2 showed similar transduction levels in PCLS and human LBOs (Table 4.2). Pre-existing immunity can also be evaded by isolation of serologically-distinct engineered variants by directed evolution, as shown in HAE cultures [315] which could be adapted to human models of the lung parenchyma, such as LBOs (discussed in Section 6.1.1, above).

Proposed rAAV gene therapies for SP-B deficiency are likely to be administered within the first months of life, therefore pre-existing immunity due to previous AAV exposure is less of an issue. However, AAV neutralising antibodies as well as immune cells are transferred to the foetus via the placenta in the third trimester, and postnatally via breastmilk [444]. The quantity of transferred antibodies varies, depending mainly on maternal antibody levels [444, 445], that are usually similar in mothers at birth and in full-term neonates, with an estimated antibody life-span of 9.9-17.4 weeks (half-life 7.0-12.1 weeks) amongst common viruses [445]. As neonatal antibody production is only significant after 15 weeks [446], there might be a window for rAAV vector dosing in non-breastfed neonates after maternal antibodies have waned and before the start of antibody production/wildtype AAV exposure. Overall, these studies suggest fewer antibodies from pre-existing immunity are expected in the first months of life compared with rAAV dosing in adults. Indeed, in clinical practice with Zolgensma, where the median patient age was 4.8 months, patient exclusion due to pre-existing immunity was uncommon (5.6 %, [447]).

Vector dosing in neonates could have another advantage by reducing the potential for the occasional severe side effects observed after high doses of rAAV (Section 1.5.3). In a clinical trial leading to the approval of Zolgensma, dosing of 2E14 GC/kg in infants with an

average age of 3.4 months reported no severe adverse effects [236] compared with a trial including patients up to 5 years of age where the 3E14 GC/kg dose (the highest rAAV dose ever applied) led to three deaths ([235], NCT03199469). Interestingly, serial intratracheal rAAV administration led to increased levels of neutralising antibodies without significant decline of transgene expression in neonatal ferrets (but not juvenile ferrets), likely because the underdeveloped immune system developed a tolerised immunological state [448].

Immune suppressants are commonly administered at the time of vector dosing to decrease humoral and cellular immunity to transgenes and vectors [437, 449]. To refine non-specific immunosuppression, Meliani and colleagues coadministered polylactic acid nanoparticles containing rapamycin together with rAAV vectors IV which led to dramatically suppressed rAAV humoral and cellular responses in mice and NHPs [450].

The tendency of rAAV to lose expression in dividing cells over time (Section 1.5.3) suggests rAAV readministration may be required. However, rAAV dosing mediates high and long-lasting levels of anti-capsid immunity (12-15 years [230]), making rAAV repeat administration a challenging endeavour [437]. Immunosuppression regimens and approaches discussed above to deplete AAV neutralising antibodies are also applicable for facilitating vector repeat administration and are in the process of clinical investigation. Indeed, both the rapamycin nanoparticles and Imlifidase approaches allowed for efficient IV vector repeat administration in the NHP model and results of a clinical trial testing repeat administration in an immunomodulatory regimen involving rapamycin and the B cell-depleting antibody rituximab are currently awaited (NCT02240407). In the lung, with high levels of host defence (Section 1.4.2), vector rAAV readministration might be particularly tricky, potentially making persistent expression a challenge for rAAV lung gene therapy (discussed in Section 6.2.2, below). This viewpoint is challenged by a 2003 study performing serial administration to the NHP lung using bronchoscopic micro-spraying.

High antibody titres were observed, after serial administration, but priming of the immune system with two doses of rAAV2.CFTR at 5 and 3 weeks before rAAV2.EGFP dosing still allowed for EGFP reporter gene expression. However, there was evidence of tissue inflammation and no control animal to assess if and by how much reporter gene expression was reduced by immune system priming [451]. Finally, as a method to escape a primed immune system, immunologically distinct capsid panels for rAAV readministration could be designed to allow for administration to the same cell type.

6.2.2 Longevity of rAAV vector expression in the lung

One of the key limitations of rAAV is a tendency to lose transgene expression over time in dividing cell populations, due to the episomal nature of the genome (Section 1.5.2). Two factors are crucial to longevity of rAAV therapeutic effect in SP-B efficiency: the turnover of the target ATII cells and the ability of the vector to target the stem/progenitor cell population(s). The current model is that ATII cells constitute the stem cells of the lung parenchyma [358], but that they are a heterogenous cell population, including a rare, long-lived, self-renewing ATII cell subtype(s) that give rise to ATI cells [452-454]. ATII stem cell heterogeneity was also verified in normal adult human lung [455]. Within ATII cells (identified by human ATII marker HTII-280 [359]) isolated from normal adult human lungs rejected for transplantation, a subpopulation (of alveolar epithelial progenitors) with conserved TM4SF1 marker expression was identified [455]. Compared with the ATII cell population, this sub-population comprising ca. 20 % of ATII cells showed higher numbers of colony formation of ATI and ATII cell-containing organoids [455] and was enriched in progenitor cell regulators including the Wnt pathway implicated in self-renewal [455, 456]; however, further experiments are needed to understand ATII cell subpopulations and their role in lung homeostasis. A potential experiment to address expression longevity would be to transduce the human LBO and PCLS models with vectors expressing EGFP transgene, followed by flow cytometry for the EGFP-positive cells within the HTII-280/TM4SF1

double-positive cell population, to investigate if lung gene therapy vectors can transduce this stem cell population. Interestingly, the alveolar epithelial progenitor cell population was also observed in mice [455], but several expression markers are not shared in humans [456], further highlighting the limits in using mice as an alveolar stem cell model.

Cell turnover is considered remarkably slow in normal lung especially compared with other organs such as skin and intestine [457, 458]. In the murine lung, the half-life of lung ciliated cells was found to be 17 months, but the authors noted that other lung epithelial cells do not necessarily have the same lifespan [459]. Following intranasal instillation of rAAV5 to the adult murine lung, luciferase transgene expression lasted at least 12 months with levels slowly declining (ca. 50 % remained at 12 months) but it was unclear if transgene expression originated from ATII cells or other cell types [218]. In the developing murine lung, transgene expression from rAAV5 was still observed at 9 months after dosing *in utero*, however, the majority of transduced cells were alveolar macrophages [341]. Most relevant is that expression from rAAV6.2FF lasted up to 210 days (measured by reporter gene) and also mediated long-term survival (median 205 days) in pups of the mouse model of SP-B deficiency [161]. Importantly, the observed therapeutic effects eventually ceased, potentially due to ATII cell turnover and the dilution of the rAAV episome from the daughter cells [161]. Overall, these data indicate a slow turnover of transduced murine ATII cells.

Strategies for further increasing the duration of rAAV expression, such as using CpG-free ITRs [129] and/or dosing with immunosuppressants ([161], Section 6.2) have not yet been explored in the lung. When the potential of rAAV5 (as well as rAAV1) to transduce ATII progenitor cells was investigated in the murine lung, transduction followed by induction of lung injury resulted in re-population of the lung parenchyma with patches of transgene-expressing cells surrounding stem cells, indicating the ability of rAAV5 to transduce murine ATII progenitor cells [460]. This highlights the possibility of long-term expression mediated

by rAAV transduction of ATII stem cells, as long as the episomal transgene is not 'diluted out' from the progenitor cell. However, stem cell homeostasis is likely to differ between species and cell turnover appears lower in pathogen-free housed laboratory mice used in animal experiments [461]; this prompts caution in extrapolating murine data to the human lung.

When rAAV transgene expression was investigated in human LBOs, a model of the developing human lung, expression lasted at least 9 weeks (Figure 4.10), however, the LBO model is not yet characterised regarding the similarity, or otherwise, of the stem cell population(s) and proliferation rate(s) to human neonatal lung. In normal adult lung, less than 1 % of the airway cells was found to be actively proliferating [462], and in future, lung organoids may help to improve our understanding of human lung homeostasis and development [463].

How the findings discussed above will translate to the developing lung of human neonates remains unclear, as we do not fully understand the role of the ATII cell stem cell subpopulation in human lung, the exact ATII cell turnover rate in humans, and how these might be influenced by continued lung growth and maturation in neonates. As a starting point, longevity of therapeutic transgene expression could be investigated in models with lungs more similar to human neonates, such as neonatal NHPs or piglets. Piglets might be suitable as they have been used as a model for ventilation of surfactant-deficient neonates [464, 465]. With the current level of research, one can only guess that rAAV-mediated expression in human ATII cells in neonates is likely to last for many months rather than for many years, but even months of therapeutic gene expression could be beneficial, by prolonging the time in which lung transplantation could occur. If the duration of transgene expression in the target cell population from a single dose of rAAV is insufficient, rAAV vector readministration may be considered (discussed in Section 6.2,

above). Alternatively, the possibility of using an integrating vector capable of transducing the ATII cell stem cell population (Section 6.2.2.1, below), or the use of gene editing of ATII cells (as attempted in Figure 5.12, Section 6.2.2.2, below), could be explored.

6.2.2.1 Integrating lentivirus as an alternative vector choice

To improve duration of expression and avoid loss of (non-integrated) episomes, integrating vectors, such as those based on rLV, could provide an alternative to rAAV vectors if they can transduce and integrate into the ATII stem cell population. The rLV vector rSIV.F/HN (Section 1.3.3) showed high transduction rates in murine lung, including ATII cells, as well as in HAE cultures; and preparations for a clinical trials using rSIV.F/HN as treatment for cystic fibrosis are underway [7]. Advantageously, much of the preclinical data on safety and toxicity (including genotoxicity) in animal toxicology studies, could perhaps be supportive of a gene therapy clinical trial for SP-B deficiency. Studies presented in Chapter 4 show that rSIV.F/HN did not transduce human LBOs (Figure 4.8), but did transduce human adult PCLS (Table 4.1). Further, rSIV.F/HN showed transduction and SP-B expression in a model of the SP-B deficient lung generated from the human adenocarcinoma cell line H441 grown in air-liquid interface [344]. Possibly, the LBO model was too immature to express the required levels of rSIV.F/HN entry receptors; it is also possible that this is a feature of the developing lung, which might therefore restrict transduction of this vector in neonatal as opposed to adult lung. Experiments using PCLS from human neonatal lung tissue could help to understand rSIV.F/HN vector transduction in the neonatal SP-B deficient patient population.

6.2.2.2 Gene editing approaches

Gene editing of mutations in the *SFTPB* gene in the ATII stem cell population is another approach that could help the therapeutic effects of gene therapy to persist. In Chapter 5, studies were presented investigating gene editing in human LBOs. Transfection of RNPs

and ssDNA donor complexed in lipid nanoparticles for HDR gene editing (Figure 5.12), did not result in correction of the 121ins2 mutation. Alternative delivery of Cas9 gene editing tools, such as the delivery of donor DNA with rAAV6 in human ESC have been shown to improve editing rate [394]. Analogously, delivery of the donor DNA with rAAV6.2 could improve editing frequency in the lung. As prolonged and high Cas9 expression levels include the increased risk for off-target editing [397, 466], rAAV vectors are less suitable for Cas9 delivery, however, RNP and mRNA approaches and development of nanoparticles for improved and more targeted delivery are attractive areas of research. Regardless of whether rAAV is used, methods to evade pre-existing immunity (Section 6.2) might be useful for Cas9 dosing, as high levels (58 %) of seroprevalence have been reported of the population, showing immunity against Cas9 from *Streptococcus pyogenes* [467, 468]. Lastly, methods to improve HDR editing rates by cell cycle synchronisation, or suppression of the NHEJ pathway, are being investigated [469].

Alternative editing strategies (summarised in Figure 6.1) could also be considered in order to boost efficiency. Amongst the more active DSB repair NHEJ pathways, classic NHEJ lacks the required accuracy for therapeutic gene editing, but can be used to create insertions via homology-independent targeted integration (HITI). Advantageously, HITI is active throughout all cell cycle phases, allowing editing of non-dividing cells [470, 471]. When using the HITI strategy for SP-B deficiency, the entire SP-B cDNA could be inserted after the *SFTPB* promoter restoring endogenous expression, and allowing for permanent SP-B expression independent of the specific *SFTPB* mutation.

It is still very early days in the field of translational gene editing with only a small number of human studies having been conducted. As a relatively safe entry point, three trials investigated CRISPR-based cell therapy, to avoid T cell fatigue of transferred T cells in cancer immunotherapies [149, 472], and treat a patient with HIV infection and acute

lymphoblastic leukaemia by transfer of edited, HIV-resistant hematopoietic stem and progenitor cells [473]. T cell editing for immunotherapy was considered generally safe in these trials (although no long-term data yet), with low rates of off-target editing, and clinically feasible despite still many opportunities for technical improvements [474]. First *in vivo* gene editing clinical trials are already underway to treat e.g. a congenital form of blindness (NCT03872479) or even systemic Cas9 delivery for a neuropathy ([475], NCT04601051). Next generation gene editing approaches such as prime editing might further improve the safety profile of therapeutic gene editing in the future [476].

As observed for gene delivery, human therapeutic gene editing has started with relatively simpler *ex vivo* approaches, and it will likely be several years until effective gene editing becomes a reality in the lung. Therapeutic gene editing is certainly exciting scientifically and translationally for applications discussed above and for gain-of-function mutations such as mutations in *SFTPC* (Section 1.1). However, in case of loss-of-function mutations with urgent patient need, such as SP-B deficiency, the focus should perhaps remain on the relatively more straightforward gene addition approaches.

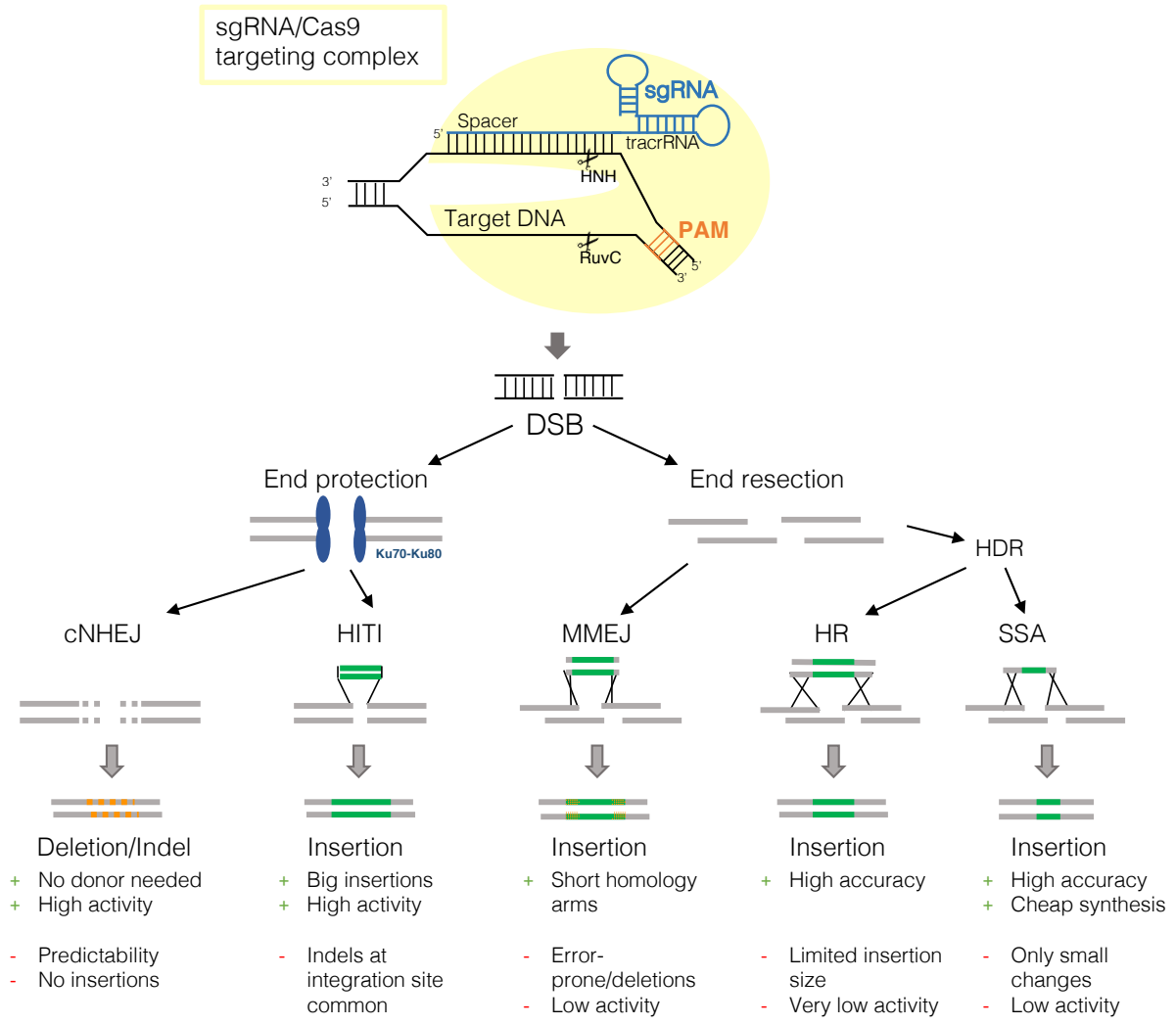


Figure 6.1 Schematic of CRISPR/Cas9 editing and DSB outcomes

Schematic of the sgRNA/Cas9 RNP complex bound to target DNA. The sgRNA consists of a tracrRNA component and spacer component for precise DNA targeting. The Cas9 nuclease domains HNH and RuvC induce a blunt end DSB. The cellular response is either 'end-protection' or 'end-resection'. DNA editing strategies are listed below together with their advantages (+) and disadvantages (-). Amongst the presented approaches, HDR and HITI are amongst the more suitable for therapeutic gene editing. The HDR pathway uses a template for repair, either the sister chromatid or in gene editing donor DNA. The initial end resection in the HDR pathway exposes long homologies between the template and the resected target DNA allowing for donor alignment and DNA strand synthesis complementary to the donor [398, 477]. The high-fidelity HDR pathway can be further divided in HR pathway using dsDNA donors and SSA pathway using ssDNA templates [477]. When using HITI as a gene editing approach, both the double stranded

DNA donor and the target DNA are cut. Depending on the direction of insertion, the sgRNA cut site is retained or disrupted, highly favouring one direction over the other instead of stochastic integration [470]. Modified from [398, 477, 478]. cNHEJ, classical non-homologous end joining; HITI, homology-independent targeted integration; MMEJ, microhomology-mediated end joining; HDR, homology directed repair; HR, homologues recombination; SSA, single stand annealing.

6.3 Acceleration of gene therapy development

Gene therapy applications are being developed to treat genetic diseases with often fatal outcomes; many of these are orphan diseases. One of the lessons learnt in the gene therapy field is that selection of the models in which to investigate and develop gene therapy approaches is key, with the 'wrong' model generating misleading results in terms of efficacy (Section 1.4.3) and safety (Section 1.5.3). In the absence of an ideal model, requirements for selection of a suitable model should be clearly defined for the research question under consideration, likely leading to the selection of different models to address different questions. This requires understanding of the factors involved (e.g. vector entry receptors) and a thorough characterisation of the model. Furthermore, differentiation between multiple models in a similar category would be helpful, such as in the case of distal phenotype lung organoid generation and HAE cultured at the air-liquid interface, where there are many (at times very similar) protocols published, resulting in difficulties in assessing studies from different laboratories and necessitating tedious comparability studies. For instance, in the selection of AAV serotypes in HAE cultures, amongst AAV1, AAV6 and AAV6.2, different candidate serotypes were selected by different groups [283, 479], with authors noting variability in donors and HAE polarisation status as reasons for contrary results [479]. Analogously, many similar but different protocols for lung organoids of the lung parenchyma are published [325, 333, 407, 409, 410, 413]. Ideally, vector lead candidate selection should be supported by a panel of different models (Table 6.1). Additionally, better comparability of titres and rAAV production methods, including improved quality control, might lead to better comparability of results between laboratories [480].

Rare genetic diseases vary in their severity and prevalence. More than 98 % of individuals affected by a rare disease, are affected by one of only 390 rare diseases with a prevalence

of between 1:100 000 and 1:2 000 [1]. This shows that developing treatments for 'only' 390 rare diseases, although this represents only a small fraction of the known 6 000 – 10 000 rare diseases, would have huge impact, helping to break down the task of treatment development for a dauntingly high number of diseases to a more accomplishable mid- to long-term goal.

To further increase the impact of gene therapy development, well-characterised diseases with high severity (e.g. SP-B deficiency) and more common, rare diseases (e.g. cystic fibrosis) could initiate the development of 'platforms' for gene therapy approaches, taking advantage of the modularity of vector design. A recent demonstration of the platform approach identified four rare diseases to be treated with rAAV9, aiming to share preclinical biodistribution and toxicology data; the authors hypothesized that this platform approach will improve the efficiency of clinical development, essentially speeding up the traditional 'one-disease-at-a-time' format for rare diseases [481]. Analogously, once an efficient vector has been developed for SP-B deficiency, it could be applied to other genetic diseases affecting ATII cells, such as SP-C deficiency, ABCA3 deficiency, hereditary PAP (Section 1.1) or alveolar microlithiasis (OMIM #265100). This provides the opportunity not only to share key animal model data (e.g. toxicology, biodistribution, etc), but also to capitalise on data in common (use of ATII cell-specific promoters) and how to deliver vector to neonates (dose, formulation, catheter delivery, etc). Success with rare diseases can also be used as a foundation or steppingstone towards developing gene therapies to tackle more complex diseases. In the case of ATII cells, there is evidence to suggest that IPF (Section 1.1) also originates in this cell type particularly its stem cell subpopulation [482]. Thus, rather than delivering a therapeutic gene to compensate for a mutated gene as in gene therapy for monogenic disorders, gene therapy could be used to regulate pathways involved in inflammation and lung repair to lessen symptoms [483].

With the recent licencing of several gene therapies for clinical use (Section 1.1), frameworks for regulatory approval and reimbursement models are likely to become more streamlined. Specific regulatory guidance for the translation of gene therapies have been developed by medicines agencies such as the European Medicines Agency, the US authority FDA, and the Medicines and Healthcare Products Regulatory Agency in the UK (EMA/CAT/80183/2014, FDA-2013-D-0576), which allows for faster translation of gene therapies. Just as important for acceleration of the development and access to gene therapies for the many rare diseases are the education of healthcare staff, access to (early) genome sequencing and the building of networks to identify, recruit and treat patients; particularly for ultrarare diseases such as SP-B deficiency.

6.4 Conclusion

SP-B deficiency is an ultrarare, well-characterised genetic lung disease, with generally no treatment offer other than palliative care. The overall aim of this thesis was to investigate how a vector for rAAV gene transfer could be selected and developed as a gene therapy for SP-B deficiency. Starting in the murine lung, the rAAV5 vector was identified as the most efficient serotype for transducing target murine ATII cells *in vivo*, but when screened in human lung organoids and *ex vivo* in human lung slices this serotype did not perform well. This highlights that the mouse model, as a gold standard for gene therapy research, has only limited capacity to model rAAV viral gene transfer to the human lung. Instead, rAAV6 and its engineered variants rAAV6.2 and rAAV6.2FF showed high levels of transduction in human lung models, indicating potential for gene transfer to the human lung parenchyma to treat SP-B deficiency and other ATII cell-involving diseases. To investigate rAAV gene transfer to a human model of SP-B deficiency, LBOs were grown from a human ESC line carrying the most common mutation in SP-B deficiency. The vector rAAV6.2.CMV.SP-B allowed for expression of therapeutic SP-B protein in this human disease model, highlighting rAAV6.2 as a translational vector and organoids as a translational disease model overall.

References

1. Nguengang Wakap, S., D.M. Lambert, A. Olry, C. Rodwell, C. Gueydan, V. Lanneau, . . . A. Rath, *Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database.* Eur J Hum Genet, 2020. **28**(2): p. 165-173.
2. Haendel, M., N. Vasilevsky, D. Unni, C. Bologa, N. Harris, H. Rehm, . . . T.I. Oprea, *How many rare diseases are there?* Nat Rev Drug Discov, 2020. **19**(2): p. 77-78.
3. *EU rare diseases research factsheet*. 2020, European Commission https://ec.europa.eu/info/sites/info/files/research_and_innovation/research_by_area/documents/ec_rtd_eu-rare-diseases-research_factsheet.pdf.
4. Harari, S., E.M. Lau, Y. Tamura, V. Cottin, G. Simonneau, and M. Humbert, *Rare (pulmonary) disease day: "feeding the breath, energy for life!".* Eur Respir J, 2015. **45**(2): p. 297-300.
5. *Approved Cellular and Gene Therapy Products*. U.S. Food and Drug Administration 28/09/2021; Available from: https://www.fda.gov/vaccines-blood-biologics/cellular-gene-therapy-products/approved-cellular-and-gene-therapy-products.
6. Miah, K.M., S.C. Hyde, and D.R. Gill, *Emerging gene therapies for cystic fibrosis.* Expert Rev Respir Med, 2019. **13**(8): p. 709-725.
7. Alton, E.W., J.M. Beekman, A.C. Boyd, J. Brand, M.S. Carlon, M.M. Connolly, . . . J. Zhu, *Preparation for a first-in-man lentivirus trial in patients with cystic fibrosis.* Thorax, 2017. **72**(2): p. 137-147.
8. Whitsett, J.A., S.E. Wert, and T.E. Weaver, *Diseases of pulmonary surfactant homeostasis.* Annu Rev Pathol, 2015. **10**: p. 371-93.
9. Wert, S.E., J.A. Whitsett, and L.M. Nogee, *Genetic disorders of surfactant dysfunction.* Pediatr Dev Pathol, 2009. **12**(4): p. 253-74.
10. Nogee, L.M., D.E. de Mello, L.P. Dehner, and H.R. Colten, *Brief report: deficiency of pulmonary surfactant protein B in congenital alveolar proteinosis.* N Engl J Med, 1993. **328**(6): p. 406-10.
11. Turcu, S., E. Ashton, L. Jenkins, A. Gupta, and Q. Mok, *Genetic testing in children with surfactant dysfunction.* Arch Dis Child, 2013. **98**(7): p. 490-5.
12. Frohlich, E., A. Mercuri, S. Wu, and S. Salar-Behzadi, *Measurements of Deposition, Lung Surface Area and Lung Fluid for Simulation of Inhaled Compounds.* Front Pharmacol, 2016. **7**: p. 181.
13. Whitsett, J.A. and T.E. Weaver, *Hydrophobic surfactant proteins in lung function and disease.* N Engl J Med, 2002. **347**(26): p. 2141-8.
14. Clements, J.A., *Surface tension of lung extracts.* Proc Soc Exp Biol Med, 1957. **95**(1): p. 170-2.
15. Autilio, C. and J. Perez-Gil, *Understanding the principle biophysics concepts of pulmonary surfactant in health and disease.* Arch Dis Child Fetal Neonatal Ed, 2019. **104**(4): p. F443-F451.
16. Veldhuizen, R., K. Nag, S. Orgeig, and F. Possmayer, *The role of lipids in pulmonary surfactant.* Biochim Biophys Acta, 1998. **1408**(2-3): p. 90-108.

17. Goerke, J., *Pulmonary surfactant: functions and molecular composition.* Biochim Biophys Acta, 1998. **1408**(2-3): p. 79-89.
18. Lopez-Rodriguez, E. and J. Perez-Gil, *Structure-function relationships in pulmonary surfactant membranes: from biophysics to therapy.* Biochim Biophys Acta, 2014. **1838**(6): p. 1568-85.
19. Perez-Gil, J. and T.E. Weaver, *Pulmonary surfactant pathophysiology: current models and open questions.* Physiology (Bethesda), 2010. **25**(3): p. 132-41.
20. Postle, A.D., E.L. Heeley, and D.C. Wilton, *A comparison of the molecular species compositions of mammalian lung surfactant phospholipids.* Comp Biochem Physiol A Mol Integr Physiol, 2001. **129**(1): p. 65-73.
21. van Meer, G., D.R. Voelker, and G.W. Feigenson, *Membrane lipids: where they are and how they behave.* Nat Rev Mol Cell Biol, 2008. **9**(2): p. 112-24.
22. Sano, H. and Y. Kuroki, *The lung collectins, SP-A and SP-D, modulate pulmonary innate immunity.* Mol Immunol, 2005. **42**(3): p. 279-87.
23. Ban, N., Y. Matsumura, H. Sakai, Y. Takanezawa, M. Sasaki, H. Arai, and N. Inagaki, *ABCA3 as a lipid transporter in pulmonary surfactant biogenesis.* J Biol Chem, 2007. **282**(13): p. 9628-34.
24. Fitzgerald, M.L., R. Xavier, K.J. Haley, R. Welti, J.L. Goss, C.E. Brown, . . . M.W. Freeman, *ABCA3 inactivation in mice causes respiratory failure, loss of pulmonary surfactant, and depletion of lung phosphatidylglycerol.* Journal of Lipid Research, 2007. **48**(3): p. 621-632.
25. Malacrida, L., S. Astrada, A. Briva, M. Bollati-Fogolin, E. Gratton, and L.A. Bagatolli, *Spectral phasor analysis of LAURDAN fluorescence in live A549 lung cells to study the hydration and time evolution of intracellular lamellar body-like structures.* Biochim Biophys Acta, 2016. **1858**(11): p. 2625-2635.
26. Martinez-Calle, M., M. Prieto, B. Olmeda, A. Fedorov, L.M.S. Loura, and J. Perez-Gil, *Pulmonary surfactant protein SP-B nanorings induce the multilamellar organization of surfactant complexes.* Biochim Biophys Acta Biomembr, 2020. **1862**(6): p. 183216.
27. Clark, J.C., S.E. Wert, C.J. Bachurski, M.T. Stahlman, B.R. Stripp, T.E. Weaver, and J.A. Whitsett, *Targeted disruption of the surfactant protein B gene disrupts surfactant homeostasis, causing respiratory failure in newborn mice.* Proc Natl Acad Sci U S A, 1995. **92**(17): p. 7794-8.
28. Ashino, Y., X. Ying, L.G. Dobbs, and J. Bhattacharya, *[Ca(2+)](i) oscillations regulate type II cell exocytosis in the pulmonary alveolus.* Am J Physiol Lung Cell Mol Physiol, 2000. **279**(1): p. L5-13.
29. Patel, A.S., D. Reigada, C.H. Mitchell, S.R. Bates, S.S. Margulies, and M. Koval, *Paracrine stimulation of surfactant secretion by extracellular ATP in response to mechanical deformation.* Am J Physiol Lung Cell Mol Physiol, 2005. **289**(3): p. L489-96.
30. Martinez-Calle, M., B. Olmeda, P. Dietl, M. Frick, and J. Perez-Gil, *Pulmonary surfactant protein SP-B promotes exocytosis of lamellar bodies in alveolar type II cells.* FASEB J, 2018. **32**(8): p. 4600-4611.
31. Olmeda, B., B. Garcia-Alvarez, M.J. Gomez, M. Martinez-Calle, A. Cruz, and J. Perez-Gil, *A model for the structure and mechanism of action of pulmonary surfactant protein B.* FASEB J, 2015. **29**(10): p. 4236-47.
32. Liekkinen, J., G. Enkavi, M. Javanainen, B. Olmeda, J. Perez-Gil, and I. Vattulainen, *Pulmonary Surfactant Lipid Reorganization Induced by the Adsorption of the Oligomeric Surfactant Protein B Complex.* J Mol Biol, 2020.

33. Williams, M.C., *Conversion of lamellar body membranes into tubular myelin in alveoli of fetal rat lungs.* J Cell Biol, 1977. **72**(2): p. 260-77.
34. Voorhout, W.F., T. Veenendaal, H.P. Haagsman, A.J. Verkleij, L.M. van Golde, and H.J. Geuze, *Surfactant protein A is localized at the corners of the pulmonary tubular myelin lattice.* J Histochem Cytochem, 1991. **39**(10): p. 1331-6.
35. Ikegami, M., T.R. Korfhagen, J.A. Whitsett, M.D. Bruno, S.E. Wert, K. Wada, and A.H. Jobe, *Characteristics of surfactant from SP-A-deficient mice.* Am J Physiol, 1998. **275**(2 Pt 1): p. L247-54.
36. Wright, J.R., *Clearance and recycling of pulmonary surfactant.* Am J Physiol, 1990. **259**(2 Pt 1): p. L1-12.
37. Dranoff, G., A.D. Crawford, M. Sadelain, B. Ream, A. Rashid, R.T. Bronson, . . . et al., *Involvement of granulocyte-macrophage colony-stimulating factor in pulmonary homeostasis.* Science, 1994. **264**(5159): p. 713-6.
38. Roldan, N., T.K.M. Nyholm, J.P. Slotte, J. Perez-Gil, and B. Garcia-Alvarez, *Effect of Lung Surfactant Protein SP-C and SP-C-Promoted Membrane Fragmentation on Cholesterol Dynamics.* Biophys J, 2016. **111**(8): p. 1703-1713.
39. Ikegami, M., C.L. Na, T.R. Korfhagen, and J.A. Whitsett, *Surfactant protein D influences surfactant ultrastructure and uptake by alveolar type II cells.* Am J Physiol Lung Cell Mol Physiol, 2005. **288**(3): p. L552-61.
40. Korfhagen, T.R., V. Sheftelyevich, M.S. Burhans, M.D. Bruno, G.F. Ross, S.E. Wert, . . . J.H. Fisher, *Surfactant protein-D regulates surfactant phospholipid homeostasis in vivo.* Journal of Biological Chemistry, 1998. **273**(43): p. 28438-28443.
41. Bates, S.R., C. Dodia, J.Q. Tao, and A.B. Fisher, *Surfactant protein-A plays an important role in lung surfactant clearance: evidence using the surfactant protein-A gene-targeted mouse.* Am J Physiol Lung Cell Mol Physiol, 2008. **294**(2): p. L325-33.
42. Lin, S., C.L. Na, H.T. Akinbi, K.S. Apsley, J.A. Whitsett, and T.E. Weaver, *Surfactant protein B (SP-B) -/- mice are rescued by restoration of SP-B expression in alveolar type II cells but not Clara cells.* J Biol Chem, 1999. **274**(27): p. 19168-74.
43. Brasch, F., G. Johnen, A. Winn-Brasch, S.H. Guttentag, A. Schmiedl, N. Kapp, . . . M. Ochs, *Surfactant protein B in type II pneumocytes and intra-alveolar surfactant forms of human lungs.* Am J Respir Cell Mol Biol, 2004. **30**(4): p. 449-58.
44. Yang, L., J. Johansson, R. Ridsdale, H. Willander, M. Fitzen, H.T. Akinbi, and T.E. Weaver, *Surfactant protein B propeptide contains a saposin-like protein domain with antimicrobial activity at low pH.* J Immunol, 2010. **184**(2): p. 975-83.
45. Lin, S., H.T. Akinbi, J.S. Breslin, and T.E. Weaver, *Structural requirements for targeting of surfactant protein B (SP-B) to secretory granules in vitro and in vivo.* J Biol Chem, 1996. **271**(33): p. 19689-95.
46. Korimilli, A., L.W. Gonzales, and S.H. Guttentag, *Intracellular localization of processing events in human surfactant protein B biosynthesis.* J Biol Chem, 2000. **275**(12): p. 8672-9.
47. Voorhout, W.F., T. Veenendaal, H.P. Haagsman, T.E. Weaver, J.A. Whitsett, L.M. van Golde, and H.J. Geuze, *Intracellular processing of pulmonary*

surfactant protein B in an endosomal/lysosomal compartment. Am J Physiol, 1992. **263**(4 Pt 1): p. L479-86.
48. Olmeda, B., B. Garcia-Alvarez, and J. Perez-Gil, *Structure-function correlations of pulmonary surfactant protein SP-B and the saposin-like family of proteins.* Eur Biophys J, 2013. **42**(2-3): p. 209-22.
49. Perez-Gil, J., A. Cruz, and C. Casals, *Solubility of hydrophobic surfactant proteins in organic solvent/water mixtures. Structural studies on SP-B and SP-C in aqueous organic solvents and lipids.* Biochim Biophys Acta, 1993. **1168**(3): p. 261-70.
50. Zaltash, S., M. Palmblad, T. Curstedt, J. Johansson, and B. Persson, *Pulmonary surfactant protein B: a structural model and a functional analogue.* Biochim Biophys Acta, 2000. **1466**(1-2): p. 179-86.
51. Beck, D.C., M. Ikegami, C.L. Na, S. Zaltash, J. Johansson, J.A. Whitsett, and T.E. Weaver, *The role of homodimers in surfactant protein B function in vivo.* J Biol Chem, 2000. **275**(5): p. 3365-70.
52. Thurm, T., E. Kaltenborn, S. Kern, M. Griese, and R. Zarbock, *SFTPC mutations cause SP-C degradation and aggregate formation without increasing ER stress.* Eur J Clin Invest, 2013. **43**(8): p. 791-800.
53. Glasser, S.W., E.A. Detmer, M. Ikegami, C.L. Na, M.T. Stahlman, and J.A. Whitsett, *Pneumonitis and emphysema in sp-C gene targeted mice.* J Biol Chem, 2003. **278**(16): p. 14291-8.
54. Glasser, S.W., T.L. Witt, A.P. Senft, J.E. Baatz, D. Folger, M.D. Maxfield, . . . T.R. Korfhagen, *Surfactant protein C-deficient mice are susceptible to respiratory syncytial virus infection.* Am J Physiol Lung Cell Mol Physiol, 2009. **297**(1): p. L64-72.
55. Vorbroker, D.K., S.A. Profitt, L.M. Nogee, and J.A. Whitsett, *Aberrant processing of surfactant protein C in hereditary SP-B deficiency.* Am J Physiol, 1995. **268**(4 Pt 1): p. L647-56.
56. Olmeda, B., M. Martinez-Calle, and J. Perez-Gil, *Pulmonary surfactant metabolism in the alveolar airspace: Biogenesis, extracellular conversions, recycling.* Ann Anat, 2017. **209**: p. 78-92.
57. Cabre, E.J., M. Martinez-Calle, M. Prieto, A. Fedorov, B. Olmeda, L.M.S. Loura, and J. Perez-Gil, *Homo- and hetero-oligomerization of hydrophobic pulmonary surfactant proteins SP-B and SP-C in surfactant phospholipid membranes.* J Biol Chem, 2018. **293**(24): p. 9399-9411.
58. Hobi, N., M. Giolai, B. Olmeda, P. Miklavc, E. Felder, P. Walther, . . . T. Haller, *A small key unlocks a heavy door: The essential function of the small hydrophobic proteins SP-B and SP-C to trigger adsorption of pulmonary surfactant lamellar bodies.* Biochim Biophys Acta, 2016. **1863**(8): p. 2124-34.
59. Cheong, N., H. Zhang, M. Madesh, M. Zhao, K. Yu, C. Dodia, . . . H. Shuman, *ABCA3 is critical for lamellar body biogenesis in vivo.* J Biol Chem, 2007. **282**(33): p. 23811-7.
60. Nogee, L.M., *Genetics of the hydrophobic surfactant proteins.* Biochim Biophys Acta, 1998. **1408**(2-3): p. 323-33.
61. Whitsett, J.A., L.M. Nogee, T.E. Weaver, and A.D. Horowitz, *Human surfactant protein B: structure, function, regulation, and genetic disease.* Physiol Rev, 1995. **75**(4): p. 749-57.
62. Guttentag, S.H., M.F. Beers, B.M. Bieler, and P.L. Ballard, *Surfactant protein B processing in human fetal lung.* Am J Physiol, 1998. **275**(3): p. L559-66.

63. Ueno, T., S. Linder, C.L. Na, W.R. Rice, J. Johansson, and T.E. Weaver, *Processing of pulmonary surfactant protein B by napsin and cathepsin H.* J Biol Chem, 2004. **279**(16): p. 16178-84.
64. Guttentag, S., *Posttranslational regulation of surfactant protein B expression.* Semin Perinatol, 2008. **32**(5): p. 367-70.
65. Gerson, K.D., C.D. Foster, P. Zhang, Z. Zhang, M.M. Rosenblatt, and S.H. Guttentag, *Pepsinogen C proteolytic processing of surfactant protein B.* J Biol Chem, 2008. **283**(16): p. 10330-8.
66. Johansson, J. and T. Curstedt, *Molecular structures and interactions of pulmonary surfactant components.* Eur J Biochem, 1997. **244**(3): p. 675-93.
67. Cole, F.S., A. Hamvas, P. Rubinstein, E. King, M. Trusgnich, L.M. Nogee, . . . H.R. Colten, *Population-based estimates of surfactant protein B deficiency.* Pediatrics, 2000. **105**(3 Pt 1): p. 538-41.
68. Garmany, T.H., J.A. Wambach, H.B. Heins, J.M. Watkins-Torry, D.J. Wegner, K. Bennet, . . . A. Hamvas, *Population and disease-based prevalence of the common mutations associated with surfactant deficiency.* Pediatr Res, 2008. **63**(6): p. 645-9.
69. Hamvas, A., L.M. Nogee, D.E. deMello, and F.S. Cole, *Pathophysiology and treatment of surfactant protein-B deficiency.* Biol Neonate, 1995. **67 Suppl 1**: p. 18-31.
70. Kurath-Koller, S., B. Resch, R. Kraschl, C. Windpassinger, and E. Eber, *Surfactant Protein B Deficiency Caused by Homozygous C248X Mutation-A Case Report and Review of the Literature.* AJP Rep, 2015. **5**(1): p. e53-9.
71. Yusen, R.D., A.H. Cohen, and A. Hamvas, *Normal lung function in subjects heterozygous for surfactant protein-B deficiency.* Am J Respir Crit Care Med, 1999. **159**(2): p. 411-4.
72. Nogee, L.M., *Alterations in SP-B and SP-C expression in neonatal lung disease.* Annu Rev Physiol, 2004. **66**: p. 601-23.
73. Nogee, L.M., G. Garnier, H.C. Dietz, L. Singer, A.M. Murphy, D.E. deMello, and H.R. Colten, *A mutation in the surfactant protein B gene responsible for fatal neonatal respiratory disease in multiple kindreds.* J Clin Invest, 1994. **93**(4): p. 1860-3.
74. Ballard, P.L., L.M. Nogee, M.F. Beers, R.A. Ballard, B.C. Planer, L. Polk, . . . W.J. Longmore, *Partial deficiency of surfactant protein B in an infant with chronic lung disease.* Pediatrics, 1995. **96**(6): p. 1046-52.
75. Lopez-Andreu, J.A., A.D. Hidalgo-Santos, M.A. Fuentes-Castello, N. Mancheno-Franch, J.A. Ceron-Perez, M.J. Esteban-Ricos, . . . L.M. Nogee, *Delayed Presentation and Prolonged Survival of a Child with Surfactant Protein B Deficiency.* J Pediatr, 2017. **190**: p. 268-270 e1.
76. Nogee, L.M., S.E. Wert, S.A. Proffit, W.M. Hull, and J.A. Whitsett, *Allelic heterogeneity in hereditary surfactant protein B (SP-B) deficiency.* Am J Respir Crit Care Med, 2000. **161**(3 Pt 1): p. 973-81.
77. Dunbar, A.E., 3rd, S.E. Wert, M. Ikegami, J.A. Whitsett, A. Hamvas, F.V. White, . . . L.M. Nogee, *Prolonged survival in hereditary surfactant protein B (SP-B) deficiency associated with a novel splicing mutation.* Pediatr Res, 2000. **48**(3): p. 275-82.
78. Schuerman, F.A., M. Griese, J.P. Gille, F. Brasch, L.A. Noorduyn, and A.H. van Kaam, *Surfactant protein B deficiency caused by a novel mutation involving multiple exons of the SP-B gene.* Eur J Med Res, 2008. **13**(6): p. 281-6.

79. Wegner, D.J., T. Hertzberg, H.B. Heins, G. Elmberger, M.J. MacCoss, C.S. Carlson, . . . A. Hamvas, *A major deletion in the surfactant protein-B gene causing lethal respiratory distress.* Acta Paediatr, 2007. **96**(4): p. 516-20.
80. Takci, S., D. Anuk-Ince, M. Louha, R. Couderc, N. Cakar, R.D. Koseoglu, and O. Ates, *A rare large mutation involving two exons of the SP-B gene in an infant with severe respiratory distress.* Turk J Pediatr, 2017. **59**(4): p. 483-486.
81. Hamvas, A., D.J. Wegner, C.S. Carlson, K.R. Bergmann, M.A. Trusgnich, L. Fulton, . . . F.S. Cole, *Comprehensive genetic variant discovery in the surfactant protein B gene.* Pediatr Res, 2007. **62**(2): p. 170-5.
82. Hamvas, A., L.M. Nogee, G.B. Mallory, Jr., T.L. Spray, C.B. Huddleston, A. August, . . . H.R. Colten, *Lung transplantation for treatment of infants with surfactant protein B deficiency.* J Pediatr, 1997. **130**(2): p. 231-9.
83. deMello, D.E., S. Heyman, D.S. Phelps, A. Hamvas, L. Nogee, S. Cole, and H.R. Colten, *Ultrastructure of lung in surfactant protein B deficiency.* Am J Respir Cell Mol Biol, 1994. **11**(2): p. 230-9.
84. Beers, M.F., A. Hamvas, M.A. Moxley, L.W. Gonzales, S.H. Guttentag, K.O. Solarin, . . . P.L. Ballard, *Pulmonary surfactant metabolism in infants lacking surfactant protein B.* Am J Respir Cell Mol Biol, 2000. **22**(3): p. 380-91.
85. Palomar, L.M., L.M. Nogee, S.C. Sweet, C.B. Huddleston, F.S. Cole, and A. Hamvas, *Long-term outcomes after infant lung transplantation for surfactant protein B deficiency related to other causes of respiratory failure.* J Pediatr, 2006. **149**(4): p. 548-53.
86. Friedmann, T. and R. Roblin, *Gene therapy for human genetic disease?* Science, 1972. **175**(4025): p. 949-55.
87. Cavazzana-Calvo, M., S. Hacein-Bey, G. de Saint Basile, F. Gross, E. Yvon, P. Nusbaum, . . . A. Fischer, *Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease.* Science, 2000. **288**(5466): p. 669-72.
88. Wang, F., Z. Qin, H. Lu, S. He, J. Luo, C. Jin, and X. Song, *Clinical translation of gene medicine.* J Gene Med, 2019. **21**(7): p. e3108.
89. Keeler, A.M., M.K. ElMallah, and T.R. Flotte, *Gene Therapy 2017: Progress and Future Directions.* Clin Transl Sci, 2017. **10**(4): p. 242-248.
90. *FDA approval brings first gene therapy to the United States*. 2017: FDA News Release.
91. Nathwani, A.C., U.M. Reiss, E.G. Tuddenham, C. Rosales, P. Chowdary, J. McIntosh, . . . A.M. Davidoff, *Long-term safety and efficacy of factor IX gene therapy in hemophilia B.* N Engl J Med, 2014. **371**(21): p. 1994-2004.
92. Perrin, G.Q., R.W. Herzog, and D.M. Markusic, *Update on clinical gene therapy for hemophilia.* Blood, 2019. **133**(5): p. 407-414.
93. Harrison, C., *First gene therapy for beta-thalassemia approved.* Nat Biotechnol, 2019. **37**(10): p. 1102-1103.
94. Grigor, E.J.M., D. Fergusson, N. Kekre, J. Montroy, H. Atkins, M.D. Seftel, . . . M.M. Lalu, *Risks and Benefits of Chimeric Antigen Receptor T-Cell (CAR-T) Therapy in Cancer: A Systematic Review and Meta-Analysis.* Transfus Med Rev, 2019. **33**(2): p. 98-110.
95. Powell, S.K., R. Rivera-Soto, and S.J. Gray, *Viral expression cassette elements to enhance transgene target specificity and expression in gene therapy.* Discov Med, 2015. **19**(102): p. 49-57.
96. Niwa, H., K. Yamamura, and J. Miyazaki, *Efficient selection for high-expression transfectants with a novel eukaryotic vector.* Gene, 1991. **108**(2): p. 193-9.

97. Hyde, S.C., I.A. Pringle, S. Abdullah, A.E. Lawton, L.A. Davies, A. Varathalingam, . . . D.R. Gill, *CpG-free plasmids confer reduced inflammation and sustained pulmonary gene expression.* Nat Biotechnol, 2008. **26**(5): p. 549-51.
98. Klein, R.L., E.M. Meyer, A.L. Peel, S. Zolotukhin, C. Meyers, N. Muzyczka, and M.A. King, *Neuron-specific transduction in the rat septohippocampal or nigrostriatal pathway by recombinant adeno-associated virus vectors.* Exp Neurol, 1998. **150**(2): p. 183-94.
99. Qin, J.Y., L. Zhang, K.L. Clift, I. Hulur, A.P. Xiang, B.Z. Ren, and B.T. Lahn, *Systematic comparison of constitutive promoters and the doxycycline-inducible promoter.* PLoS One, 2010. **5**(5): p. e10611.
100. Gray, S.J., S.B. Foti, J.W. Schwartz, L. Bachaboina, B. Taylor-Blake, J. Coleman, . . . R.J. Samulski, *Optimizing promoters for recombinant adeno-associated virus-mediated gene expression in the peripheral and central nervous system using self-complementary vectors.* Hum Gene Ther, 2011. **22**(9): p. 1143-53.
101. Hsu, C.C., H.P. Li, Y.H. Hung, Y.W. Leu, W.H. Wu, F.S. Wang, . . . S.H. Hsiao, *Targeted methylation of CMV and E1A viral promoters.* Biochem Biophys Res Commun, 2010. **402**(2): p. 228-34.
102. Gruh, I., S. Wunderlich, M. Winkler, K. Schwanke, J. Heinke, U. Blomer, . . . U. Martin, *Human CMV immediate-early enhancer: a useful tool to enhance cell-type-specific expression from lentiviral vectors.* J Gene Med, 2008. **10**(1): p. 21-32.
103. Kuzmin, D.A., M.V. Shutova, N.R. Johnston, O.P. Smith, V.V. Fedorin, Y.S. Kukushkin, . . . E.C. Johnstone, *The clinical landscape for AAV gene therapies.* Nat Rev Drug Discov, 2021. **20**(3): p. 173-174.
104. Gentner, B. and L. Naldini, *Exploiting microRNA regulation for genetic engineering.* Tissue Antigens, 2012. **80**(5): p. 393-403.
105. Furger, A., J.M. O'Sullivan, A. Binnie, B.A. Lee, and N.J. Proudfoot, *Promoter proximal splice sites enhance transcription.* Genes Dev, 2002. **16**(21): p. 2792-9.
106. Huang, M.T. and C.M. Gorman, *Intervening sequences increase efficiency of RNA 3' processing and accumulation of cytoplasmic RNA.* Nucleic Acids Res, 1990. **18**(4): p. 937-47.
107. Barrett, L.W., S. Fletcher, and S.D. Wilton, *Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements.* Cell Mol Life Sci, 2012. **69**(21): p. 3613-34.
108. Bianchi, M., R. Crinelli, E. Giacomini, E. Carloni, and M. Magnani, *A potent enhancer element in the 5'-UTR intron is crucial for transcriptional regulation of the human ubiquitin C gene.* Gene, 2009. **448**(1): p. 88-101.
109. Ostedgaard, L.S., T. Rokhlina, P.H. Karp, P. Lashmit, S. Afione, M. Schmidt, . . . M.J. Welsh, *A shortened adeno-associated virus expression cassette for CFTR gene transfer to cystic fibrosis airway epithelia.* Proc Natl Acad Sci U S A, 2005. **102**(8): p. 2952-7.
110. Ronzitti, G., G. Bortolussi, R. van Dijk, F. Collaud, S. Charles, C. Leborgne, . . . F. Mingozzi, *A translationally optimized AAV-UGT1A1 vector drives safe and long-lasting correction of Crigler-Najjar syndrome.* Mol Ther Methods Clin Dev, 2016. **3**: p. 16049.

111. Choi, J.H., N.K. Yu, G.C. Baek, J. Bakes, D. Seo, H.J. Nam, . . . B.K. Kaang, *Optimization of AAV expression cassettes to improve packaging capacity and transgene expression in neurons.* Mol Brain, 2014. **7**: p. 17.
112. Xu, D.H., X.Y. Wang, Y.L. Jia, T.Y. Wang, Z.W. Tian, X. Feng, and Y.N. Zhang, *SV40 intron, a potent strong intron element that effectively increases transgene expression in transfected Chinese hamster ovary cells.* J Cell Mol Med, 2018. **22**(4): p. 2231-2239.
113. Yew, N.S., D.M. Wysokenski, K.X. Wang, R.J. Ziegler, J. Marshall, D. McNeilly, . . . S.H. Cheng, *Optimization of plasmid vectors for high-level expression in lung epithelial cells.* Hum Gene Ther, 1997. **8**(5): p. 575-84.
114. Lu, J., J.A. Williams, J. Luke, F. Zhang, K. Chu, and M.A. Kay, *A 5' Noncoding Exon Containing Engineered Intron Enhances Transgene Expression from Recombinant AAV Vectors in vivo.* Hum Gene Ther, 2017. **28**(1): p. 125-134.
115. Wu, Z., J. Sun, T. Zhang, C. Yin, F. Yin, T. Van Dyke, . . . P.E. Monahan, *Optimization of self-complementary AAV vectors for liver-directed expression results in sustained correction of hemophilia B at low vector dose.* Mol Ther, 2008. **16**(2): p. 280-9.
116. Donello, J.E., J.E. Loeb, and T.J. Hope, *Woodchuck hepatitis virus contains a tripartite posttranscriptional regulatory element.* J Virol, 1998. **72**(6): p. 5085-92.
117. Huang, Z.M. and T.S. Yen, *Hepatitis B virus RNA element that facilitates accumulation of surface gene transcripts in the cytoplasm.* J Virol, 1994. **68**(5): p. 3193-9.
118. Huang, Z.M. and T.S. Yen, *Role of the hepatitis B virus posttranscriptional regulatory element in export of intronless transcripts.* Mol Cell Biol, 1995. **15**(7): p. 3864-9.
119. Ramezani, A., T.S. Hawley, and R.G. Hawley, *Lentiviral vectors for enhanced gene expression in human hematopoietic cells.* Mol Ther, 2000. **2**(5): p. 458-69.
120. Fagoe, N.D., R. Eggers, J. Verhaagen, and M.R. Mason, *A compact dual promoter adeno-associated viral vector for efficient delivery of two genes to dorsal root ganglion neurons.* Gene Ther, 2014. **21**(3): p. 242-52.
121. Patricio, M.I., A.R. Barnard, H.O. Orlans, M.E. McClements, and R.E. MacLaren, *Inclusion of the Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element Enhances AAV2-Driven Transduction of Mouse and Human Retina.* Mol Ther Nucleic Acids, 2017. **6**: p. 198-208.
122. Klein, R., B. Ruttkowski, E. Knapp, B. Salmons, W.H. Gunzburg, and C. Hohenadl, *WPRE-mediated enhancement of gene expression is promoter and cell line specific.* Gene, 2006. **372**: p. 153-61.
123. Schambach, A., H. Wodrich, M. Hildinger, J. Bohne, H.G. Krausslich, and C. Baum, *Context dependence of different modules for posttranscriptional enhancement of gene expression from retroviral vectors.* Mol Ther, 2000. **2**(5): p. 435-45.
124. Ill, C.R. and H.C. Chiou, *Gene therapy progress and prospects: recent progress in transgene and RNAi expression cassettes.* Gene Ther, 2005. **12**(10): p. 795-802.
125. Chen, Z.Y., C.Y. He, L. Meuse, and M.A. Kay, *Silencing of episomal transgene expression by plasmid bacterial DNA elements in vivo.* Gene Ther, 2004. **11**(10): p. 856-64.

126. Kako, K., M. Nishikawa, H. Yoshida, and Y. Takakura, *Effects of inflammatory response on in vivo transgene expression by plasmid DNA in mice.* J Pharm Sci, 2008. **97**(8): p. 3074-83.
127. Hodges, B.L., K.M. Taylor, M.F. Joseph, S.A. Bourgeois, and R.K. Scheule, *Long-term transgene expression from plasmid DNA gene therapy vectors is negatively affected by CpG dinucleotides.* Mol Ther, 2004. **10**(2): p. 269-78.
128. Mitsui, M., M. Nishikawa, L. Zang, M. Ando, K. Hattori, Y. Takahashi, . . . Y. Takakura, *Effect of the content of unmethylated CpG dinucleotides in plasmid DNA on the sustainability of transgene expression.* J Gene Med, 2009. **11**(5): p. 435-43.
129. Faust, S.M., P. Bell, B.J. Cutler, S.N. Ashley, Y. Zhu, J.E. Rabinowitz, and J.M. Wilson, *CpG-depleted adeno-associated virus vectors evade immune detection.* J Clin Invest, 2013. **123**(7): p. 2994-3001.
130. Bauer, A.P., D. Leikam, S. Krinner, F. Notka, C. Ludwig, G. Langst, and R. Wagner, *The impact of intragenic CpG content on gene expression.* Nucleic Acids Res, 2010. **38**(12): p. 3891-908.
131. Kosovac, D., J. Wild, C. Ludwig, S. Meissner, A.P. Bauer, and R. Wagner, *Minimal doses of a sequence-optimized transgene mediate high-level and long-term EPO expression in vivo: challenging CpG-free gene design.* Gene Ther, 2011. **18**(2): p. 189-98.
132. Krinner, S., A. Heitzer, B. Asbach, and R. Wagner, *Interplay of Promoter Usage and Intragenic CpG Content: Impact on GFP Reporter Gene Expression.* Hum Gene Ther, 2015. **26**(12): p. 826-40.
133. Mueller, W.F., L.S. Larsen, A. Garibaldi, G.W. Hatfield, and K.J. Hertel, *The Silent Sway of Splicing by Synonymous Substitutions.* J Biol Chem, 2015. **290**(46): p. 27700-11.
134. Cheng, D., P.S. MacArthur, S. Rong, J.S. Parks, and G.S. Shelness, *Alternative splicing attenuates transgenic expression directed by the apolipoprotein E promoter-enhancer based expression vector pLIV11.* J Lipid Res, 2010. **51**(4): p. 849-55.
135. Jansen, R., J.D. Embden, W. Gaastra, and L.M. Schouls, *Identification of genes that are associated with DNA repeats in prokaryotes.* Mol Microbiol, 2002. **43**(6): p. 1565-75.
136. Bolotin, A., B. Quinquis, A. Sorokin, and S.D. Ehrlich, *Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin.* Microbiology, 2005. **151**(Pt 8): p. 2551-2561.
137. Makarova, K.S., N.V. Grishin, S.A. Shabalina, Y.I. Wolf, and E.V. Koonin, *A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action.* Biol Direct, 2006. **1**: p. 7.
138. Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, and E. Charpentier, *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.
139. Mali, P., L. Yang, K.M. Esvelt, J. Aach, M. Guell, J.E. DiCarlo, . . . G.M. Church, *RNA-guided human genome engineering via Cas9.* Science, 2013. **339**(6121): p. 823-6.

140. Cong, L., F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, . . . F. Zhang, *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.
141. Jinek, M., A. East, A. Cheng, S. Lin, E. Ma, and J. Doudna, *RNA-programmed genome editing in human cells.* Elife, 2013. **2**: p. e00471.
142. Li, D., Z. Qiu, Y. Shao, Y. Chen, Y. Guan, M. Liu, . . . M. Liu, *Heritable gene targeting in the mouse and rat using a CRISPR-Cas system.* Nat Biotechnol, 2013. **31**(8): p. 681-3.
143. Chen, Y., Y. Zheng, Y. Kang, W. Yang, Y. Niu, X. Guo, . . . X.J. Li, *Functional disruption of the dystrophin gene in rhesus monkey using CRISPR/Cas9.* Hum Mol Genet, 2015. **24**(13): p. 3764-74.
144. Long, C., J.R. McAnally, J.M. Shelton, A.A. Mireault, R. Bassel-Duby, and E.N. Olson, *Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA.* Science, 2014. **345**(6201): p. 1184-1188.
145. Long, C., L. Amoasii, A.A. Mireault, J.R. McAnally, H. Li, E. Sanchez-Ortiz, . . . E.N. Olson, *Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy.* Science, 2016. **351**(6271): p. 400-3.
146. Schwank, G., B.K. Koo, V. Sasselli, J.F. Dekkers, I. Heo, T. Demircan, . . . H. Clevers, *Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients.* Cell Stem Cell, 2013. **13**(6): p. 653-8.
147. Song, C.Q., D. Wang, T. Jiang, K. O'Connor, Q. Tang, L. Cai, . . . W. Xue, *In Vivo Genome Editing Partially Restores Alpha1-Antitrypsin in a Murine Model of AAT Deficiency.* Hum Gene Ther, 2018. **29**(8): p. 853-860.
148. Jacinto, F.V., W. Link, and B.I. Ferreira, *CRISPR/Cas9-mediated genome editing: From basic research to translational medicine.* J Cell Mol Med, 2020.
149. Stadtmauer, E.A., J.A. Fraietta, M.M. Davis, A.D. Cohen, K.L. Weber, E. Lancaster, . . . C.H. June, *CRISPR-engineered T cells in patients with refractory cancer.* Science, 2020. **367**(6481).
150. Buck, J., P. Grossen, P.R. Cullis, J. Huwyler, and D. Witzigmann, *Lipid-Based DNA Therapeutics: Hallmarks of Non-Viral Gene Delivery.* ACS Nano, 2019. **13**(4): p. 3754-3782.
151. Hacein-Bey-Abina, S., C. Von Kalle, M. Schmidt, M.P. McCormack, N. Wulffraat, P. Leboulch, . . . M. Cavazzana-Calvo, *LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1.* Science, 2003. **302**(5644): p. 415-9.
152. Braun, C.J., K. Boztug, A. Paruzynski, M. Witzel, A. Schwarzer, M. Rothe, . . . C. Klein, *Gene therapy for Wiskott-Aldrich syndrome--long-term efficacy and genotoxicity.* Sci Transl Med, 2014. **6**(227): p. 227ra33.
153. Herman, S.A. and J.M. Coffin, *Efficient packaging of readthrough RNA in ALV: implications for oncogene transduction.* Science, 1987. **236**(4803): p. 845-8.
154. Dull, T., R. Zufferey, M. Kelly, R.J. Mandel, M. Nguyen, D. Trono, and L. Naldini, *A third-generation lentivirus vector with a conditional packaging system.* J Virol, 1998. **72**(11): p. 8463-71.
155. Zufferey, R., T. Dull, R.J. Mandel, A. Bukovsky, D. Quiroz, L. Naldini, and D. Trono, *Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery.* J Virol, 1998. **72**(12): p. 9873-80.
156. Kobayashi, N., N. Bagheri, J.G. Nedrud, R.M. Strieter, Y. Tomino, M.E. Lamm, and S.N. Emancipator, *Differential effects of Sendai virus infection on mediator*

synthesis by mesangial cells from two mouse strains. Kidney Int, 2003. **64**(5): p. 1675-84.
157. Wold, W.S. and K. Toth, *Adenovirus vectors for gene therapy, vaccination and cancer gene therapy.* Curr Gene Ther, 2013. **13**(6): p. 421-33.
158. Baker, A.T., C. Aguirre-Hernandez, G. Hallden, and A.L. Parker, *Designer Oncolytic Adenovirus: Coming of Age.* Cancers (Basel), 2018. **10**(6).
159. Zincarelli, C., S. Soltys, G. Rengo, and J.E. Rabinowitz, *Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection.* Mol Ther, 2008. **16**(6): p. 1073-80.
160. Majowicz, A., D. Salas, N. Zabaleta, E. Rodriguez-Garcia, G. Gonzalez-Aseguinolaza, H. Petry, and V. Ferreira, *Successful Repeated Hepatic Gene Delivery in Mice and Non-human Primates Achieved by Sequential Administration of AAV5(ch) and AAV1.* Mol Ther, 2017. **25**(8): p. 1831-1842.
161. Kang, M.H., L.P. van Lieshout, L. Xu, J.M. Domm, A. Vadivel, L. Renesme, . . . B. Thebaud, *A lung tropic AAV vector improves survival in a mouse model of surfactant B deficiency.* Nat Commun, 2020. **11**(1): p. 3929.
162. Kurosaki, F., R. Uchibori, N. Mato, Y. Sehara, Y. Saga, M. Urabe, . . . A. Kume, *Optimization of adeno-associated virus vector-mediated gene transfer to the respiratory tract.* Gene Ther, 2017. **24**(5): p. 290-297.
163. van Haasteren, J., S.C. Hyde, and D.R. Gill, *Lessons learned from lung and liver in-vivo gene therapy: implications for the future.* Expert Opin Biol Ther, 2018. **18**(9): p. 959-972.
164. Bostick, B., A. Ghosh, Y. Yue, C. Long, and D. Duan, *Systemic AAV-9 transduction in mice is influenced by animal age but not by the route of administration.* Gene Ther, 2007. **14**(22): p. 1605-9.
165. Laube, B.L., *Aerosolized Medications for Gene and Peptide Therapy.* Respir Care, 2015. **60**(6): p. 806-21; discussion 821-4.
166. Duncan, G.A., J. Jung, J. Hanes, and J.S. Suk, *The Mucus Barrier to Inhaled Gene Therapy.* Mol Ther, 2016. **24**(12): p. 2043-2053.
167. Kim, N., G.A. Duncan, J. Hanes, and J.S. Suk, *Barriers to inhaled gene therapy of obstructive lung diseases: A review.* J Control Release, 2016. **240**: p. 465-488.
168. Button, B., L.H. Cai, C. Ehre, M. Kesimer, D.B. Hill, J.K. Sheehan, . . . M. Rubinstein, *A periciliary brush promotes the lung health by separating the mucus layer from airway epithelia.* Science, 2012. **337**(6097): p. 937-41.
169. Kesimer, M., C. Ehre, K.A. Burns, C.W. Davis, J.K. Sheehan, and R.J. Pickles, *Molecular organization of the mucins and glycocalyx underlying mucus transport over mucosal surfaces of the airways.* Mucosal Immunol, 2013. **6**(2): p. 379-92.
170. Tsan, M.F., G.L. Tsan, and J.E. White, *Surfactant inhibits cationic liposome-mediated gene transfer.* Hum Gene Ther, 1997. **8**(7): p. 817-25.
171. Duncan, J.E., J.A. Whitsett, and A.D. Horowitz, *Pulmonary surfactant inhibits cationic liposome-mediated gene delivery to respiratory epithelial cells in vitro.* Hum Gene Ther, 1997. **8**(4): p. 431-8.
172. Ernst, N., S. Ulrichskotter, W.A. Schmalix, J. Radler, R. Galneder, E. Mayer, . . . J. Rosenecker, *Interaction of liposomal and polycationic transfection complexes with pulmonary surfactant.* J Gene Med, 1999. **1**(5): p. 331-40.

173. Nayak, A., E. Dodagatta-Marri, A.G. Tsolaki, and U. Kishore, *An Insight into the Diverse Roles of Surfactant Proteins, SP-A and SP-D in Innate and Adaptive Immunity.* Front Immunol, 2012. **3**: p. 131.
174. Alton, E., D.K. Armstrong, D. Ashby, K.J. Bayfield, D. Bilton, E.V. Bloomfield, . . . U.K.C.F.G.T. Consortium, *Repeated nebulisation of non-viral CFTR gene therapy in patients with cystic fibrosis: a randomised, double-blind, placebo-controlled, phase 2b trial.* Lancet Respir Med, 2015. **3**(9): p. 684-691.
175. Flotte, T., B. Carter, C. Conrad, W. Guggino, T. Reynolds, B. Rosenstein, . . . R. Wetzel, *A phase I study of an adeno-associated virus-CFTR gene vector in adult CF patients with mild lung disease.* Hum Gene Ther, 1996. **7**(9): p. 1145-59.
176. Wagner, J.A., T. Reynolds, M.L. Moran, R.B. Moss, J.J. Wine, T.R. Flotte, and P. Gardner, *Efficient and persistent gene transfer of AAV-CFTR in maxillary sinus.* Lancet, 1998. **351**(9117): p. 1702-3.
177. Moss, R.B., C. Milla, J. Colombo, F. Accurso, P.L. Zeitlin, J.P. Clancy, . . . A.E. Heald, *Repeated aerosolized AAV-CFTR for treatment of cystic fibrosis: a randomized placebo-controlled phase 2B trial.* Hum Gene Ther, 2007. **18**(8): p. 726-32.
178. Moss, R.B., D. Rodman, L.T. Spencer, M.L. Aitken, P.L. Zeitlin, D. Waltz, . . . A.E. Heald, *Repeated adeno-associated virus serotype 2 aerosol-mediated cystic fibrosis transmembrane regulator gene transfer to the lungs of patients with cystic fibrosis: a multicenter, double-blind, placebo-controlled trial.* Chest, 2004. **125**(2): p. 509-21.
179. Wagner, J.A., I.B. Nepomuceno, A.H. Messner, M.L. Moran, E.P. Batson, S. Dimiceli, . . . P. Gardner, *A phase II, double-blind, randomized, placebo-controlled clinical trial of tgAAVCF using maxillary sinus delivery in patients with cystic fibrosis with antrostomies.* Hum Gene Ther, 2002. **13**(11): p. 1349-59.
180. Hamilton, B.A., X. Li, A.A. Pezzulo, M.H. Abou Alaiwa, and J. Zabner, *Polarized AAVR expression determines infectivity by AAV gene therapy vectors.* Gene Ther, 2019. **26**(6): p. 240-249.
181. Yan, Z., D.C. Lei-Butters, X. Liu, Y. Zhang, L. Zhang, M. Luo, . . . J.F. Engelhardt, *Unique biologic properties of recombinant AAV1 transduction in polarized human airway epithelia.* J Biol Chem, 2006. **281**(40): p. 29684-92.
182. Cmielewski, P., D.S. Anson, and D.W. Parsons, *Lysophosphatidylcholine as an adjuvant for lentiviral vector mediated gene transfer to airway epithelium: effect of acyl chain length.* Respir Res, 2010. **11**: p. 84.
183. de Graaf, M. and R.A.M. Fouchier, *Role of receptor binding specificity in influenza A virus transmission and pathogenesis.* Embo Journal, 2014. **33**(8): p. 823-841.
184. Hida, K., S.K. Lai, J.S. Suk, S.Y. Won, M.P. Boyle, and J. Hanes, *Common gene therapy viral vectors do not efficiently penetrate sputum from cystic fibrosis patients.* PLoS One, 2011. **6**(5): p. e19919.
185. Schuster, B.S., A.J. Kim, J.C. Kays, M.M. Kanzawa, W.B. Guggino, M.P. Boyle, . . . J. Hanes, *Overcoming the cystic fibrosis sputum barrier to leading adeno-associated virus gene therapy vectors.* Mol Ther, 2014. **22**(8): p. 1484-1493.
186. Atchison, R.W., B.C. Casto, and W.M. Hammon, *Adenovirus-Associated Defective Virus Particles.* Science, 1965. **149**(3685): p. 754-6.

187. Blacklow, N.R., M.D. Hoggan, and W.P. Rowe, *Isolation of adenovirus-associated viruses from man.* Proc Natl Acad Sci U S A, 1967. **58**(4): p. 1410-5.
188. Srivastava, A., E.W. Lusby, and K.I. Berns, *Nucleotide sequence and organization of the adeno-associated virus 2 genome.* J Virol, 1983. **45**(2): p. 555-64.
189. King, J.A., R. Dubielzig, D. Grimm, and J.A. Kleinschmidt, *DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids.* EMBO J, 2001. **20**(12): p. 3282-91.
190. Samulski, R.J. and N. Muzyczka, *AAV-Mediated Gene Therapy for Research and Therapeutic Purposes.* Annu Rev Virol, 2014. **1**(1): p. 427-51.
191. Sonntag, F., K. Kother, K. Schmidt, M. Weghofer, C. Raupp, K. Nieto, . . . J.A. Kleinschmidt, *The assembly-activating protein promotes capsid assembly of different adeno-associated virus serotypes.* J Virol, 2011. **85**(23): p. 12686-97.
192. Sonntag, F., K. Schmidt, and J.A. Kleinschmidt, *A viral assembly factor promotes AAV2 capsid formation in the nucleolus.* Proc Natl Acad Sci U S A, 2010. **107**(22): p. 10220-5.
193. Earley, L.F., J.M. Powers, K. Adachi, J.T. Baumgart, N.L. Meyer, Q. Xie, . . . H. Nakai, *Adeno-associated Virus (AAV) Assembly-Activating Protein Is Not an Essential Requirement for Capsid Assembly of AAV Serotypes 4, 5, and 11.* J Virol, 2017. **91**(3).
194. Cao, M., H. You, and P.L. Hermonat, *The X gene of adeno-associated virus 2 (AAV2) is involved in viral DNA replication.* PLoS One, 2014. **9**(8): p. e104596.
195. Ogden, P.J., E.D. Kelsic, S. Sinai, and G.M. Church, *Comprehensive AAV capsid fitness landscape reveals a viral gene and enables machine-guided design.* Science, 2019. **366**(6469): p. 1139-1143.
196. Hauswirth, W.W. and K.I. Berns, *Origin and termination of adeno-associated virus DNA replication.* Virology, 1977. **78**(2): p. 488-99.
197. Becerra, S.P., F. Koczot, P. Fabisch, and J.A. Rose, *Synthesis of adeno-associated virus structural proteins requires both alternative mRNA splicing and alternative initiations from a single transcript.* J Virol, 1988. **62**(8): p. 2745-54.
198. Xiao, X., J. Li, and R.J. Samulski, *Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus.* J Virol, 1998. **72**(3): p. 2224-32.
199. Nony, P., G. Chadeuf, J. Tessier, P. Moullier, and A. Salvetti, *Evidence for packaging of rep-cap sequences into adeno-associated virus (AAV) type 2 capsids in the absence of inverted terminal repeats: a model for generation of rep-positive AAV particles.* J Virol, 2003. **77**(1): p. 776-81.
200. Samulski, R.J., L.S. Chang, and T. Shenk, *Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression.* J Virol, 1989. **63**(9): p. 3822-8.
201. Zhang, H., B. Yang, X. Mu, S.S. Ahmed, Q. Su, R. He, . . . G. Gao, *Several rAAV vectors efficiently cross the blood-brain barrier and transduce neurons and astrocytes in the neonatal mouse central nervous system.* Mol Ther, 2011. **19**(8): p. 1440-8.
202. Nonnenmacher, M. and T. Weber, *Intracellular transport of recombinant adeno-associated virus vectors.* Gene Ther, 2012. **19**(6): p. 649-58.
203. Sonntag, F., S. Bleker, B. Leuchs, R. Fischer, and J.A. Kleinschmidt, *Adeno-associated virus type 2 capsids with externalized VP1/VP2 trafficking domains*

are generated prior to passage through the cytoplasm and are maintained until uncoating occurs in the nucleus. J Virol, 2006. **80**(22): p. 11040-54.
204. Li, H., Y. Yang, W. Hong, M. Huang, M. Wu, and X. Zhao, *Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects.* Signal Transduct Target Ther, 2020. **5**(1): p. 1.
205. Ferrari, F.K., T. Samulski, T. Shenk, and R.J. Samulski, *Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors.* J Virol, 1996. **70**(5): p. 3227-34.
206. Duan, D., P. Sharma, J. Yang, Y. Yue, L. Dudus, Y. Zhang, . . . J.F. Engelhardt, *Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue.* J Virol, 1998. **72**(11): p. 8568-77.
207. Miller, D.G., L.M. Petek, and D.W. Russell, *Adeno-associated virus vectors integrate at chromosome breakage sites.* Nat Genet, 2004. **36**(7): p. 767-73.
208. Deyle, D.R. and D.W. Russell, *Adeno-associated virus vector integration.* Curr Opin Mol Ther, 2009. **11**(4): p. 442-7.
209. Hanlon, K.S., B.P. Kleinstiver, S.P. Garcia, M.P. Zaborowski, A. Volak, S.E. Spirig, . . . B. Gyorgy, *High levels of AAV vector integration into CRISPR-induced DNA breaks.* Nat Commun, 2019. **10**(1): p. 4439.
210. Gruntman, A.M., L. Su, Q. Su, G. Gao, C. Mueller, and T.R. Flotte, *Stability and compatibility of recombinant adeno-associated virus under conditions commonly encountered in human gene therapy trials.* Hum Gene Ther Methods, 2015. **26**(2): p. 71-6.
211. MacLoughlin, R.J., B.D. Higgins, J. Devaney, D. O'Toole, J.G. Laffey, and T. O'Brien, *Aerosol-mediated delivery of AAV2/6-IkappaBalpha attenuates lipopolysaccharide-induced acute lung injury in rats.* Hum Gene Ther, 2015. **26**(1): p. 36-46.
212. Leung, K., E. Louca, K. Munson, B. Dutzar, P. Anklesaria, and A.L. Coates, *Calculating expected lung deposition of aerosolized administration of AAV vector in human clinical studies.* J Gene Med, 2007. **9**(1): p. 10-21.
213. Batty, P. and D. Lillicrap, *Advances and challenges for hemophilia gene therapy.* Hum Mol Genet, 2019. **28**(R1): p. R95-R101.
214. Keeler, A.M. and T.R. Flotte, *Recombinant Adeno-Associated Virus Gene Therapy in Light of Luxturna (and Zolgensma and Glybera): Where Are We, and How Did We Get Here?* Annu Rev Virol, 2019. **6**(1): p. 601-621.
215. Wang, D., P.W.L. Tai, and G.P. Gao, *Adeno-associated virus vector as a platform for gene therapy delivery.* Nature Reviews Drug Discovery, 2019. **18**(5): p. 358-378.
216. Gottlieb, S., *Statement from FDA Commissioner Scott Gottlieb, M.D. and Peter Marks, M.D., Ph.D., Director of the Center for Biologics Evaluation and Research on new policies to advance development of safe and effective cell and gene therapies.* 2019, FDA.
217. Zinn, E. and L.H. Vandenberghe, *Adeno-associated virus: fit to serve.* Curr Opin Virol, 2014. **8**: p. 90-7.
218. Sumner-Jones, S.G., L.A. Davies, A. Varathalingam, D.R. Gill, and S.C. Hyde, *Long-term persistence of gene expression from adeno-associated virus serotype 5 in the mouse airways.* Gene Ther, 2006. **13**(24): p. 1703-13.

219. Wu, Z.J., H.Y. Yang, and P. Colosi, *Effect of Genome Size on AAV Vector Packaging.* Molecular Therapy, 2010. **18**(1): p. 80-86.
220. England, S.B., L.V. Nicholson, M.A. Johnson, S.M. Forrest, D.R. Love, E.E. Zubrzycka-Gaarn, . . . K.E. Davies, *Very mild muscular dystrophy associated with the deletion of 46% of dystrophin.* Nature, 1990. **343**(6254): p. 180-2.
221. Gao, G., L.H. Vandenberghe, M.R. Alvira, Y. Lu, R. Calcedo, X. Zhou, and J.M. Wilson, *Clades of Adeno-associated viruses are widely disseminated in human tissues.* J Virol, 2004. **78**(12): p. 6381-8.
222. Chen, C.L., R.L. Jensen, B.C. Schnepp, M.J. Connell, R. Shell, T.J. Sferra, . . . P.R. Johnson, *Molecular characterization of adeno-associated viruses infecting children.* J Virol, 2005. **79**(23): p. 14781-92.
223. Boutin, S., V. Monteilhet, P. Veron, C. Leborgne, O. Benveniste, M.F. Montus, and C. Masurier, *Prevalence of serum IgG and neutralizing factors against adeno-associated virus (AAV) types 1, 2, 5, 6, 8, and 9 in the healthy population: implications for gene therapy using AAV vectors.* Hum Gene Ther, 2010. **21**(6): p. 704-12.
224. Bello, A., A. Chand, J. Aviles, G. Soule, A. Auricchio, and G.P. Kobinger, *Novel Adeno-associated Viruses Derived From Pig Tissues Transduce Most Major Organs in Mice.* Scientific Reports, 2014. **4**.
225. Leborgne, C., E. Barbon, J.M. Alexander, H. Hanby, S. Delignat, D.M. Cohen, . . . F. Mingozzi, *IgG-cleaving endopeptidase enables in vivo gene therapy in the presence of anti-AAV neutralizing antibodies.* Nat Med, 2020.
226. Hordeaux, J., Q. Wang, N. Katz, E.L. Buza, P. Bell, and J.M. Wilson, *The Neurotropic Properties of AAV-PHP.B Are Limited to C57BL/6J Mice.* Mol Ther, 2018. **26**(3): p. 664-668.
227. Hinderer, C., N. Katz, E.L. Buza, C. Dyer, T. Goode, P. Bell, . . . J.M. Wilson, *Severe toxicity in nonhuman primates and piglets following high-dose intravenous administration of an AAV vector expressing human SMN.* Hum Gene Ther, 2018.
228. Hordeaux, J., E.L. Buza, C. Dyer, T. Goode, T.W. Mitchell, L. Richman, . . . J.M. Wilson, *Adeno-Associated Virus-Induced Dorsal Root Ganglion Pathology.* Hum Gene Ther, 2020. **31**(15-16): p. 808-818.
229. Qin, W., G. Xu, P.W.L. Tai, C. Wang, L. Luo, C. Li, . . . G. Gao, *Large-scale molecular epidemiological analysis of AAV in a cancer patient population.* Oncogene, 2021. **40**(17): p. 3060-3071.
230. George, L.A., M.V. Ragni, J.E.J. Rasko, L.J. Raffini, B.J. Samelson-Jones, M. Ozelo, . . . K.A. High, *Long-Term Follow-Up of the First in Human Intravascular Delivery of AAV for Gene Transfer: AAV2-hFIX16 for Severe Hemophilia B.* Mol Ther, 2020. **28**(9): p. 2073-2082.
231. de Jong, Y.P. and R.W. Herzog, *Liver gene therapy and hepatocellular carcinoma: A complex web.* Mol Ther, 2021. **29**(4): p. 1353-1354.
232. *Zolgensma® data shows rapid, significant, clinically meaningful benefit in SMA including prolonged event-free survival, motor milestone achievement and durability now up to 5 years post-dosing.* 2020; Cumulative Safety Data as of December 31, 2019]. Available from: https://www.novartis.com/news/media-releases/zolgensma-data-shows-rapid-significant-clinically-meaningful-benefit-sma-including-prolonged-event-free-survival-motor-milestone-achievement-and-durability-now.

233. Sender, R., S. Fuchs, and R. Milo, *Revised Estimates for the Number of Human and Bacteria Cells in the Body.* PLoS Biol, 2016. **14**(8): p. e1002533.
234. *High-dose AAV gene therapy deaths.* Nat Biotechnol, 2020. **38**(8): p. 910.
235. Philippidis, A., *After Third Death, Audentes' AT132 Remains on Clinical Hold.* Hum Gene Ther, 2020. **31**(17-18): p. 908-910.
236. Srivastava, A., *AAV vectors: Are they safe?* Hum Gene Ther, 2020.
237. Gao, G.P., M.R. Alvira, L. Wang, R. Calcedo, J. Johnston, and J.M. Wilson, *Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy.* Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11854-9.
238. Wu, Z., A. Asokan, J.C. Grieger, L. Govindasamy, M. Agbandje-McKenna, and R.J. Samulski, *Single amino acid changes can influence titer, heparin binding, and tissue tropism in different adeno-associated virus serotypes.* Journal of Virology, 2006. **80**(22): p. 11393-11397.
239. Grieger, J.C., S. Snowdy, and R.J. Samulski, *Separate basic region motifs within the adeno-associated virus capsid proteins are essential for infectivity and assembly.* J Virol, 2006. **80**(11): p. 5199-210.
240. Kronenberg, S., B. Bottcher, C.W. von der Lieth, S. Bleker, and J.A. Kleinschmidt, *A conformational change in the adeno-associated virus type 2 capsid leads to the exposure of hidden VP1 N termini.* J Virol, 2005. **79**(9): p. 5296-303.
241. Agbandje-McKenna, M. and J. Kleinschmidt, *AAV capsid structure and cell interactions.* Methods Mol Biol, 2011. **807**: p. 47-92.
242. Govindasamy, L., E. Padron, R. McKenna, N. Muzyczka, N. Kaludov, J.A. Chiorini, and M. Agbandje-McKenna, *Structurally mapping the diverse phenotype of adeno-associated virus serotype 4.* J Virol, 2006. **80**(23): p. 11556-70.
243. Nam, H.J., M.D. Lane, E. Padron, B. Gurda, R. McKenna, E. Kohlbrenner, . . . M. Agbandje-McKenna, *Structure of adeno-associated virus serotype 8, a gene therapy vector.* J Virol, 2007. **81**(22): p. 12260-71.
244. Vandenberghe, L.H., E. Breous, H.J. Nam, G. Gao, R. Xiao, A. Sandhu, . . . J.M. Wilson, *Naturally occurring singleton residues in AAV capsid impact vector performance and illustrate structural constraints.* Gene Ther, 2009. **16**(12): p. 1416-28.
245. Kern, A., K. Schmidt, C. Leder, O.J. Muller, C.E. Wobus, K. Bettinger, . . . J.A. Kleinschmidt, *Identification of a heparin-binding motif on adeno-associated virus type 2 capsids.* J Virol, 2003. **77**(20): p. 11072-81.
246. Opie, S.R., K.H. Warrington, Jr., M. Agbandje-McKenna, S. Zolotukhin, and N. Muzyczka, *Identification of amino acid residues in the capsid proteins of adeno-associated virus type 2 that contribute to heparan sulfate proteoglycan binding.* J Virol, 2003. **77**(12): p. 6995-7006.
247. Huang, L.Y., A. Patel, R. Ng, E.B. Miller, S. Halder, R. McKenna, . . . M. Agbandje-McKenna, *Characterization of the Adeno-Associated Virus 1 and 6 Sialic Acid Binding Site.* J Virol, 2016. **90**(11): p. 5219-5230.
248. Excoffon, K.J., J.T. Koerber, D.D. Dickey, M. Murtha, S. Keshavjee, B.K. Kaspar, . . . D.V. Schaffer, *Directed evolution of adeno-associated virus to an infectious respiratory virus.* Proc Natl Acad Sci U S A, 2009. **106**(10): p. 3865-70.

249. Pillay, S., N.L. Meyer, A.S. Puschnik, O. Davulcu, J. Diep, Y. Ishikawa, . . . J.E. Carette, *An essential receptor for adeno-associated virus infection.* Nature, 2016. **530**(7588): p. 108-12.
250. Dudek, A.M., N. Zabaleta, E. Zinn, S. Pillay, J. Zengel, C. Porter, . . . L.H. Vandenberghe, *GPR108 Is a Highly Conserved AAV Entry Factor.* Molecular Therapy, 2020. **28**(2): p. 367-381.
251. Yan, Z., Z. Feng, X. Sun, Y. Zhang, W. Zou, Z. Wang, . . . J.F. Engelhardt, *Human Bocavirus Type-1 Capsid Facilitates the Transduction of Ferret Airways by Adeno-Associated Virus Genomes.* Hum Gene Ther, 2017. **28**(8): p. 612-625.
252. Liu, X., Z. Yan, M. Luo, and J.F. Engelhardt, *Species-specific differences in mouse and human airway epithelial biology of recombinant adeno-associated virus transduction.* Am J Respir Cell Mol Biol, 2006. **34**(1): p. 56-64.
253. Beck, S.E., B.L. Laube, C.I. Barberena, A.C. Fischer, R.J. Adams, K. Chesnut, . . . W.B. Guggino, *Deposition and expression of aerosolized rAAV vectors in the lungs of Rhesus macaques.* Mol Ther, 2002. **6**(4): p. 546-54.
254. Flotte, T.R., A.C. Fischer, J. Goetzmann, C. Mueller, L. Cebotaru, Z. Yan, . . . J.F. Engelhardt, *Dual reporter comparative indexing of rAAV pseudotyped vectors in chimpanzee airway.* Mol Ther, 2010. **18**(3): p. 594-600.
255. Guggino, W.B., J. Benson, J. Seagrave, Z. Yan, J. Engelhardt, G. Gao, . . . L. Cebotaru, *A Preclinical Study in Rhesus Macaques for Cystic Fibrosis to Assess Gene Transfer and Transduction by AAV1 and AAV5 with a Dual-Luciferase Reporter System.* Hum Gene Ther Clin Dev, 2017. **28**(3): p. 145-156.
256. Wu, Z., E. Miller, M. Agbandje-McKenna, and R.J. Samulski, *Alpha2,3 and alpha2,6 N-linked sialic acids facilitate efficient binding and transduction by adeno-associated virus types 1 and 6.* J Virol, 2006. **80**(18): p. 9093-103.
257. Ng, R., L. Govindasamy, B.L. Gurda, R. McKenna, O.G. Kozyreva, R.J. Samulski, . . . M. Agbandje-McKenna, *Structural characterization of the dual glycan binding adeno-associated virus serotype 6.* J Virol, 2010. **84**(24): p. 12945-57.
258. Hadaczek, P., L. Stanek, A. Ciesielska, V. Sudhakar, L. Samaranch, P. Pivirotto, . . . L.S. Shihabuddin, *Widespread AAV1- and AAV2-mediated transgene expression in the nonhuman primate brain: implications for Huntington's disease.* Mol Ther Methods Clin Dev, 2016. **3**: p. 16037.
259. Brantly, M.L., J.D. Chulay, L. Wang, C. Mueller, M. Humphries, L.T. Spencer, . . . T.R. Flotte, *Sustained transgene expression despite T lymphocyte responses in a clinical trial of rAAV1-AAT gene therapy.* Proc Natl Acad Sci U S A, 2009. **106**(38): p. 16363-8.
260. Su, H., Y. Yeghiazarians, A. Lee, Y. Huang, J. Arakawa-Hoyt, J. Ye, . . . Y.W. Kan, *AAV serotype 1 mediates more efficient gene transfer to pig myocardium than AAV serotype 2 and plasmid.* J Gene Med, 2008. **10**(1): p. 33-41.
261. Aguero, J., K. Ishikawa, L. Hadri, C.G. Santos-Gallego, K.M. Fish, E. Kohlbrenner, . . . R.J. Hajjar, *Intratracheal Gene Delivery of SERCA2a Ameliorates Chronic Post-Capillary Pulmonary Hypertension: A Large Animal Model.* J Am Coll Cardiol, 2016. **67**(17): p. 2032-46.
262. McClain, L.E., M.G. Davey, P.W. Zoltick, M.P. Limberis, A.W. Flake, and W.H. Peranteau, *Vector serotype screening for use in ovine perinatal lung gene therapy.* J Pediatr Surg, 2016. **51**(6): p. 879-84.

263. Summerford, C. and R.J. Samulski, *Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions.* Journal of Virology, 1998. **72**(2): p. 1438-1445.
264. Kashiwakura, Y., K. Tamayose, K. Iwabuchi, Y. Hirai, T. Shimada, K. Matsumoto, . . . H. Daida, *Hepatocyte growth factor receptor is a coreceptor for adeno-associated virus type 2 infection.* J Virol, 2005. **79**(1): p. 609-14.
265. Qing, K., C. Mah, J. Hansen, S.Z. Zhou, V. Dwarki, and A. Srivastava, *Human fibroblast growth factor receptor 1 is a co-receptor for infection by adeno-associated virus 2.* Nature Medicine, 1999. **5**(1): p. 71-77.
266. Akache, B., D. Grimm, K. Pandey, S.R. Yant, H. Xu, and M.A. Kay, *The 37/67-kilodalton laminin receptor is a receptor for adeno-associated virus serotypes 8, 2, 3, and 9.* J Virol, 2006. **80**(19): p. 9831-6.
267. Asokan, A., J.B. Hamra, L. Govindasamy, M. Agbandje-McKenna, and R.J. Samulski, *Adeno-associated virus type 2 contains an integrin alpha5beta1 binding domain essential for viral cell entry.* J Virol, 2006. **80**(18): p. 8961-9.
268. Kurzeder, C., B. Koppold, G. Sauer, S. Pabst, R. Kreienberg, and H. Deissler, *CD9 promotes adeno-associated virus type 2 infection of mammary carcinoma cells with low cell surface expression of heparan sulphate proteoglycans.* Int J Mol Med, 2007. **19**(2): p. 325-33.
269. Manno, C.S., A.J. Chew, S. Hutchison, P.J. Larson, R.W. Herzog, V.R. Arruda, . . . B. Glader, *AAV-mediated factor IX gene transfer to skeletal muscle in patients with severe hemophilia B.* Blood, 2003. **101**(8): p. 2963-72.
270. Kay, M.A., C.S. Manno, M.V. Ragni, P.J. Larson, L.B. Couto, A. McClelland, . . . K.A. High, *Evidence for gene transfer and expression of factor IX in haemophilia B patients treated with an AAV vector.* Nat Genet, 2000. **24**(3): p. 257-61.
271. Rabinowitz, J.E., F. Rolling, C. Li, H. Conrath, W. Xiao, X. Xiao, and R.J. Samulski, *Cross-packaging of a single adeno-associated virus (AAV) type 2 vector genome into multiple AAV serotypes enables transduction with broad specificity.* J Virol, 2002. **76**(2): p. 791-801.
272. Blackburn, S.D., R.A. Steadman, and F.B. Johnson, *Attachment of adeno-associated virus type 3H to fibroblast growth factor receptor 1.* Arch Virol, 2006. **151**(3): p. 617-23.
273. Ling, C., Y. Lu, J.K. Kalsi, G.R. Jayandharan, B. Li, W. Ma, . . . A. Srivastava, *Human hepatocyte growth factor receptor is a cellular coreceptor for adeno-associated virus serotype 3.* Hum Gene Ther, 2010. **21**(12): p. 1741-7.
274. Glushakova, L.G., M.J. Lisankie, E.B. Eruslanov, C. Ojano-Dirain, I. Zolotukhin, C. Liu, . . . P.W. Stacpoole, *AAV3-mediated transfer and expression of the pyruvate dehydrogenase E1 alpha subunit gene causes metabolic remodeling and apoptosis of human liver cancer cells.* Mol Genet Metab, 2009. **98**(3): p. 289-99.
275. Kaludov, N., K.E. Brown, R.W. Walters, J. Zabner, and J.A. Chiorini, *Adeno-associated virus serotype 4 (AAV4) and AAV5 both require sialic acid binding for hemagglutination and efficient transduction but differ in sialic acid linkage specificity.* Journal of Virology, 2001. **75**(15): p. 6884-6893.
276. Dudek, A.M., S. Pillay, A.S. Puschnik, C.M. Nagamine, F. Cheng, J. Qiu, . . . L.H. Vandenberghe, *An Alternate Route for Adeno-associated Virus (AAV) Entry Independent of AAV Receptor.* J Virol, 2018. **92**(7).

277. Wiley, L.A., E.R. Burnight, E.E. Kaalberg, C. Jiao, M.J. Riker, J.A. Halder, . . . R.F. Mullins, *Assessment of Adeno-Associated Virus Serotype Tropism in Human Retinal Explants.* Hum Gene Ther, 2018. **29**(4): p. 424-436.
278. Walters, R.W., S.M. Yi, S. Keshavjee, K.E. Brown, M.J. Welsh, J.A. Chiorini, and J. Zabner, *Binding of adeno-associated virus type 5 to 2,3-linked sialic acid is required for gene transfer.* J Biol Chem, 2001. **276**(23): p. 20610-6.
279. Zhang, R., G. Xu, L. Cao, Z. Sun, Y. He, M. Cui, . . . Z. Lou, *Divergent engagements between adeno-associated viruses with their cellular receptor AAVR.* Nat Commun, 2019. **10**(1): p. 3760.
280. Di Pasquale, G., B.L. Davidson, C.S. Stein, I. Martins, D. Scudiero, A. Monks, and J.A. Chiorini, *Identification of PDGFR as a receptor for AAV-5 transduction.* Nat Med, 2003. **9**(10): p. 1306-12.
281. Colle, M.A., F. Piguet, L. Bertrand, S. Raoul, I. Bieche, L. Dubreil, . . . C. Sevin, *Efficient intracerebral delivery of AAV5 vector encoding human ARSA in non-human primate.* Hum Mol Genet, 2010. **19**(1): p. 147-58.
282. Zabner, J., M. Seiler, R. Walters, R.M. Kotin, W. Fulgeras, B.L. Davidson, and J.A. Chiorini, *Adeno-associated virus type 5 (AAV5) but not AAV2 binds to the apical surfaces of airway epithelia and facilitates gene transfer.* J Virol, 2000. **74**(8): p. 3852-8.
283. Limberis, M.P., L.H. Vandenberghe, L. Zhang, R.J. Pickles, and J.M. Wilson, *Transduction efficiencies of novel AAV vectors in mouse airway epithelium in vivo and human ciliated airway epithelium in vitro.* Mol Ther, 2009. **17**(2): p. 294-301.
284. Weller, M.L., P. Amornphimoltham, M. Schmidt, P.A. Wilson, J.S. Gutkind, and J.A. Chiorini, *Epidermal growth factor receptor is a co-receptor for adeno-associated virus serotype 6.* Nat Med, 2010. **16**(6): p. 662-4.
285. Towne, C., B.L. Schneider, D. Kieran, D.E. Redmond, Jr., and P. Aebischer, *Efficient transduction of non-human primate motor neurons after intramuscular delivery of recombinant AAV serotype 6.* Gene Ther, 2010. **17**(1): p. 141-6.
286. Raake, P.W., R. Hinkel, S. Muller, S. Delker, R. Kreuzpointner, C. Kupatt, . . . O.J. Muller, *Cardio-specific long-term gene expression in a porcine model after selective pressure-regulated retroinfusion of adeno-associated viral (AAV) vectors.* Gene Ther, 2008. **15**(1): p. 12-7.
287. Duncan, G.A., N. Kim, Y. Colon-Cortes, J. Rodriguez, M. Mazur, S.E. Birket, . . . J.S. Suk, *An Adeno-Associated Viral Vector Capable of Penetrating the Mucus Barrier to Inhaled Gene Therapy.* Mol Ther Methods Clin Dev, 2018. **9**: p. 296-304.
288. Samaranch, L., E.A. Salegio, W. San Sebastian, A.P. Kells, J.R. Bringas, J. Forsayeth, and K.S. Bankiewicz, *Strong cortical and spinal cord transduction after AAV7 and AAV9 delivery into the cerebrospinal fluid of nonhuman primates.* Hum Gene Ther, 2013. **24**(5): p. 526-32.
289. Gao, G., Y. Lu, R. Calcedo, R.L. Grant, P. Bell, L. Wang, . . . J.M. Wilson, *Biology of AAV serotype vectors in liver-directed gene transfer to nonhuman primates.* Mol Ther, 2006. **13**(1): p. 77-87.
290. Muraine, L., M. Bensalah, J. Dhiab, G. Cordova, L. Arandel, A. Marhic, . . . C. Trollet, *Transduction Efficiency of Adeno-Associated Virus Serotypes After Local Injection in Mouse and Human Skeletal Muscle.* Hum Gene Ther, 2020. **31**(3-4): p. 233-240.

291. Rodino-Klapac, L.R., C.L. Montgomery, W.G. Bremer, K.M. Shontz, V. Malik, N. Davis, . . . L.G. Chicoine, *Persistent expression of FLAG-tagged micro dystrophin in nonhuman primates following intramuscular and vascular delivery.* Mol Ther, 2010. **18**(1): p. 109-17.
292. Shen, S., K.D. Bryant, S.M. Brown, S.H. Randell, and A. Asokan, *Terminal N-linked galactose is the primary receptor for adeno-associated virus 9.* J Biol Chem, 2011. **286**(15): p. 13532-40.
293. Green, F., L. Samaranch, H.S. Zhang, A. Manning-Bog, K. Meyer, J. Forsayeth, and K.S. Bankiewicz, *Axonal transport of AAV9 in nonhuman primate brain.* Gene Ther, 2016. **23**(6): p. 520-6.
294. Gray, S.J., S. Nagabhushan Kalburgi, T.J. McCown, and R. Jude Samulski, *Global CNS gene delivery and evasion of anti-AAV-neutralizing antibodies by intrathecal AAV administration in non-human primates.* Gene Ther, 2013. **20**(4): p. 450-9.
295. Mori, S., L. Wang, T. Takeuchi, and T. Kanda, *Two novel adeno-associated viruses from cynomolgus monkey: pseudotyping characterization of capsid protein.* Virology, 2004. **330**(2): p. 375-83.
296. Ai, J., J. Li, D.J. Gessler, Q. Su, Q. Wei, H. Li, and G. Gao, *Adeno-associated virus serotype rh.10 displays strong muscle tropism following intraperitoneal delivery.* Sci Rep, 2017. **7**: p. 40336.
297. Yang, B., S. Li, H. Wang, Y. Guo, D.J. Gessler, C. Cao, . . . G. Gao, *Global CNS transduction of adult mice by intravenously delivered rAAVrh.8 and rAAVrh.10 and nonhuman primates by rAAVrh.10.* Mol Ther, 2014. **22**(7): p. 1299-1309.
298. Schmidt, M., A. Voutetakis, S. Afione, C. Zheng, D. Mandikian, and J.A. Chiorini, *Adeno-associated virus type 12 (AAV12): a novel AAV serotype with sialic acid- and heparan sulfate proteoglycan-independent transduction activity.* J Virol, 2008. **82**(3): p. 1399-406.
299. Schmidt, M., L. Govindasamy, S. Afione, N. Kaludov, M. Agbandje-McKenna, and J.A. Chiorini, *Molecular characterization of the heparin-dependent transduction domain on the capsid of a novel adeno-associated virus isolate, AAV(VR-942).* J Virol, 2008. **82**(17): p. 8911-6.
300. Asokan, A., D.V. Schaffer, and R.J. Samulski, *The AAV vector toolkit: poised at the clinical crossroads.* Mol Ther, 2012. **20**(4): p. 699-708.
301. Naso, M.F., B. Tomkowicz, W.L. Perry, 3rd, and W.R. Strohl, *Adeno-Associated Virus (AAV) as a Vector for Gene Therapy.* BioDrugs, 2017. **31**(4): p. 317-334.
302. Li, C. and R.J. Samulski, *Engineering adeno-associated virus vectors for gene therapy.* Nat Rev Genet, 2020. **21**(4): p. 255-272.
303. Srivastava, A., *In vivo tissue-tropism of adeno-associated viral vectors.* Curr Opin Virol, 2016. **21**: p. 75-80.
304. Pillay, S. and J.E. Carette, *Host determinants of adeno-associated viral vector entry.* Curr Opin Virol, 2017. **24**: p. 124-131.
305. Herrmann, A.K. and D. Grimm, *High-Throughput Dissection of AAV-Host Interactions: The Fast and the Curious.* J Mol Biol, 2018. **430**(17): p. 2626-2640.
306. Bowles, D.E., S.W. McPhee, C. Li, S.J. Gray, J.J. Samulski, A.S. Camp, . . . R.J. Samulski, *Phase 1 gene therapy for Duchenne muscular dystrophy using a translational optimized AAV vector.* Mol Ther, 2012. **20**(2): p. 443-55.
307. Asokan, A., J.C. Conway, J.L. Phillips, C. Li, J. Hegge, R. Sinnott, . . . R.J. Samulski, *Reengineering a receptor footprint of adeno-associated virus enables*

selective and systemic gene transfer to muscle. Nat Biotechnol, 2010. **28**(1): p. 79-82.
308. Lochrie, M.A., G.P. Tatsuno, B. Christie, J.W. McDonnell, S. Zhou, R. Surosky, . . . P. Colosi, *Mutations on the external surfaces of adeno-associated virus type 2 capsids that affect transduction and neutralization.* J Virol, 2006. **80**(2): p. 821-34.
309. Aydemir, F., M. Salganik, J. Resztak, J. Singh, A. Bennett, M. Agbandje-McKenna, and N. Muzyczka, *Mutants at the 2-Fold Interface of Adeno-associated Virus Type 2 (AAV2) Structural Proteins Suggest a Role in Viral Transcription for AAV Capsids.* J Virol, 2016. **90**(16): p. 7196-7204.
310. van Lieshout, L.P., J.M. Domm, T.N. Rindler, K.L. Frost, D.L. Sorensen, S.J. Medina, . . . S.K. Wootton, *A Novel Triple-Mutant AAV6 Capsid Induces Rapid and Potent Transgene Expression in the Muscle and Respiratory Tract of Mice.* Mol Ther Methods Clin Dev, 2018. **9**: p. 323-329.
311. Marsic, D., L. Govindasamy, S. Currlin, D.M. Markusic, Y.S. Tseng, R.W. Herzog, . . . S. Zolotukhin, *Vector design Tour de Force: integrating combinatorial and rational approaches to derive novel adeno-associated virus variants.* Mol Ther, 2014. **22**(11): p. 1900-9.
312. Zinn, E., S. Pacouret, V. Khaychuk, H.T. Turunen, L.S. Carvalho, E. Andres-Mateos, . . . L.H. Vandenberghe, *In Silico Reconstruction of the Viral Evolutionary Lineage Yields a Potent Gene Therapy Vector.* Cell Rep, 2015. **12**(6): p. 1056-68.
313. Maheshri, N., J.T. Koerber, B.K. Kaspar, and D.V. Schaffer, *Directed evolution of adeno-associated virus yields enhanced gene delivery vectors.* Nat Biotechnol, 2006. **24**(2): p. 198-204.
314. Koerber, J.T., J.H. Jang, and D.V. Schaffer, *DNA shuffling of adeno-associated virus yields functionally diverse viral progeny.* Mol Ther, 2008. **16**(10): p. 1703-9.
315. Fakhiri, J., M.A. Schneider, J. Puschhof, M. Stanifer, V. Schildgen, S. Holderbach, . . . D. Grimm, *Novel Chimeric Gene Therapy Vectors Based on Adeno-Associated Virus and Four Different Mammalian Bocaviruses.* Mol Ther Methods Clin Dev, 2019. **12**: p. 202-222.
316. Dickey, D.D., K.J. Excoffon, J.T. Koerber, J. Bergen, B. Steines, J. Klesney-Tait, . . . J. Zabner, *Enhanced sialic acid-dependent endocytosis explains the increased efficiency of infection of airway epithelia by a novel adeno-associated virus.* J Virol, 2011. **85**(17): p. 9023-30.
317. Steines, B., D.D. Dickey, J. Bergen, K.J. Excoffon, J.R. Weinstein, X. Li, . . . J. Zabner, *CFTR gene transfer with AAV improves early cystic fibrosis pig phenotypes.* JCI Insight, 2016. **1**(14): p. e88728.
318. Korst, R.J., B. Bewig, and R.G. Crystal, *In vitro and in vivo transfer and expression of human surfactant SP-A- and SP-B-associated protein cDNAs mediated by replication-deficient, recombinant adenoviral vectors.* Hum Gene Ther, 1995. **6**(3): p. 277-87.
319. Yei, S., C.J. Bachurski, T.E. Weaver, S.E. Wert, B.C. Trapnell, and J.A. Whitsett, *Adenoviral-mediated gene transfer of human surfactant protein B to respiratory epithelial cells.* Am J Respir Cell Mol Biol, 1994. **11**(3): p. 329-36.
320. Melton, K.R., L.L. Nesslein, M. Ikegami, J.W. Tichelaar, J.C. Clark, J.A. Whitsett, and T.E. Weaver, *SP-B deficiency causes respiratory failure in adult mice.* Am J Physiol Lung Cell Mol Physiol, 2003. **285**(3): p. L543-9.

321. Barnett, R.C., X. Lin, M. Barravecchia, R.A. Norman, K.L. de Mesy Bentley, F. Fazal, . . . D.A. Dean, *Featured Article: Electroporation-mediated gene delivery of surfactant protein B (SP-B) restores expression and improves survival in mouse model of SP-B deficiency.* Exp Biol Med (Maywood), 2017. **242**(13): p. 1345-1354.
322. Kormann, M.S., G. Hasenpusch, M.K. Aneja, G. Nica, A.W. Flemmer, S. Herber-Jonat, . . . C. Rudolph, *Expression of therapeutic proteins after delivery of chemically modified mRNA in mice.* Nat Biotechnol, 2011. **29**(2): p. 154-7.
323. Mahiny, A.J., A. Dewerth, L.E. Mays, M. Alkhaled, B. Mothes, E. Malaeksefat, . . . M.S. Kormann, *In vivo genome editing using nuclease-encoding mRNA corrects SP-B deficiency.* Nat Biotechnol, 2015. **33**(6): p. 584-6.
324. Jacob, A., M. Morley, F. Hawkins, K.B. McCauley, J.C. Jean, H. Heins, . . . D.N. Kotton, *Differentiation of Human Pluripotent Stem Cells into Functional Lung Alveolar Epithelial Cells.* Cell Stem Cell, 2017. **21**(4): p. 472-488 e10.
325. Leibel, S.L., A. Winquist, I. Tseu, J. Wang, D. Luo, S. Shojaie, . . . M. Post, *Reversal of Surfactant Protein B Deficiency in Patient Specific Human Induced Pluripotent Stem Cell Derived Lung Organoids by Gene Therapy.* Sci Rep, 2019. **9**(1): p. 13450.
326. Pensiero, M., M.K.L. Collins, F.-L. Cosset, Y. Takeuchi, and R.A. Weiss, *Retroviral vectors produced by producer cell lines resistant to lysis by human serum*. 1999, Google Patents.
327. Giard, D.J., S.A. Aaronson, G.J. Todaro, P. Arnstein, J.H. Kersey, H. Dosik, and W.P. Parks, *In vitro cultivation of human tumors: establishment of cell lines derived from a series of solid tumors.* J Natl Cancer Inst, 1973. **51**(5): p. 1417-23.
328. Lacoste, A., F. Berenshteyn, and A.H. Brivanlou, *An efficient and reversible transposable system for gene delivery and lineage-specific differentiation in human embryonic stem cells.* Cell Stem Cell, 2009. **5**(3): p. 332-42.
329. Hsiau, T., D. Conant, N. Rossi, T. Maures, K. Waite, J. Yang, . . . R. Stoner, *Inference of CRISPR Edits from Sanger Trace Data.* bioRxiv, 2019: p. 251082.
330. de Silva, S.R., M.E. McClements, M.W. Hankins, and R.E. MacLaren, *Adeno-Associated Viral Gene Therapy for Retinal Disorders*, in *Gene Delivery and Therapy for Neurological Disorders*, X. Bo and J. Verhaagen, Editors. 2015, Springer New York: New York, NY. p. 203-228.
331. Rosales Gerpe, M.C., J.P. van Vloten, L.A. Santry, J. de Jong, R.C. Mould, A. Pelin, . . . S.K. Wootton, *Use of Precision-Cut Lung Slices as an Ex Vivo Tool for Evaluating Viruses and Viral Vectors for Gene and Oncolytic Therapy.* Mol Ther Methods Clin Dev, 2018. **10**: p. 245-256.
332. Meyer-Berg, H., L. Zhou Yang, M. Pilar de Lucas, A. Zambrano, S.C. Hyde, and D.R. Gill, *Identification of AAV serotypes for lung gene therapy in human embryonic stem cell-derived lung organoids.* Stem Cell Res Ther, 2020. **11**(1): p. 448.
333. Chen, Y.W., S.X. Huang, A. de Carvalho, S.H. Ho, M.N. Islam, S. Volpi, . . . H.W. Snoeck, *A three-dimensional model of human lung development and disease from pluripotent stem cells.* Nat Cell Biol, 2017. **19**(5): p. 542-549.
334. Wu, X., H.Y. Fay Wang, N. Su, L.C. Wang, and Y. Luo, *Ultra sensitive method for in situ detection of nucleic acids*. 2014, Advanced Cell Diagnostics Inc: USA.

335. Billinton, N. and A.W. Knight, *Seeing the wood through the trees: A review of techniques for distinguishing green fluorescent protein from endogenous autofluorescence.* Analytical Biochemistry, 2001. **291**(2): p. 175-197.
336. Pilot-Matias, T.J., S.E. Kister, J.L. Fox, K. Kropp, S.W. Glasser, and J.A. Whitsett, *Structure and organization of the gene encoding human pulmonary surfactant proteolipid SP-B.* DNA, 1989. **8**(2): p. 75-86.
337. Kozak, M., *An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs.* Nucleic Acids Res, 1987. **15**(20): p. 8125-48.
338. Fehrholz, M., I. Bersani, B.W. Kramer, C.P. Speer, and S. Kunzmann, *Synergistic effect of caffeine and glucocorticoids on expression of surfactant protein B (SP-B) mRNA.* PLoS One, 2012. **7**(12): p. e51575.
339. Vieira, F., J.W. Kung, and F. Bhatti, *Structure, genetics and function of the pulmonary associated surfactant proteins A and D: The extra-pulmonary role of these C type lectins.* Ann Anat, 2017. **211**: p. 184-201.
340. Haagsman, H.P., R.T. White, J. Schilling, K. Lau, B.J. Benson, J. Golden, . . . J.A. Clements, *Studies of the structure of lung surfactant protein SP-A.* Am J Physiol, 1989. **257**(6 Pt 1): p. L421-9.
341. Joyeux, L., E. Danzer, M.P. Limberis, P.W. Zoltick, A. Radu, A.W. Flake, and M.G. Davey, *In utero lung gene transfer using adeno-associated viral and lentiviral vectors in mice.* Hum Gene Ther Methods, 2014. **25**(3): p. 197-205.
342. Auricchio, A., E. O'Connor, D. Weiner, G.P. Gao, M. Hildinger, L. Wang, . . . J.M. Wilson, *Noninvasive gene transfer to the lung for systemic delivery of therapeutic proteins.* J Clin Invest, 2002. **110**(4): p. 499-504.
343. Crapo, J.D., B.E. Barry, P. Gehr, M. Bachofen, and E.R. Weibel, *Cell number and cell characteristics of the normal human lung.* Am Rev Respir Dis, 1982. **126**(2): p. 332-7.
344. Munis, A.M., S.C. Hyde, and D.R. Gill, *A human surfactant B deficiency air-liquid interface cell culture model suitable for gene therapy applications.* Mol Ther Methods Clin Dev, 2021. **20**: p. 237-246.
345. Mauro, V.P., *Codon Optimization in the Production of Recombinant Biotherapeutics: Potential Risks and Considerations.* BioDrugs, 2018. **32**(1): p. 69-81.
346. Mauro, V.P. and S.A. Chappell, *A critical analysis of codon optimization in human therapeutics.* Trends Mol Med, 2014. **20**(11): p. 604-13.
347. Fehrholz, M., S. Seidenspinner, and S. Kunzmann, *Expression of surfactant protein B is dependent on cell density in H441 lung epithelial cells.* PLoS One, 2017. **12**(9): p. e0184556.
348. Maunder, H.E., J. Wright, B.R. Kolli, C.R. Vieira, T.T. Mkandawire, S. Tatoris, . . . D.C. Farley, *Enhancing titres of therapeutic viral vectors using the transgene repression in vector production (TRiP) system.* Nat Commun, 2017. **8**: p. 14834.
349. Degiulio, J.V., C.D. Kaufman, and D.A. Dean, *The SP-C promoter facilitates alveolar type II epithelial cell-specific plasmid nuclear import and gene expression.* Gene Ther, 2010. **17**(4): p. 541-9.
350. Liu, Y., V.S. Kumar, W. Zhang, J. Rehman, and A.B. Malik, *Activation of type II cells into regenerative stem cell antigen-1(+) cells during alveolar repair.* Am J Respir Cell Mol Biol, 2015. **53**(1): p. 113-24.
351. Canadas, O., B. Olmeda, A. Alonso, and J. Perez-Gil, *Lipid-Protein and Protein-Protein Interactions in the Pulmonary Surfactant System and Their Role in Lung Homeostasis.* Int J Mol Sci, 2020. **21**(10).

352. Loring, H.S., M.K. ElMallah, and T.R. Flotte, *Development of rAAV2-CFTR: History of the First rAAV Vector Product to be Used in Humans.* Hum Gene Ther Methods, 2016. **27**(2): p. 49-58.
353. Ibricevic, A., A. Pekosz, M.J. Walter, C. Newby, J.T. Battaile, E.G. Brown, . . . S.L. Brody, *Influenza virus receptor specificity and cell tropism in mouse and human airway epithelial cells.* J Virol, 2006. **80**(15): p. 7469-80.
354. Nicholls, J.M., A.J. Bourne, H. Chen, Y. Guan, and J.S. Peiris, *Sialic acid receptor detection in the human respiratory tract: evidence for widespread distribution of potential binding sites for human and avian influenza viruses.* Respir Res, 2007. **8**: p. 73.
355. Mao, P., S. Wu, J. Li, W. Fu, W. He, X. Liu, . . . Y. Li, *Human alveolar epithelial type II cells in primary culture.* Physiol Rep, 2015. **3**(2).
356. Lancaster, M.A. and J.A. Knoblich, *Organogenesis in a dish: modeling development and disease using organoid technologies.* Science, 2014. **345**(6194): p. 1247125.
357. Barkauskas, C.E., M.I. Chung, B. Fioret, X. Gao, H. Katsura, and B.L. Hogan, *Lung organoids: current uses and future promise.* Development, 2017. **144**(6): p. 986-997.
358. Barkauskas, C.E., M.J. Cronce, C.R. Rackley, E.J. Bowie, D.R. Keene, B.R. Stripp, . . . B.L. Hogan, *Type 2 alveolar cells are stem cells in adult lung.* J Clin Invest, 2013. **123**(7): p. 3025-36.
359. Gonzalez, R.F., L. Allen, L. Gonzales, P.L. Ballard, and L.G. Dobbs, *HTII-280, a biomarker specific to the apical plasma membrane of human lung alveolar type II cells.* J Histochem Cytochem, 2010. **58**(10): p. 891-901.
360. Gotoh, S., I. Ito, T. Nagasaki, Y. Yamamoto, S. Konishi, Y. Korogi, . . . M. Mishima, *Generation of alveolar epithelial spheroids via isolated progenitor cells from human pluripotent stem cells.* Stem Cell Reports, 2014. **3**(3): p. 394-403.
361. Dye, B.R., D.R. Hill, M.A. Ferguson, Y.H. Tsai, M.S. Nagy, R. Dyal, . . . J.R. Spence, *In vitro generation of human pluripotent stem cell derived lung organoids.* Elife, 2015. **4**.
362. Jain, R., C.E. Barkauskas, N. Takeda, E.J. Bowie, H. Aghajanian, Q. Wang, . . . J.A. Epstein, *Plasticity of Hopx(+) type I alveolar cells to regenerate type II cells in the lung.* Nat Commun, 2015. **6**: p. 6727.
363. Whitsett, J.A., S.E. Wert, and T.E. Weaver, *Alveolar surfactant homeostasis and the pathogenesis of pulmonary disease.* Annu Rev Med, 2010. **61**: p. 105-19.
364. Stahlman, M.T., M.P. Gray, M.W. Falconieri, J.A. Whitsett, and T.E. Weaver, *Lamellar body formation in normal and surfactant protein B-deficient fetal mice.* Lab Invest, 2000. **80**(3): p. 395-403.
365. Dye, B.R., P.H. Dedhia, A.J. Miller, M.S. Nagy, E.S. White, L.D. Shea, and J.R. Spence, *A bioengineered niche promotes in vivo engraftment and maturation of pluripotent stem cell derived human lung organoids.* Elife, 2016. **5**.
366. Dye, B.R., A.J. Miller, and J.R. Spence, *How to Grow a Lung: Applying Principles of Developmental Biology to Generate Lung Lineages from Human Pluripotent Stem Cells.* Curr Pathobiol Rep, 2016. **4**: p. 47-57.
367. Morrisey, E.E. and B.L. Hogan, *Preparing for the first breath: genetic and cellular mechanisms in lung development.* Dev Cell, 2010. **18**(1): p. 8-23.
368. Iliodromiti, Z., D. Zygouris, S. Sifakis, K.I. Pappa, P. Tsikouras, N. Salakos, . . . N. Vrachnis, *Acute lung injury in preterm fetuses and neonates: mechanisms*

and molecular pathways. J Matern Fetal Neonatal Med, 2013. **26**(17): p. 1696-704.
369. Ahmed, E., C. Sansac, S. Assou, D. Gras, A. Petit, I. Vachier, . . . A. Bourdin, *Lung development, regeneration and plasticity: From disease physiopathology to drug design using induced pluripotent stem cells.* Pharmacol Ther, 2018. **183**: p. 58-77.
370. Hussain, M., C. Xu, M. Ahmad, Y. Yang, M. Lu, X. Wu, . . . X. Wu, *Notch Signaling: Linking Embryonic Lung Development and Asthmatic Airway Remodeling.* Mol Pharmacol, 2017. **92**(6): p. 676-693.
371. Herriges, M. and E.E. Morrisey, *Lung development: orchestrating the generation and regeneration of a complex organ.* Development, 2014. **141**(3): p. 502-13.
372. Liu, G.H., C. Betts, D.M. Cunoosamy, P.M. Aberg, J.J. Hornberg, K.B. Sivars, and T.S. Cohen, *Use of precision cut lung slices as a translational model for the study of lung biology.* Respiratory Research, 2019. **20**.
373. Sanderson, M.J., *Exploring lung physiology in health and disease with lung slices.* Pulm Pharmacol Ther, 2011. **24**(5): p. 452-65.
374. Temann, A., T. Golovina, V. Neuhaus, C. Thompson, J.A. Chichester, A. Braun, and V. Yusibov, *Evaluation of inflammatory and immune responses in long-term cultured human precision-cut lung slices.* Hum Vaccin Immunother, 2017. **13**(2): p. 351-358.
375. Krimmling, T. and C. Schwegmann-Wessels, *Comparison of mono- and co-infection by swine influenza A viruses and porcine respiratory coronavirus in porcine precision-cut lung slices.* Res Vet Sci, 2017. **115**: p. 470-477.
376. Goris, K., S. Uhlenbruck, C. Schwegmann-Wessels, W. Kohl, F. Niedorf, M. Stern, . . . G. Herrler, *Differential sensitivity of differentiated epithelial cells to respiratory viruses reveals different viral strategies of host infection.* J Virol, 2009. **83**(4): p. 1962-8.
377. Zhu, Z.B., A.A. Rivera, S.K. Makhija, B. Lu, M. Wang, M. Izumi, . . . D.T. Curiel, *Targeting lung cancer using an infectivity enhanced CXCR4-CRAd.* Lung Cancer, 2007. **55**(2): p. 145-56.
378. Huang, S.X., M.D. Green, A.T. de Carvalho, M. Mumau, Y.W. Chen, S.L. D'Souza, and H.W. Snoeck, *The in vitro generation of lung and airway progenitor cells from human pluripotent stem cells.* Nat Protoc, 2015. **10**(3): p. 413-25.
379. Kubo, A., K. Shinozaki, J.M. Shannon, V. Kouskoff, M. Kennedy, S. Woo, . . . G. Keller, *Development of definitive endoderm from embryonic stem cells in culture.* Development, 2004. **131**(7): p. 1651-62.
380. Geisler, C. and D.L. Jarvis, *Effective glycoanalysis with Maackia amurensis lectins requires a clear understanding of their binding specificities.* Glycobiology, 2011. **21**(8): p. 988-93.
381. Haeger, S.M., Y. Yang, and E.P. Schmidt, *Heparan Sulfate in the Developing, Healthy, and Injured Lung.* Am J Respir Cell Mol Biol, 2016. **55**(1): p. 5-11.
382. Hoffmann, M., H. Kleine-Weber, S. Schroeder, N. Kruger, T. Herrler, S. Erichsen, . . . S. Pohlmann, *SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor.* Cell, 2020. **181**(2): p. 271-280 e8.
383. Neuhaus, V., D. Schaudien, T. Golovina, U.A. Temann, C. Thompson, T. Lippmann, . . . A. Braun, *Assessment of long-term cultivated human precision-*

cut lung slices as an ex vivo system for evaluation of chronic cytotoxicity and functionality. J Occup Med Toxicol, 2017. **12**: p. 13.
384. Bhowmick, R. and H. Gappa-Fahlenkamp, *Cells and Culture Systems Used to Model the Small Airway Epithelium.* Lung, 2016. **194**(3): p. 419-28.
385. Haeger, S.M., X. Liu, X. Han, J.B. McNeil, K. Oshima, S.A. McMurtry, . . . E.P. Schmidt, *Epithelial Heparan Sulfate Contributes to Alveolar Barrier Function and Is Shed during Lung Injury.* Am J Respir Cell Mol Biol, 2018. **59**(3): p. 363-374.
386. Westergren-Thorsson, G., U. Hedstrom, A. Nybom, E. Tykesson, E. Ahrman, M. Hornfelt, . . . O. Hallgren, *Increased deposition of glycosaminoglycans and altered structure of heparan sulfate in idiopathic pulmonary fibrosis.* Int J Biochem Cell Biol, 2017. **83**: p. 27-38.
387. Cerna, A., P. Janega, P. Martanovic, M. Lisy, and P. Babal, *Changes in sialic acid expression in the lung during intrauterine development of the human fetus.* Acta Histochem, 2002. **104**(4): p. 339-42.
388. Vincent, M.C., B.C. Trapnell, R.P. Baughman, S.E. Wert, J.A. Whitsett, and H.S. Iwamoto, *Adenovirus-mediated gene transfer to the respiratory tract of fetal sheep in utero.* Hum Gene Ther, 1995. **6**(8): p. 1019-28.
389. Markwell, M.A. and J.C. Paulson, *Sendai virus utilizes specific sialyloligosaccharides as host cell receptor determinants.* Proc Natl Acad Sci U S A, 1980. **77**(10): p. 5693-7.
390. Horowitz, S., R.H. Watkins, R.L. Auten, Jr., C.E. Mercier, and E.R. Cheng, *Differential accumulation of surfactant protein A, B, and C mRNAs in two epithelial cell types of hyperoxic lung.* Am J Respir Cell Mol Biol, 1991. **5**(6): p. 511-5.
391. Mead, B.E. and J.M. Karp, *All models are wrong, but some organoids may be useful.* Genome Biol, 2019. **20**(1): p. 66.
392. Concordet, J.P. and M. Haeussler, *CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens.* Nucleic Acids Res, 2018. **46**(W1): p. W242-W245.
393. Doench, J.G., N. Fusi, M. Sullender, M. Hegde, E.W. Vaimberg, K.F. Donovan, . . . D.E. Root, *Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9.* Nat Biotechnol, 2016. **34**(2): p. 184-191.
394. Martin, R.M., K. Ikeda, M.K. Cromer, N. Uchida, T. Nishimura, R. Romano, . . . M.H. Porteus, *Highly Efficient and Marker-free Genome Editing of Human Pluripotent Stem Cells by CRISPR-Cas9 RNP and AAV6 Donor-Mediated Homologous Recombination.* Cell Stem Cell, 2019. **24**(5): p. 821-828 e5.
395. Brasch, F., M. Ochs, T. Kahne, S. Guttentag, V. Schauer-Vukasinovic, M. Derrick, . . . F. Buhling, *Involvement of napsin A in the C- and N-terminal processing of surfactant protein B in type-II pneumocytes of the human lung.* J Biol Chem, 2003. **278**(49): p. 49006-14.
396. Korogi, Y., S. Gotoh, S. Ikeo, Y. Yamamoto, N. Sone, K. Tamai, . . . T. Hirai, *In Vitro Disease Modeling of Hermansky-Pudlak Syndrome Type 2 Using Human Induced Pluripotent Stem Cell-Derived Alveolar Organoids.* Stem Cell Reports, 2019. **12**(3): p. 431-440.
397. Kim, S., D. Kim, S.W. Cho, J. Kim, and J.S. Kim, *Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins.* Genome Res, 2014. **24**(6): p. 1012-9.

398. Danner, E., S. Bashir, S. Yumlu, W. Wurst, B. Wefers, and R. Kuhn, *Control of gene editing by manipulation of DNA repair mechanisms.* Mamm Genome, 2017. **28**(7-8): p. 262-274.
399. Aitken, M.L., R.B. Moss, D.A. Waltz, M.E. Dovey, M.R. Tonelli, S.C. McNamara, . . . T.C. Reynolds, *A phase I study of aerosolized administration of tgAAVCF to cystic fibrosis subjects with mild lung disease.* Hum Gene Ther, 2001. **12**(15): p. 1907-16.
400. Mercuri, E., M.C. Pera, M. Scoto, R. Finkel, and F. Muntoni, *Spinal muscular atrophy - insights and challenges in the treatment era.* Nat Rev Neurol, 2020. **16**(12): p. 706-715.
401. Mendell, J.R., S.A. Al-Zaidy, K.J. Lehman, M. McColly, L.P. Lowes, L.N. Alfano, . . . R. Shell, *Five-Year Extension Results of the Phase 1 START Trial of Onasemnogene Abeparvovec in Spinal Muscular Atrophy.* JAMA Neurol, 2021.
402. Mendell, J.R., S. Al-Zaidy, R. Shell, W.D. Arnold, L.R. Rodino-Klapac, T.W. Prior, . . . B.K. Kaspar, *Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy.* N Engl J Med, 2017. **377**(18): p. 1713-1722.
403. Chen, O.G., S.E. Mather, C.M. Brommel, B.A. Hamilton, A. Ehler, R. Villacreses, . . . X. Li, *Transduction of Pig Small Airway Epithelial Cells and Distal Lung Progenitor Cells by AAV4.* Cells, 2021. **10**(5).
404. Kotterman, M.A. and D.V. Schaffer, *Engineering adeno-associated viruses for clinical gene therapy.* Nat Rev Genet, 2014. **15**(7): p. 445-51.
405. Davidsson, M., G. Wang, P. Aldrin-Kirk, T. Cardoso, S. Nolbrant, M. Hartnor, . . . T. Bjorklund, *A systematic capsid evolution approach performed in vivo for the design of AAV vectors with tailored properties and tropism.* Proc Natl Acad Sci U S A, 2019.
406. Marques, A.D., M. Kummer, O. Kondratov, A. Banerjee, O. Moskalenko, and S. Zolotukhin, *Applying machine learning to predict viral assembly for adeno-associated virus capsid libraries.* Mol Ther Methods Clin Dev, 2021. **20**: p. 276-286.
407. Miller, A.J., B.R. Dye, D. Ferrer-Torres, D.R. Hill, A.W. Overeem, L.D. Shea, and J.R. Spence, *Generation of lung organoids from human pluripotent stem cells in vitro.* Nat Protoc, 2019. **14**(2): p. 518-540.
408. Huang, S.X., M.N. Islam, J. O'Neill, Z. Hu, Y.G. Yang, Y.W. Chen, . . . H.W. Snoeck, *Efficient generation of lung and airway epithelial cells from human pluripotent stem cells.* Nat Biotechnol, 2014. **32**(1): p. 84-91.
409. Miller, A.J., D.R. Hill, M.S. Nagy, Y. Aoki, B.R. Dye, A.M. Chin, . . . J.R. Spence, *In Vitro Induction and In Vivo Engraftment of Lung Bud Tip Progenitor Cells Derived from Human Pluripotent Stem Cells.* Stem Cell Reports, 2018. **10**(1): p. 101-119.
410. Leibel, S.L., R.N. McVicar, A.M. Winquist, and E.Y. Snyder, *Generation of 3D Whole Lung Organoids from Induced Pluripotent Stem Cells for Modeling Lung Developmental Biology and Disease.* J Vis Exp, 2021(170).
411. Heo, H.R., J. Kim, W.J. Kim, S.R. Yang, S.S. Han, S.J. Lee, . . . S.H. Hong, *Human pluripotent stem cell-derived alveolar epithelial cells are alternatives for in vitro pulmotoxicity assessment.* Sci Rep, 2019. **9**(1): p. 505.
412. Kim, J.H., G.H. An, J.Y. Kim, R. Rasaei, W.J. Kim, X. Jin, . . . S.H. Hong, *Human pluripotent stem-cell-derived alveolar organoids for modeling pulmonary fibrosis and drug testing.* Cell Death Discov, 2021. **7**(1): p. 48.

413. Lamers, M.M., J. van der Vaart, K. Knoops, S. Riesebosch, T.I. Breugem, A.Z. Mykytyn, . . . B.L. Haagmans, *An organoid-derived bronchioalveolar model for SARS-CoV-2 infection of human alveolar type II-like cells.* EMBO J, 2021. **40**(5): p. e105912.
414. Lemaitre, J., T. Naninck, B. Delache, J. Creppy, P. Huber, M. Holzapfel, . . . P. Maisonnasse, *Non-human primate models of human respiratory infections.* Mol Immunol, 2021. **135**: p. 147-164.
415. Kitano, M., Y. Itoh, M. Kodama, H. Ishigaki, M. Nakayama, T. Nagata, . . . K. Ogasawara, *Establishment of a cynomolgus macaque model of influenza B virus infection.* Virology, 2010. **407**(2): p. 178-84.
416. Cross, R.W., K.N. Agans, A.N. Prasad, V. Borisevich, C. Woolsey, D.J. Deer, . . . T.W. Geisbert, *Intranasal exposure of African green monkeys to SARS-CoV-2 results in acute phase pneumonia with shedding and lung injury still present in the early convalescence phase.* Virol J, 2020. **17**(1): p. 125.
417. Babu, P.G., A. Selvan, S. Christuraj, J. David, T.J. John, and E.A. Simoes, *A primate model of respiratory syncytial virus infection.* Indian J Exp Biol, 1998. **36**(8): p. 758-62.
418. Chen, J., F. Zeng, S.J. Forrester, S. Eguchi, M.Z. Zhang, and R.C. Harris, *Expression and Function of the Epidermal Growth Factor Receptor in Physiology and Disease.* Physiol Rev, 2016. **96**(3): p. 1025-1069.
419. Plopper, C.G., J.A. St George, L.C. Read, S.J. Nishio, A.J. Weir, L. Edwards, . . . et al., *Acceleration of alveolar type II cell differentiation in fetal rhesus monkey lung by administration of EGF.* Am J Physiol, 1992. **262**(3 Pt 1): p. L313-21.
420. Miettinen, P.J., D. Warburton, D. Bu, J.S. Zhao, J.E. Berger, P. Minoo, . . . R. Derynck, *Impaired lung branching morphogenesis in the absence of functional EGF receptor.* Dev Biol, 1997. **186**(2): p. 224-36.
421. Playford, R.J., A.M. Hanby, S. Gschmeissner, L.P. Peiffer, N.A. Wright, and T. McGarrity, *The epidermal growth factor receptor (EGF-R) is present on the basolateral, but not the apical, surface of enterocytes in the human gastrointestinal tract.* Gut, 1996. **39**(2): p. 262-6.
422. Katz, M.G., A.S. Fargnoli, S.M. Gubara, K. Fish, T. Weber, C.R. Bridges, . . . K. Ishikawa, *Targeted Gene Delivery through the Respiratory System: Rationale for Intratracheal Gene Transfer.* J Cardiovasc Dev Dis, 2019. **6**(1).
423. Cooney, A.L. and P.L. Sinn, *Intratracheal aerosolization of viral vectors to newborn pig airways.* Biotechniques, 2020. **68**(5): p. 235-239.
424. Bernardino de la Serna, J., J. Perez-Gil, A.C. Simonsen, and L.A. Bagatolli, *Cholesterol rules: direct observation of the coexistence of two fluid phases in native pulmonary surfactant membranes at physiological temperatures.* J Biol Chem, 2004. **279**(39): p. 40715-22.
425. Zhao, J., Y. Yue, A. Patel, L. Wasala, J.F. Karp, K. Zhang, . . . Y. Lai, *High-Resolution Histological Landscape of AAV DNA Distribution in Cellular Compartments and Tissues following Local and Systemic Injection.* Mol Ther Methods Clin Dev, 2020. **18**: p. 856-868.
426. Strayer, M., R.C. Savani, L.W. Gonzales, A. Zaman, Z. Cui, E. Veszelovszky, . . . P.L. Ballard, *Human surfactant protein B promoter in transgenic mice: temporal, spatial, and stimulus-responsive regulation.* Am J Physiol Lung Cell Mol Physiol, 2002. **282**(3): p. L394-404.

427. Strayer, M.S., S.H. Guttentag, and P.L. Ballard, *Targeting type II and Clara cells for adenovirus-mediated gene transfer using the surfactant protein B promoter.* Am J Respir Cell Mol Biol, 1998. **18**(1): p. 1-11.
428. Szegezdi, E., S.E. Logue, A.M. Gorman, and A. Samali, *Mediators of endoplasmic reticulum stress-induced apoptosis.* EMBO Rep, 2006. **7**(9): p. 880-5.
429. Kennedy, D., A. Samali, and R. Jager, *Methods for studying ER stress and UPR markers in human cells.* Methods Mol Biol, 2015. **1292**: p. 3-18.
430. Gill, D.R., S.E. Smyth, C.A. Goddard, I.A. Pringle, C.F. Higgins, W.H. Colledge, and S.C. Hyde, *Increased persistence of lung gene expression using plasmids containing the ubiquitin C or elongation factor 1 alpha promoter.* Gene Therapy, 2001. **8**(20): p. 1539-1546.
431. Pringle, I.A., S. Raman, W.W. Sharp, S.H. Cheng, S.C. Hyde, and D.R. Gill, *Detection of plasmid DNA vectors following gene transfer to the murine airways.* Gene Ther, 2005. **12**(15): p. 1206-14.
432. Zabner, J., S.C. Wadsworth, A.E. Smith, and M.J. Welsh, *Adenovirus-mediated generation of cAMP-stimulated Cl- transport in cystic fibrosis airway epithelia in vitro: effect of promoter and administration method.* Gene Ther, 1996. **3**(5): p. 458-65.
433. Brooks, A.R., R.N. Harkins, P. Wang, H.S. Qian, P. Liu, and G.M. Rubanyi, *Transcriptional silencing is associated with extensive methylation of the CMV promoter following adenoviral gene delivery to muscle.* J Gene Med, 2004. **6**(4): p. 395-404.
434. Chen, P., J. Tian, I. Kovesdi, and J.T. Bruder, *Promoters influence the kinetics of transgene expression following adenovector gene delivery.* J Gene Med, 2008. **10**(2): p. 123-31.
435. Liu, B.H., X. Wang, Y.X. Ma, and S. Wang, *CMV enhancer/human PDGF-beta promoter for neuron-specific transgene expression.* Gene Ther, 2004. **11**(1): p. 52-60.
436. Naldini, L., U. Blomer, F.H. Gage, D. Trono, and I.M. Verma, *Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector.* Proc Natl Acad Sci U S A, 1996. **93**(21): p. 11382-8.
437. Verdera, H.C., K. Kuranda, and F. Mingozzi, *AAV Vector Immunogenicity in Humans: A Long Journey to Successful Gene Transfer.* Mol Ther, 2020. **28**(3): p. 723-746.
438. Mingozzi, F., Y. Chen, S.C. Edmonson, S. Zhou, R.M. Thurlings, P.P. Tak, . . . M.J. Vervoordeldonk, *Prevalence and pharmacological modulation of humoral immunity to AAV vectors in gene transfer to synovial tissue.* Gene Ther, 2013. **20**(4): p. 417-24.
439. de Back, D.Z., M.M. Neyrinck, and H. Vrielink, *Therapeutic plasma apheresis: Expertise and indications.* Transfus Apher Sci, 2019. **58**(3): p. 254-257.
440. Bertin, B., P. Veron, C. Leborgne, J.Y. Deschamps, S. Moullec, Y. Fromes, . . . F. Mingozzi, *Capsid-specific removal of circulating antibodies to adeno-associated virus vectors.* Sci Rep, 2020. **10**(1): p. 864.
441. Chicoine, L.G., C.L. Montgomery, W.G. Bremer, K.M. Shontz, D.A. Griffin, K.N. Heller, . . . L.R. Rodino-Klapac, *Plasmapheresis eliminates the negative impact of AAV antibodies on microdystrophin gene expression following vascular delivery.* Mol Ther, 2014. **22**(2): p. 338-347.

442. Orlowski, A., M.G. Katz, S.M. Gubara, A.S. Fargnoli, K.M. Fish, and T. Weber, *Successful Transduction with AAV Vectors after Selective Depletion of Anti-AAV Antibodies by Immunoadsorption.* Mol Ther Methods Clin Dev, 2020. **16**: p. 192-203.
443. Luisetti, M., G. Rodi, C. Perotti, I. Campo, F. Mariani, E. Pozzi, and B.C. Trapnell, *Plasmapheresis for treatment of pulmonary alveolar proteinosis.* Eur Respir J, 2009. **33**(5): p. 1220-2.
444. Albrecht, M. and P.C. Arck, *Vertically Transferred Immunity in Neonates: Mothers, Mechanisms and Mediators.* Front Immunol, 2020. **11**: p. 555.
445. Pou, C., D. Nkulikiyimfura, E. Henckel, A. Olin, T. Lakshmikanth, J. Mikes, . . . P. Brodin, *The repertoire of maternal anti-viral antibodies in human newborns.* Nat Med, 2019. **25**(4): p. 591-596.
446. Conway, S.P., P.R. Dear, and I. Smith, *Immunoglobulin profile of the preterm baby.* Arch Dis Child, 1985. **60**(3): p. 208-12.
447. Day, J.W., R.S. Finkel, E. Mercuri, K.J. Swoboda, M. Menier, R. van Olden, . . . J.R. Mendell, *Adeno-associated virus serotype 9 antibodies in patients screened for treatment with onasemnogene abeparvovec.* Mol Ther Methods Clin Dev, 2021. **21**: p. 76-82.
448. Tang, Y., Z. Yan, S. Lin, E.D. Huntemann, Z. Feng, S.Y. Park, . . . J.F. Engelhardt, *Repeat Dosing of AAV2.5T to Ferret Lungs Elicits an Antibody Response That Diminishes Transduction in an Age-Dependent Manner.* Mol Ther Methods Clin Dev, 2020. **19**: p. 186-200.
449. Ronzitti, G., D.A. Gross, and F. Mingozzi, *Human Immune Responses to Adeno-Associated Virus (AAV) Vectors.* Front Immunol, 2020. **11**: p. 670.
450. Meliani, A., F. Boisgerault, R. Hardet, S. Marmier, F. Collaud, G. Ronzitti, . . . F. Mingozzi, *Antigen-selective modulation of AAV immunogenicity with tolerogenic rapamycin nanoparticles enables successful vector re-administration.* Nat Commun, 2018. **9**(1): p. 4098.
451. Fischer, A.C., S.E. Beck, C.I. Smith, B.L. Laube, F.B. Askin, S.E. Guggino, . . . W.B. Guggino, *Successful transgene expression with serial doses of aerosolized rAAV2 vectors in rhesus macaques.* Mol Ther, 2003. **8**(6): p. 918-26.
452. Desai, T.J., D.G. Brownfield, and M.A. Krasnow, *Alveolar progenitor and stem cells in lung development, renewal and cancer.* Nature, 2014. **507**(7491): p. 190-+.
453. Chen, Q. and Y. Liu, *Heterogeneous groups of alveolar type II cells in lung homeostasis and repair.* Am J Physiol Cell Physiol, 2020. **319**(6): p. C991-C996.
454. Chen, Q., V. Suresh Kumar, J. Finn, D. Jiang, J. Liang, Y.Y. Zhao, and Y. Liu, *CD44(high) alveolar type II cells show stem cell properties during steady-state alveolar homeostasis.* Am J Physiol Lung Cell Mol Physiol, 2017. **313**(1): p. L41-L51.
455. Zacharias, W.J., D.B. Frank, J.A. Zepp, M.P. Morley, F.A. Alkhaleel, J. Kong, . . . E.E. Morrisey, *Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor.* Nature, 2018.
456. Travaglini, K.J., A.N. Nabhan, L. Penland, R. Sinha, A. Gillich, R.V. Sit, . . . M.A. Krasnow, *A molecular cell atlas of the human lung from single-cell RNA sequencing.* Nature, 2020. **587**(7835): p. 619-625.
457. Whitsett, J.A., T.V. Kalin, Y. Xu, and V.V. Kalinichenko, *Building and Regenerating the Lung Cell by Cell.* Physiol Rev, 2019. **99**(1): p. 513-554.

458. Wansleeben, C., C.E. Barkauskas, J.R. Rock, and B.L. Hogan, *Stem cells of the adult lung: their development and role in homeostasis, regeneration, and disease.* Wiley Interdiscip Rev Dev Biol, 2013. **2**(1): p. 131-48.
459. Rawlins, E.L. and B.L. Hogan, *Ciliated epithelial cell lifespan in the mouse trachea and lung.* Am J Physiol Lung Cell Mol Physiol, 2008. **295**(1): p. L231-4.
460. Liu, X., M. Luo, C. Guo, Z. Yan, Y. Wang, D.C. Lei-Butters, and J.F. Engelhardt, *Analysis of adeno-associated virus progenitor cell transduction in mouse lung.* Mol Ther, 2009. **17**(2): p. 285-93.
461. Wells, A.B., *The kinetics of cell proliferation in the tracheobronchial epithelia of rats with and without chronic respiratory disease.* Cell Tissue Kinet, 1970. **3**(2): p. 185-206.
462. Leigh, M.W., J.E. Kylander, J.R. Yankaskas, and R.C. Boucher, *Cell proliferation in bronchial epithelium and submucosal glands of cystic fibrosis patients.* Am J Respir Cell Mol Biol, 1995. **12**(6): p. 605-12.
463. Liao, D. and H. Li, *Dissecting the Niche for Alveolar Type II Cells With Alveolar Organoids.* Front Cell Dev Biol, 2020. **8**: p. 419.
464. Eiby, Y.A., L.L. Wright, V.P. Kalanjati, S.M. Miller, S.T. Bjorkman, H.L. Keates, . . . B.E. Lingwood, *A pig model of the preterm neonate: anthropometric and physiological characteristics.* PLoS One, 2013. **8**(7): p. e68763.
465. Rey-Santano, C., V.E. Mielgo, M.A. Gomez-Solaetxe, F. Ricci, F. Bianco, F. Salomone, . . . Y.G.J.L. de Heredia, *Non-invasive ventilation and surfactant treatment as the primary mode of respiratory support in surfactant-deficient newborn piglets.* Pediatr Res, 2018.
466. Petris, G., A. Casini, C. Montagna, F. Lorenzin, D. Prandi, A. Romanel, . . . A. Cereseto, *Hit and go CAS9 delivered through a lentiviral based self-limiting circuit.* Nat Commun, 2017. **8**: p. 15334.
467. Charlesworth, C.T., P.S. Deshpande, D.P. Dever, J. Camarena, V.T. Lemgart, M.K. Cromer, . . . M.H. Porteus, *Identification of preexisting adaptive immunity to Cas9 proteins in humans.* Nat Med, 2019. **25**(2): p. 249-254.
468. Wagner, D.L., L. Amini, D.J. Wendering, L.M. Burkhardt, L. Akyuz, P. Reinke, . . . M. Schmueck-Henneresse, *High prevalence of Streptococcus pyogenes Cas9-reactive T cells within the adult human population.* Nat Med, 2019. **25**(2): p. 242-248.
469. Bischoff, N., S. Wimberger, M. Maresca, and C. Brakebusch, *Improving Precise CRISPR Genome Editing by Small Molecules: Is there a Magic Potion?* Cells, 2020. **9**(5).
470. Suzuki, K., Y. Tsunekawa, R. Hernandez-Benitez, J. Wu, J. Zhu, E.J. Kim, . . . J.C. Belmonte, *In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration.* Nature, 2016. **540**(7631): p. 144-149.
471. Suzuki, K. and J.C.I. Belmonte, *In vivo genome editing via the HITI method as a tool for gene therapy.* Journal of Human Genetics, 2018. **63**(2): p. 157-164.
472. Lu, Y., J. Xue, T. Deng, X. Zhou, K. Yu, L. Deng, . . . T. Mok, *Safety and feasibility of CRISPR-edited T cells in patients with refractory non-small-cell lung cancer.* Nat Med, 2020. **26**(5): p. 732-740.
473. Xu, L., J. Wang, Y. Liu, L. Xie, B. Su, D. Mou, . . . H. Chen, *CRISPR-Edited Stem Cells in a Patient with HIV and Acute Lymphocytic Leukemia.* N Engl J Med, 2019. **381**(13): p. 1240-1247.

474. He, S., *The first human trial of CRISPR-based cell therapy clears safety concerns as new treatment for late-stage lung cancer.* Signal Transduct Target Ther, 2020. **5**(1): p. 168.
475. *First systemic CRISPR agent in humans.* Nat Biotechnol, 2020. **38**(12): p. 1364.
476. Anzalone, A.V., P.B. Randolph, J.R. Davis, A.A. Sousa, L.W. Koblan, J.M. Levy, . . . D.R. Liu, *Search-and-replace genome editing without double-strand breaks or donor DNA.* Nature, 2019. **576**(7785): p. 149-157.
477. Yeh, C.D., C.D. Richardson, and J.E. Corn, *Advances in genome editing through control of DNA repair pathways.* Nat Cell Biol, 2019. **21**(12): p. 1468-1478.
478. Knott, G.J. and J.A. Doudna, *CRISPR-Cas guides the future of genetic engineering.* Science, 2018. **361**(6405): p. 866-869.
479. Yan, Z., D.C. Lei-Butters, N.W. Keiser, and J.F. Engelhardt, *Distinct transduction difference between adeno-associated virus type 1 and type 6 vectors in human polarized airway epithelia.* Gene Ther, 2013. **20**(3): p. 328-37.
480. Furuta-Hanawa, B., T. Yamaguchi, and E. Uchida, *Two-Dimensional Droplet Digital PCR as a Tool for Titration and Integrity Evaluation of Recombinant Adeno-Associated Viral Vectors.* Hum Gene Ther Methods, 2019. **30**(4): p. 127-136.
481. Brooks, P.J., E.A. Ottinger, D. Portero, R.M. Lomash, A. Alimardanov, P. Terse, . . . D.C. Lo, *The Platform Vector Gene Therapies Project: Increasing the Efficiency of Adeno-Associated Virus Gene Therapy Clinical Trial Startup.* Hum Gene Ther, 2020. **31**(19-20): p. 1034-1042.
482. Nemeth, J., A. Schundner, and M. Frick, *Insights Into Development and Progression of Idiopathic Pulmonary Fibrosis From Single Cell RNA Studies.* Front Med (Lausanne), 2020. **7**: p. 611728.
483. Carneiro, A., H. Lee, L. Lin, J. van Haasteren, and D.V. Schaffer, *Novel Lung Tropic Adeno-Associated Virus Capsids for Therapeutic Gene Delivery.* Hum Gene Ther, 2020. **31**(17-18): p. 996-1009.