

University of Nottingham

UK | CHINA | MALAYSIA

Developing a chemically defined, xeno-free platform to differentiate human induced pluripotent stem cells to type 2 alveolar epithelial cells of the lung

Submitted by:

Liam Allen Reed

MSc, BSc

School of Medicine

Division of Cancer and Stem Cells

The University of Nottingham

Principal supervisor: Associate professor Nicholas R.F. Hannan

Secondary Supervisor: Professor Cathy Merry

October 2023

Thesis submitted to The University of Nottingham for the

Degree of Doctor in Philosophy

Acknowledgements

I would first like to thank my principal supervisor Associate Professor Nicholas R.F. Hannan for his continued guidance, support, and encouragement throughout this project. His ever-calm temperament and willingness to help was crucial in sustaining my motivation during some of the more challenging times. I would like to express my appreciation to my secondary supervisor Professor Cathy Merry for her expertise, knowledge and for always being there to listen. I also want to thank my co-supervisor, Dr Sara Cuevas Ocaña, for her support, advice, the scientific conversations that continued to motivate me and her significant contribution to the project.

I would like to show my gratitude to the flow cytometry team, Dr David Onion, Dr Nicola Croxall, and Maria Haig for their willingness to always help and support along the project. I would like to thank all of the Biodiscovery Institute staff for enabling this work to happen, especially Paula Olsen who was always there to help with almost anything.

I was fortunate to be part of a wonderful research group, and I want to thank all the past and current members of the Hannan Lab for their unwavering support. To Dr Peggy Lo, for her positivity and encouragement at the start of my PhD. I would like to express my gratitude to Carlos Sainz Zuñiga for all of his help collecting data and processing all of the RNA-seq. To Ana Lilia Serna Valverde, who was there for me no matter what, every step of the journey, during the best and worst moments. I will be forever grateful for that.

To the people that became close friends, Paulina Durczak, Peggy Lo, Analy Serna, Kate Vo, Rizal Rizal, Betul Ayanoglu Ozturk, Carlos Sainz Zuñiga, Sara Bodbin, Maria Eleftheriou, Emma Collins, Sara Pijuan-Galito. Many happy memories were made, and I will always look back on them fondly. I want to thank my friends back home for keeping me going, in particular my friend Kurtis, who was there for me when I needed it most.

Finally, I would like to express my appreciation to my family who are the best support network anyone could wish for. To my brother Luke, and Maddy, for always supporting me and picking me up when I most needed it. To my aunty Helen, for all of her love and encouragement. To my cousins Jamie, Bryony and Daniel, and their partners, who are there for me no matter what, and Albie and Rosie for all the joy they bring me. I want to thank my grandparents Janet and Peter, and my nan Margaret for everything they have taught me. Lastly, to my parents, Stephen and Karla, for their unconditional love and support. I will be eternally grateful for everything they have done for me. Without them, this would not have been possible.

Declaration

I hereby declare that this thesis has been composed by myself and has not been submitted for any other degree previously. This project was supervised by Associate Professor Nicholas R.R. Hannan, Professor Cathy Merry and Dr Sara Cuevas Ocaña. Acknowledgements of specific procedures not performed by myself are stated; otherwise, the work described is my own.

Signature:



Liam Allen Reed

Date:

30th October 2023

Table of Contents

Acknowledgements i

Declaration ii

List of figures ix

List of tables xiv

Abstract 1

1. General Introduction 3

1.1. Introduction to the lungs 3

1.2. Embryonic lung development 3

1.2.1. Gastrulation and formation of definitive endoderm 3

1.2.2. Molecular aspects of definitive endoderm development 4

1.2.3. Definitive endoderm patterning 5

1.2.4. Specification of anterior foregut endoderm 5

1.2.5. Lung bud formation and branching morphogenesis 6

1.2.6. Development of the alveoli 8

1.2.7. Specification of alveolar epithelial cells 8

1.2.8. Human models of alveolar development 9

1.3. Lung structure and function 12

1.3.1. Structure of the lung 12

1.3.2. Proximal airway cell types and function 12

1.3.3. Alveolar cell types and function 13

1.3.4. Pulmonary surfactant 14

1.3.5. Signalling in the adult alveolar niche 15

1.3.6. Human primary alveolar cells and organoids 16

1.4. Major diseases of the lung 17

1.4.1. Idiopathic pulmonary fibrosis 17

1.4.2. Models of lung fibrosis 18
1.4.3. Cystic fibrosis 19
1.4.4. Chronic obstructive pulmonary dysfunction 19
1.5. Human pluripotent stem cells and *in vitro* differentiation to type 2 AECs of the lung 20
1.5.1. Introduction to human pluripotent stem cells 20
1.5.2. hPSC differentiation to DE and patterning of the gut tube *in vitro* 22
1.5.3. Specification of NKX2.1 lung epithelial progenitor cells 24
1.5.4. Induction of type 2 AECs from hPSC-derived lung progenitor cells 27
1.5.5. Human lung organoids and modelling branching morphogenesis 28
1.6. hiPSC-derived type 2 AECs for disease modelling 29
1.6.1. Modelling Hermansky-Pudlak syndrome using hiPSC-type 2 AECs 29
1.6.2. Modelling SFTPB deficiency with hiPSC-type 2 AECs 30
1.7. Aims and objectives of the project 31
2. Methods and Materials 32
2.1. Cell culture 32
2.1.1. Coating cell culture plasticware with Matrigel™ 32
2.1.2. Coating cell culture plasticware with Biolamina-521 32
2.1.3. Basal medium composition 34
2.1.4. Derivation of hiPSC lines 35
2.1.5. hiPSC cell lines 36
2.1.6. Genetic modification of hiPSCs using base editing 37
2.1.7. Karyotype analysis of hiPSCs 38
2.1.8. Maintenance of human induced pluripotent stem cells 39
2.1.9. Cryopreservation of hiPSCs 39
2.1.10. Pre-optimised differentiation of hiPSCs to lung organoids 39
2.1.11. Directed differentiation of hiPSCs to DE 40

2.1.12. Patterning of DE to anterior foregut endoderm 41
2.1.13. Specification of lung epithelial progenitor cells 41
2.1.14. Differentiation to type 2 AEC organoids 41
2.1.15. Passaging type 2 AEC organoids 42
2.2. Fluorescence activated cell sorting (FACS) and flow cytometry 43
2.2.1. Flow cytometry reagents 43
2.2.2. Flow cytometry analysis of DE cells 44
2.2.3. FACS of 2D lung progenitor cells 44
2.2.4. FACS of 3D type 2 alveolar organoids 45
2.2.5. Intracellular flow cytometry 46
2.3. Molecular biology 46
2.3.1. RNA extraction and cDNA synthesis 46
2.3.2. Quantitative real-time polymerase chain reaction (qRT-PCR) 47
2.3.3. Immunocytochemistry in 2D 49
2.3.4. Immunocytochemistry in 3D 49
2.3.5. RNA sequencing 50
2.3.6. Statistical analysis 51
3. Development of a protocol for differentiation of patient-derived hiPSCs to type 2 AECs of the lung 52
3.1. Introduction 52
3.2. Results 56
3.2.1. Overview of the strategy to generate type 2 AECs 56
3.2.2. Gene-editing patient-derived hiPSCs 58
62
3.2.3. Maintenance and characterisation of hiPSC lines 63
3.2.4. Differentiation of REBL-PAT control hiPSCs to lung progenitor cells using a previously established protocol 66

3.2.5. Differentiation of IPF1.12 hiPSCs to lung organoids using a previously established protocol 69

3.2.6. Generation of three-dimensional organoids 72

3.2.7. Single cell passaging of organoids and enrichment via FACS for CPM and NaPi2B..... 75

3.2.8. Lung progenitor cells in monolayer culture lack CPM expression 75

3.2.9. Growth factor optimisation of definitive endoderm specification 79

3.2.10. Addition of LY294002 significantly improves differentiation efficiency to definitive endoderm 85

3.2.11. Transient Wnt stimulation via CHIR99021 improves endoderm specification..... 88

3.2.12. Characterisation of definitive endoderm cells..... 95

3.2.13. Optimisation of the differentiation of endoderm to lung epithelial progenitor cells using carboxypeptidase M as a surrogate marker for NKX2.1-positive cells..... 97

3.2.14. CPM-positive cells express markers of non-lung endoderm 102

104

3.2.15. Patterning of anterior foregut endoderm using dual-SMAD inhibition increases expression of lung progenitor markers 105

3.2.16. Splitting definitive endoderm cells leads to emergence of a high CPM-expressing cell population 107

3.2.17. CPM 'high' cells are enriched for markers of lung epithelial progenitors 110

3.2.18. Robust induction of CPM+ cells after splitting following dual-SMAD inhibition..... 113

3.2.19. Development of a strategy to generate and isolate hiPSC-type 2 AECs 115

3.2.20. NaPi2B+ cells and organoids express markers of type 2 AECs..... 116

3.3. Discussion 121

4. Optimisation of a protocol to derive hiPSC-type 2 AECs in chemically defined medium 129

4.1. Introduction 129

4.2. Results 132

4.2.1. hiPSC differentiation to DE is robust and reproducible in XFL medium .. 132

4.2.2. Transcriptomic profiling of hiPSC-derived DE cells 138

4.2.3. DE cells can be differentiated to CPM-expressing lung progenitor cells in XFL medium 141

4.2.4. Specification of lung progenitor cells is modulated by changes in activation of signalling pathways 145

4.2.5. Changes in small molecule inhibition during foregut endoderm patterning impact differentiation to lung epithelial progenitor cells 147

4.2.6. A dose-dependent increase in recombinant WNT3a and R-spondin concentration improves specification of CPM(hi) cells 151

4.2.7. Removal of retinoic acid and activation of FGF10 signalling impairs lung epithelial progenitor cell differentiation 153

4.2.8. CPM(hi) cells generated in defined medium with an optimised differentiation protocol express a lung progenitor gene signature 155

4.2.9. Optimised differentiation with recombinant WNT3a and R-spondin in XFL medium is reproducible 157

4.2.10. The percentage of CPM(hi) cells peaks early during lung differentiation 159

4.2.11. Defining the transcriptomic profile of hPSC-derived lung epithelial progenitor cells 161

4.2.12. CPM(hi) cells give rise to NAPI2B+/EPCAM+ type 2 AEC-like cells in XFL medium 165

4.2.13. Characterising the transcriptome of NaPi2B+/EPCAM+ cells 167

4.2.14. Comparison of the transcriptomic profiles of NaPi2B populations to SFTPC-tdTomato enriched cells 173

4.2.15. Manipulation of TGF-β and BMP signalling impacts the specification of NAPI2B+ cells in organoid culture 177

4.2.16. NAPI2B+ epithelial cells grown with WNT and FGF signalling and SMAD inhibition express multiple lung gene markers 182

4.2.17. Benching the transcriptome of NaPi2B+/EPCAM+ cells to human primary fetal and adult lung 189

4.2.18. hiPSC-derived type 2 AEC organoids display mature functional characteristics and can recapitulate aspects of lung disease 193

4.3. Discussion 195

5. Differentiation of hiPSCs to type 2 AECs in fully defined, xeno-free condition 203

5.1. Introduction 203

5.2. Results 206

5.2.1. hiPSCs can be propagated on recombinant laminin 206

5.2.2. Highly reproducible differentiation of hiPSCs to DE in xeno-fee conditions 208

5.2.3. Transcriptomic profiling of hiPSC-derived DE in xeno-free conditions 212

5.2.4. Laminin-521 is a suitable substrate for induction of CPM(hi) cells from hPSC-derived DE 215

5.2.5. Defining the transcriptome of CPM(hi) lung progenitor cells derived in xeno-free conditions 217

5.2.6. Comparable transcriptomic profiles of CPM(hi) cells differentiated on laminin-521 or Matrigel™ 220

5.2.7. Generation of NaPi2B+/EPCAM+ cells in xeno-free conditions using recombinant 3D biosilk fibres 223

5.3. Discussion 225

6. General Discussion 230

References 238

List of figures

Figure 1-1. Major stages of lung development. 7
Figure 1-2. Schematic of current hypothesis of alveolar cell differentiation during human embryonic development. 11
Figure 1-3. Epithelial cell types of the human lung and their respective marker genes. 14
Figure 1-4. Schematic of TGFβ and BMP signalling pathways. 24
Figure 1-5. Wnt/β-catenin signalling pathway. 26
Figure 3-1. Overview of the project strategy to generate type 2 AECs for disease modelling and therapeutic purposes. 57
Figure 3-2. Optimisation of nucleofection efficiency using an eGFP plasmid vector. . 59
Figure 3-3. Plasmid maps for base editing. 60
Figure 3-4. Base edits can be detected at the population level using the restriction digestion HpyCH4V assay. 61
Figure 3-5. Karyotype analysis and validation of base-editing of IPF1.12 hiPSCs. 62
Figure 3-6. Morphology of human induced pluripotent stem cells. 64
Figure 3-7. Expression of pluripotency markers in hiPSCs. 65
Figure 3-8. Differentiation of REBL-PAT hiPSCs to alveolar epithelial type 2 cells. 67
Figure 3-9. Expression of stage-specific genes during lung differentiation of REBL-PAT hiPSCs 68
Figure 3-10. Gene expression during lung differentiation of IPF1.12 ($SFTPC^{WT/MUT}$) hiPSCs 70
Figure 3-11. Expression of SOX17 varies across hiPSC lines. 71
Figure 3-12. Generation of three-dimensional lung organoids. 73
Figure 3-13. Organoids can be recovered after cryopreservation. 74
Figure 3-14. Organoids can be dissociated into single cells for FACS. 77
Figure 3-15. CPM is expressed at very low levels after 8 days of lung differentiation. 78
Figure 3-16. Gating strategy for flow cytometry analysis of CXCR4 and CKIT. 80
Figure 3-17. Optimisation of definitive endoderm specification using recombinant proteins. 81
Figure 3-18. Induction of BE.32 ($SFTPC^{WT/WT}$) CXCR4-positive endoderm cells is sub-optimal using recombinant proteins. 82
Figure 3-19. Induction of BE.32 ($SFTPC^{WT/WT}$) CKIT-positive endoderm cells is sub-optimal using recombinant proteins. 83
Figure 3-20. Definitive endoderm cells specified with recombinant proteins express endoderm gene markers. 84
Figure 3-21. Optimisation of definitive endoderm specification with the addition of LY294002 in BE.32 ($SFTPC^{WT/WT}$) hiPSCs. 86
Figure 3-22. Addition of LY294002 improves induction of BE.32 ($SFTPC^{WT/WT}$) CXCR4-expressing endoderm cells. 87
Figure 3-23. Addition of LY294002 improves induction of BE.32 ($SFTPC^{WT/WT}$) CKIT-expressing endoderm cells. 88

Figure 3-24. Optimisation of definitive endoderm differentiation using the small molecule inhibitor CHIR99021 in BE.32 (SFTPC$^{WT/WT}$) hiPSCs. 90
Figure 3-25. Induction of BE.32 (SFTPC$^{WT/WT}$) CXCR4-positive cells are promoted by CHIR99021. 91
Figure 3-26. Induction of BE.32 (SFTPC$^{WT/WT}$) CKIT-positive cells are promoted by CHIR99021. 92
Figure 3-27. Flow cytometry gating strategy for co-staining of CXCR4-PE and CKIT-APC. 93
Figure 3-28. Quantification of CXCR4 and CKIT expression after 4 days of optimised endoderm differentiation. 94
Figure 3-29. Definitive endoderm genes are expressed with an optimised endoderm differentiation protocol. 96
Figure 3-30. Protein expression of definitive endoderm markers. 97
Figure 3-31. Optimisation of the differentiation of definitive endoderm to lung epithelial progenitor cells. 99
Figure 3-32. Each lung progenitor cell differentiation protocol induces CPM-positive cells. 101
Figure 3-33. BE.32 (SFTPC$^{WT/WT}$) CPM-positive cells express markers of non-lung endoderm. 103
Figure 3-34. Sorted CPM cells express NKX2.1. 104
Figure 3-35. Anteriorisation of endoderm via dual-SMAD inhibition preceding lung progenitor differentiation promotes induction of lung markers. 106
Figure 3-36. Splitting endoderm cells prior to lung differentiation induces three CPM populations. 109
Figure 3-37. Splitting endoderm cells before lung differentiation upregulates lung progenitor specific markers. 110
Figure 3-38. BE.31 (SFTPC$^{WT/WT}$) CPM-high expressing cells are enriched for lung epithelial progenitor cell markers. 111
Figure 3-39. Cells on day 8 of lung differentiation express lung progenitor cell markers. 112
Figure 3-40. Splitting cells after anteriorisation of foregut induces a high percentage of CPM(hi) cells in BE.31 (SFTPC$^{WT/WT}$) hiPSCs. 114
Figure 3-41. Strategy for enriching NaPi2B/EPCAM double-positive cells. 116
Figure 3-42. BE.31 (SFTPC$^{WT/WT}$) NAPI2B+/EPCAM+ sorted cells are enriched for type 2 AEC markers. 118
Figure 3-43. NaPi2B-enriched cells and organoids express markers of type 2 AECs. 119
Figure 3-44. Cells enriched for NaPi2B express markers of type 2 AECs. 120

Figure 4-1. Schematic of the optimised directed differentiation of hiPSCs to type 2 AECs. 134
Figure 4-2. Optimised endoderm differentiation generates the highest number of CXCR4+/CKIT+ cells in XFL medium. 135

Figure 4-3. Optimisation of cell seeding density for differentiation of hiPSCs to endoderm in XFL medium on Matrigel™. 136
Figure 4-4. hiPSCs can be robustly differentiated to definitive endoderm in chemically defined medium. 137
Figure 4-5. Differentiation of hiPSCs to DE elicits transcriptomic changes. 139
Figure 4-6. Transcriptomic profiling of hiPSC-derived DE cells. 140
Figure 4-7. Confluency of cells significantly impacts specification of CPM(hi) cells. . 143
Figure 4-8. CPM(hi) cells differentiated in XFL medium express lung epithelial progenitor genes. 144
Figure 4-9. Modulation of growth factors during lung progenitor cell differentiation impacts specification of CPM(hi) cells. 146
Figure 4-10. CHIR99021 induces over-proliferation during lung progenitor specification. 147
Figure 4-11. Effect of dual-SMAD and Wnt signalling inhibition during foregut patterning on induction of CPM(hi)-expressing cells. 149
Figure 4-12. BMP signalling inhibition via LDN193189 during foregut patterning reduces specification of CPM(hi) cells. 150
Figure 4-13. Increasing concentrations of recombinant Wnt3a and R-spondin increase induction of CPM(hi) cell numbers. 152
Figure 4-14. Efficiency of differentiation to CPM(high) lung progenitor cells is reduced with removal of retinoic acid and addition of FGF10. 154
Figure 4-15. CPM(high) cells specified by dual-SMAD inhibition alone or with Wnt inhibition before lung differentiation express markers of lung progenitor cells. 156
Figure 4-16. CPM(hi) cells differentiated in XFL medium express NKX2.1. 157
Figure 4-17. Specification of CPM cells on Matrigel and laminin in animal free medium. 158
Figure 4-18. Induction of CPM(Hi) cells during early differentiation. 160
Figure 4-19. Transcriptomic profiling of hiPSC-derived cell types in at developmental major stages of differentiation to type 2 AECs. 162
Figure 4-20. Gene expression profiling of cells during lung differentiation and CPM populations. 163
Figure 4-21. Transcriptomic analysis reveals enrichment of lung progenitor cell transcripts in CPM(hi) cells. 164
Figure 4-22. CPM(hi) cells can give rise to NaPi2B+/EPCAM+ cells in organoids maintained in XFL medium. 166
Figure 4-23. Transcriptomic profiling of isolated NaPi2B+/EPCAM+ cells. 170
Figure 4-24. Transcriptomic analysis reveals enrichment of type 2 AEC genes in NaPi2B+/EPCAM+ cells. 171
Figure 4-25. NaPi2B+/EPCAM+ cells do not strongly express transcripts related to proximal airway cells. 172
Figure 4-26. NaPi2B+/EPCAM+ cells have a similar transcriptomic profile to hPSC-SFTPC+ cells. 174

Figure 4-27. Gene expression analysis of type 2 AEC cell states and type 1 AEC genes in NaPi2B- and SFTPCtdTomato-isolated populations. 176
Figure 4-28. Manipulation of TGF-β and BMP signalling leads to variance in differentiation to NAPI2B+/EPCAM+ cells. 179
Figure 4-29. Organoids differentiated in the presence of CHIR99021, FGF7, EGF and dual-SMAD inhibition express non-lung endoderm markers without FACS enrichment. 180
Figure 4-30. Dual-SMAD inhibition with Wnt and FGF signalling induces NAPI2B+ epithelial cells in organoid culture. 181
Figure 4-31. Airway, alveolar and non-lung endoderm genes and markers..... 184
Figure 4-32. Dual-SMAD inhibition specifies organoids with NAPI2B+ cells that express a mix of proximal and distal airway markers. 186
Figure 4-33. NAPI2B+/EPCAM+ cells from organoids maintained with dual-SMAD inhibition express markers of alveolar and airway cell identity. 188
Figure 4-34. Comparison of transcriptomic profiles of NaPi2B cells and primary lung samples. 190
Figure 4-35. Benchmarking the transcriptomic profile of hiPSC-type 2 AECs to human fetal and adult lung samples..... 191
Figure 4-36. Gene expression analysis of non-type 2 AEC lung markers in NaPi2B cells and primary lung samples. 192
Figure 4-37. hiPSC-derived type 2 AECs are functionally mature and can recapitulate aspects of lung disease. 194

Figure 5-1. hiPSCs maintained on recombinant laminin express pluripotency markers. 207
Figure 5-2. Schematic of the fully optimised differentiation of hiPSCs to type 2 AECs in defined medium. 209
Figure 5-3. Different hiPSC seeding densities lead to robust induction of CXCR4+/CKIT+ cells in XFL medium on recombinant laminin. 210
Figure 5-4. Differentiation of hiPSCs to DE is reproducible in XFL medium on recombinant laminin. 211
Figure 5-5. Transcriptomic profiling of hPSC-derived DE cells in xeno-free conditions. 213
Figure 5-6. Transcriptome profiles of DE cells derived from hiPSCs in XFL medium on Matrigel and laminin-521 are comparable. 214
Figure 5-7. CPM(hi) cells can be specified from hiPSCs in fully defined, xeno-free conditions. 216
Figure 5-8. The transcriptome of lung cells and CPM populations differentiated in xeno-free conditions cluster together. 218
Figure 5-9. Transcriptional profiling of differentiated lung cells and CPM isolated populations. 219
Figure 5-10. CPM(hi) cells derived from hiPSCs on laminin-521 and Matrigel™ are not associated with significantly different transcriptomes. 221

Figure 5-11. The transcriptome of CPM(hi) cells differentiated on laminin-521 is enriched for lung epithelial progenitor transcripts. 222
Figure 5-12. NaPi2B+/EPCAM+ cells can be generated in xeno-free conditions using recombinant biosilk. 224

List of tables

Table 2. 1. Composition of cell culture basal media 32
Table 2. 2. List of growth factors and small molecules used for hiPSC maintenance and differentiation..... 35
Table 2. 3. Antibodies for flow cytometry and FACS 43
Table 2. 4. Primer sequences for qRT-PCR..... 47
Table 2. 5. Antibodies for immunocytochemistry..... 50

Abstract

Type 2 alveolar epithelial cells (AECs) are the facultative progenitors of the lung alveolar epithelium, as they self-renew and dedifferentiate into type 1 AECs under homeostatic conditions and in response to chronic injury. They are responsible for the synthesis and secretion of a lipoprotein rich surfactant to aid in the mechanics of breathing, preventing alveolar collapse at periods of low lung volume during respiration. Surfactant protein C (SFTPC) is synthesised exclusively in type 2 AECs and enhances phospholipid adsorption and spreading. Damage to type 2 AECs is implicated in diseases such as idiopathic pulmonary fibrosis (IPF), and several genetic and environmental risk factors have been identified. This includes mutations in SFTPC identified in familial cases of pulmonary fibrosis. Access to type 2 AECs for research is limited, and so animal models have primarily been used. However, animal models lack many features of human development and disease and as a result newly developed therapeutics for lung disease often fail in clinical trials.

The advent of human induced pluripotent stem cell technology and directed differentiation has circumvented some of these problems, providing an unlimited source of adult cells to model human development and disease, and for clinical translation. However, one major limitation of current protocols to differentiate human pluripotent stem cells (hPSCs) to type 2 AECs is the presence of animal-derived components in medium and the widespread use of basement membrane matrices like Matrigel™. This can cause significant biological variability and limits the therapeutic application of derived cells.

To address these issues, we reprogramme fibroblasts from a patient diagnosed with familial idiopathic pulmonary fibrosis (fIPF) harbouring a heterozygous mutation in the BRICHOS domain of SFTPC to human induced pluripotent stem cells (hiPSCs). Using base-editing technology, we generate homozygous-corrected and homozygous-mutant isogenic control hiPSC lines. We describe the development a novel, xeno-free protocol for the directed differentiation of hiPSCs to type 2 AEC organoids using chemically defined medium and human recombinant laminin. To overcome the need for fluorescence reporter cell lines, we develop a strategy to enrich for lung epithelial progenitor cells and type 2 AECs via cell surface markers by fluorescence activated cell

sorting (FACS). Importantly, type 2 AEC organoids derived from healthy and mutant hiPSCs are able to recapitulate features of human lung disease. We perform detailed transcriptomic profiling across the differentiation and confirm the enrichment of genes related to major stages of embryonic lung development. Finally, we benchmark our type 2 AECs to primary lung samples and confirm similar gene expression profiles to late human fetal lung. These findings provide a platform to investigate lung disease and generate type 2 AECs with clinical applicability.

1. General Introduction

1.1. Introduction to the lungs

The lungs are the major organ of the respiratory system and form an arborized network of conducting and respiratory airways which terminate in the alveoli, the site of gas-exchange at the alveolar-capillary barrier. Specialised epithelial cells lining the lung derive from a population of embryonic endoderm cells at gastrulation and facilitate several critical functions in coordination with a diverse range of endothelial, smooth muscle and stromal cells, including gas-exchange, physical protection from inhaled noxious stimuli, removal of pathogens and roles in regulating innate immunity to maintain lung homeostasis. Development of the mammalian respiratory system occurs across 5 major stages: the embryonic, pseudoglandular, canalicular, saccular and alveolar stage (Parker & Prince, 2011; Schittny, 2017). Much of our current understanding of lung development derives from animal models. However, despite their value, they fail to capture the fundamental differences between animal and human lung development, function, and disease. The prognosis for major respiratory diseases such as chronic obstructive pulmonary disease (COPD) and idiopathic pulmonary fibrosis (IPF) remains considerably poor as newly developed therapeutics often fail to make it through clinical trials, largely owing to poor predictive models (Barnes et al., 2015). This highlights an unmet need for human-relevant platforms of disease for drug screening purposes. Below, I review our current knowledge of embryonic development of the lung, and I also discuss the advancements in human cell models to further our understanding of human respiratory development and disease.

1.2. Embryonic lung development

1.2.1. Gastrulation and formation of definitive endoderm

Model organisms such as drosophila, xenopus, chick and mouse have provided the foundation for our current understanding of the processes regulating embryonic development of the lung (Nowotschin et al., 2019). Therefore, the majority of the findings discussed in this section refer to development of the mouse, but differences to human development are highlighted.

Gastrulation constitutes a major event during embryogenesis where the three primary germ layers; the endoderm, mesoderm and ectoderm, are specified from the pluripotent epiblast, establishing the foundation for development of all future tissues and organs (Tam et al., 1993). The onset of gastrulation is marked by the formation of an elongated groove in the posterior epiblast known as the primitive streak (PS) (Tam & Behringer, 1997). Cells of the epiblast fated to become embryonic endoderm, also known as definitive endoderm (DE), ingress through the anterior region of the PS, emerge, and intercalate with an epithelial layer of visceral endoderm (VE) overlying the epiblast (Burtscher & Lickert, 2009; Kwon et al., 2008; Lawson et al., 1991; Viotti et al., 2014). Findings in *Drosophila* and zebrafish model systems have suggested prospective DE cells undergo epithelial-mesenchymal transition (EMT) and move through the PS followed by re-acquisition of an epithelial identity (Burtscher & Lickert, 2009; Tam & Behringer, 1997; Viotti et al., 2014). However, the extent of EMT in mammals remains uncertain, as the DE layer can form independent of a complete switch to a mesenchymal state by suppression of EMT factors. As gastrulation concludes, the DE has formed a continuous epithelium at the innermost layer of the embryo (Scheibner et al., 2021; Tam & Behringer, 1997).

1.2.2. Molecular aspects of definitive endoderm development

During embryogenesis, communication between developing cells and tissues through a limited number of highly conserved signalling pathways, such as Wnt/β-catenin, bone morphogenic protein (BMP)/transforming growth factor-beta (TGFβ), Notch, Retinoic acid (RA), fibroblast growth factor (FGF) and Hedgehog pathways, drive diverse cell fate decisions. Integration of these pathways establishes dynamic transcriptional networks that govern cell specification (Perrimon et al., 2012). In the posterior epiblast region, activation of Wnt/β-catenin and high levels of Nodal, a secreted ligand of the TGF-β family of signalling proteins, is critical for PS formation, early embryonic patterning and specification of DE cells (Arnold & Robertson, 2009; Engert et al., 2013; Vincent et al., 2003). Nodal and Wnt/β-catenin signalling input induces the expression of several transcription factors to establish the genetic program of DE. This includes *Foxa2* which is expressed in prospective DE cells in the epiblast and is critical for development of descendants of the DE layer (Ang & Rossant, 1994; Burtscher & Lickert, 2009; Scheibner et al., 2021), *Sox17* which is indispensable

for DE formation as it enables DE cells to intercalate in the VE layer (Kanai-Azuma et al., 2002; Scheibner et al., 2021; Viotti et al., 2014), Gata zinc finger transcription factor family members including *Gata4* and *Gata6* (Molkentin et al., 1997; Morrisey et al., 1998), and *Eomes*, as knockout mouse embryos fail to develop embryonic endoderm (Arnold et al., 2008).

1.2.3. Definitive endoderm patterning

As development proceeds, the embryo undergoes major morphogenetic movements, and the single-layered DE epithelium folds into a gut tube surrounded by embryonic mesoderm and is patterned along its anterior-posterior axis (Kraus & Grapin-Botton, 2012). FGF, BMP and Wnt/β-catenin signalling ligands derived from adjacent mesenchyme posteriorize endoderm, whereas inhibition of TGF-β/BMP and Wnt signalling induce anterior endoderm identity (Zorn & Wells, 2009). RA also plays a key role in axis specification of endoderm in a concentration-dependent manner (Kelly & Drysdale, 2015). This segregates the gut tube broadly into foregut, midgut and hindgut domains, marked by transcription factor expression of *Sox2*, *Pdx1* and *Cdx2*, respectively (Zorn & Wells, 2009). Precursors for organs of the respiratory and gastrointestinal tract emerge at discrete regions along the gut tube as patterning continues and distinct transcriptional signatures are established. The foregut region gives rise to the trachea, oesophagus, lung, stomach, liver and pancreas, and the small and large intestine derive from the midgut and hindgut, respectively (Kraus & Grapin-Botton, 2012; Zorn & Wells, 2009).

1.2.4. Specification of anterior foregut endoderm

The onset of respiratory system specification is signified by localised expression of NKX2 Homeobox 1 (*Nkx2.1*) (also known as thyroid-specific transcription factor 1(*Ttf1)*), a master regulator of normal lung development, in a population of cells in the ventral anterior foregut endoderm (Lazzaro et al., 1991; Minoo et al., 1999). This is driven by inductive signals from the surrounding mesenchyme that were previously refractory to anterior endoderm fate. Temporal control of RA signalling primes ventral foregut endoderm cells to commit to the respiratory lineage in response to Wnt and Bmp ligands expressed later by splanchnic mesenchyme (Rankin et al., 2016; Rankin et al., 2018). RA induces sonic hedgehog (Shh) signalling in the ventral foregut and a

feedback mechanism is established where Shh signals to the mesenchyme to potentiate Wnt and Bmp signalling to promote induction of respiratory-related genes in the foregut (Rankin et al., 2016). Wnt signalling proteins induce *Nkx2.1* expression and are necessary to specify prospective respiratory cells (Goss et al., 2009). In parallel, signalling through Bmp receptors 1A and 1B establishes the dorsal-ventral axis of the anterior foregut and represses dorsally expressed *Sox2* to maintain ventral expression of *Nkx2.1* (Domyan et al., 2011; Rankin et al., 2018). Fgf signalling from neighbouring mesenchyme also plays a role in specifying the respiratory domain (Serls et al., 2005).

1.2.5. Lung bud formation and branching morphogenesis

From the Nkx2.1+ ventral foregut region emerge two lung buds shortly followed by septation of the trachea and oesophagus (Schittny, 2017). Initial outgrowth is dependent on Fgf10 ligands from distal mesenchyme signalling through Fgf receptor 2b (Fgfr2b) in adjacent ventral foregut endoderm to elicit the onset of cell migration (Cardoso & Lü, 2006). As in earlier development, RA is critical in facilitating this process. In the mesenchyme, RA inhibits Tgf-β-mediated repression of *Fgf10*, and inhibits Dickkopf 1 (*Dkk1*), an inhibitor of Wnt signalling. This enables Wnt proteins to activate *Fgf10*, and also signal to the ventral foregut endoderm independently to promote the onset of the budding response. This signifies the conclusion of the embryonic period of lung development (F. Chen et al., 2010; Chen et al., 2007; Desai et al., 2004).

During pseudoglandular development, the network of conducting airways (bronchi and bronchioles) is established during a highly stereotyped process called branching morphogenesis (Swarr & Morrisey, 2015). Compared to mice (13-17 branch generations), humans undergo extended branching (17-21 branch generations) (Irvin & Bates, 2003). The interplay of Wnt, Fgf and Bmp pathways, and negative feedback mechanisms controlled by Shh, Tgf-β, and the Sprouty signalling proteins, coordinate Fgf10-Fgfr2b signalling between the mesenchyme and foregut endoderm to regulate repeated bud elongation and bifurcation into the surrounding mesenchyme (reviewed in Cardoso and Lü (2006), Morrisey and Hogan (2010), Swarr and Morrisey (2015)). When human fetal lung explants are cultured in the presence of FGF10 they fail to undergo branching, suggesting possible contrasting roles for FGF10 between species (Danopoulos et al., 2019). Branching of the airways is driven by a multipotent

progenitor population in the distal tip epithelium of each developing bud that can give rise to all cell types of the lung. In mice, these cells are marked bySox9 and inhibitor of DNA binding 2 (Id2), and human tips also express SOX2 (Nikolić et al., 2017; Rawlins et al., 2009). In mouse, Sox9+/Id2+ cells leave the proliferative tip region, downregulate *Sox9* and upregulate *Sox2* to become airway progenitor cells of the stalk. Towards the end of the pseudoglandular stage of development, specialised airway cells can be identified (**Fig 1-1**) (Swarr & Morrisey, 2015).

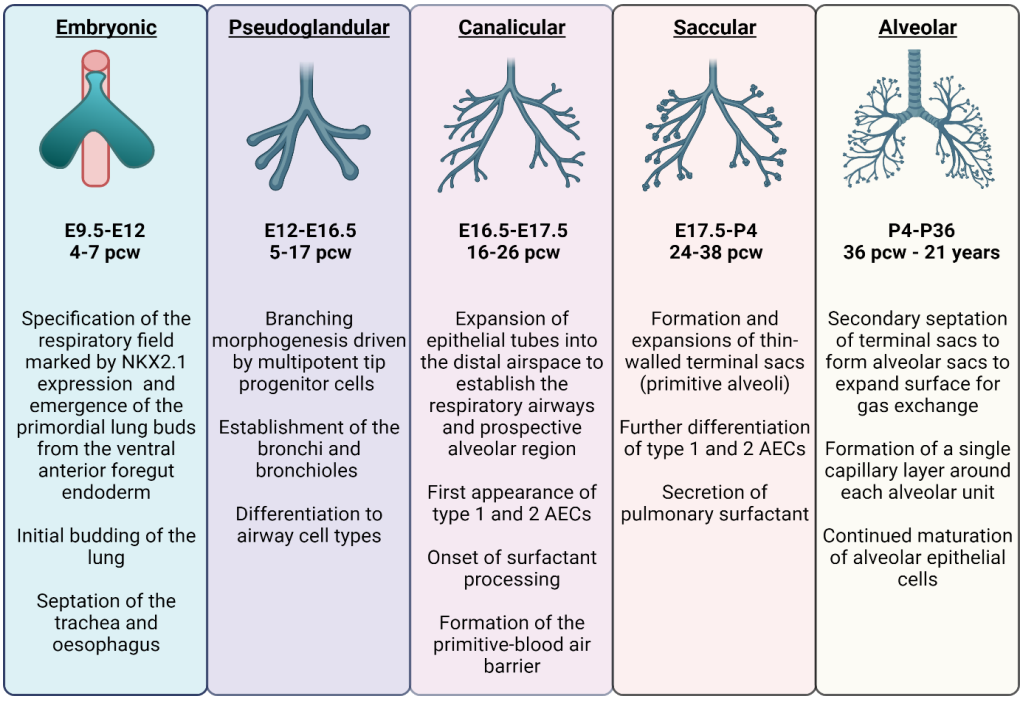


Figure 1-1. Major stages of lung development.

The 5 major stages of lung development. Development of the lung begins with emergence of the first two lung buds and separation of the developing trachea and oesophagus around embryonic day 9.5 in mouse and post-conception week (pcw) 4 in humans. During pseudoglandular development stereotypical branching is driven by reciprocal signalling between the epithelium and surrounding mesenchyme to form a highly branched airway structure which further expands into distal airspaces during the canalicular phase of development. The formation of narrow lung sacs preceding formation of the alveoli during sacculation occurs postnatally in the mouse whereas the majority of lung morphogenesis is completed prior to birth in humans. Maturation of the alveolar compartment concludes around post-natal day 30 in mice but continues into early adulthood in humans. Figure generated using Biorender.

1.2.6. Development of the alveoli

Branching morphogenesis continues during the canalicular stage of development and distal tubules expand into the airspace to establish the respiratory airways and prospective alveolar region. A capillary network develops around the airspaces of the distal epithelium and forms a primitive blood-air barrier. At this stage, bud tip progenitor cells begin to differentiate into the first type 1 and 2 alveolar epithelial cells (AECs). The specialised organelles of type 2 AECs responsible for pulmonary surfactant storage and secretion, called lamellar bodies (LBs), can be detected at this stage (Mammoto & Mammoto, 2019; Rawlins et al., 2009; Schittny, 2017).

The saccular phase of development marks the end of branching morphogenesis as clusters of thin-walled terminal sacs (primitive alveoli) develop and further expand into the distal airspaces. Terminal sacs are divided by primary septa and surrounded by a double capillary layer as the space between the blood-air barrier is reduced. Type 1 and 2 AECs undergo further differentiation and the machinery for surfactant processing matures resulting in secretion of pulmonary surfactant into the luminal space (Mammoto & Mammoto, 2019; Nikolić et al., 2018). Development of the lung concludes with alveolarization. Secondary septation further divides the terminal sacs into alveoli to increase the surface area for gas exchange. Type 1 AECs flatten and cover 95% of the alveolar epithelial surface, secondary septa thin and a single-layered capillary network is formed around the alveoli to enable efficient gas exchange at birth (Mammoto & Mammoto, 2019; Nikolić et al., 2018; Schittny, 2017). In mice, alveolarization starts after birth and concludes around post-natal day 36, but in humans it begins before birth and continues into early adulthood (Narayanan et al., 2012).

1.2.7. Specification of alveolar epithelial cells

In the mouse, two models have been proposed to explain the specification of alveolar cells. Some findings suggest there exists a bipotent alveolar progenitor that can give rise to type 1 and 2 AECs; however, it is likely this reflects a rare cell type in the developing lung that does not contribute extensively to alveolar lineages. Instead, distinct progenitor populations for both type 1 and 2 AECs appear to be specified from bud tip progenitor cells during early stages of branching morphogenesis in a spatial

manner, with type 2 AECs emerging first from the Sox9+ distal tip epithelium region followed by type 1 AECs in the stalk of the branching lung (Desai et al., 2014; Frank et al., 2019).

Important roles for several signalling pathways governing alveolar development have been identified. Glucocorticoid steroids are commonly used to treat preterm infants to promote maturation of the surfactant processing machinery in type 2 AECs (Bolt et al., 2001). Although studies in mice have suggested glucocorticoid signalling is not essential for determining alveolar fate, it can promote alveolar differentiation, potentially by signalling to the mesenchyme which in turn signals to Sox9+ bud tip epithelial cells (Bridges et al., 2020; Laresgoiti et al., 2016). Similarly, enhanced cyclic adenosine monophosphate (cAMP) signalling can promote expression of surfactant protein-related genes, but its impact on alveolar development is unknown (Mendelson & Boggaram, 1989). Fgf7/10 activation of Fgfr2b has been shown to selectively induce specification of type 2 AECs in mice. Instead, lack of Fgfr2b signalling, and possibly activation of the Notch pathway, favours type 1 AEC fate (Brownfield et al., 2022). Active Wnt signalling is also important for proliferation and expansion of type 2 AECs during alveolar development in mice, and Wnt inhibition promotes type 1 AEC differentiation (Frank et al., 2016).

Nkx2.1 is a master regulator of alveolar development. Nkx2.1 can bind to the promoter region of type 1 and 2 AEC-specific genes and can promote differentiation to type 1 AECs with the transcriptional co-factor YAP/TAZ, or favour type 2 AEC fate via differential chromatin binding (Little et al., 2019; Little et al., 2021). These epigenetic features are acquired in Sox9+ progenitor cells and persist during alveolar cell specification, providing a mechanistic explanation for how Nkx2.1 regulates opposing cell fate choices (Little et al., 2021).

1.2.8. Human models of alveolar development

Limited access to human fetal tissue from later developmental timepoints has made it difficult to investigate human alveolar development beyond early stages. However, advances in techniques such as single cell RNA-sequencing (scRNA-seq) have permitted more extensive analysis of human fetal lung samples when available. Bud tip progenitor cells have been isolated from pseudoglandular stage human lungs and

were found to be analogous to mouse bud tips. They expressed SOX9, ID2 and NKX2.1, but in contrast to mice, they also co-expressed SOX2, and could be maintained *in vitro* long-term as self-renewing organoids in the presence of WNT, FGF and EGF signalling, and TGF-β and BMP inhibition. However, mouse tips could not be propagated under these conditions, highlighting functional differences between species. Human tips retained multipotency in culture and could give rise to airway and alveolar cells (Nikolić et al., 2017).

Recent findings using scRNAseq datasets of human fetal lung development for trajectory inference have shown canalicular stage bud tip progenitor cells downregulate tip markers and enter a stalk 'state' of the branch before acquiring type 1 and 2 AEC identity (**Fig. 1-2**) (He et al., 2022). This model is in contrast with current hypotheses for mouse alveolar development which suggest early fate restriction into type 1 and 2 AEC progenitor cells or bipotent alveolar progenitors capable of giving rise to both cell types (Frank et al., 2019; Treutlein et al., 2014). Isolation of bud tip cells from 17-21 post-conception week (pcw) human lungs can also generate self-renewing organoids *in vitro* but co-express progenitor and type 2 AEC markers. These cells retain the capacity to give rise to airway-fated cells, but readily differentiate to type 2 AECs. This model showed WNT signalling induces NKX2.1 activation which could bind to alveolar gene targets to drive differentiation to type 2 AECs whilst repressing airway lineages (Lim et al., 2023). Furthermore, this study indicated FGF signalling favoured airway over alveolar development (Lim et al., 2023). Despite this, FGF ligands are commonly used to promote maintenance of human type 2 AECs *in vitro*, and there is evidence to suggest FGF signalling is required to differentiate human bud tip progenitors to alveolar cells in conjunction with other signalling pathways (Frum et al., 2023; Jacob et al., 2017). Indeed, key roles for BMP signalling and TGF-β inhibition in promoting human bud tip differentiation to type 2 AECs have just recently been identified (Frum et al., 2023). These models are valuable tools for investigating early developmental control of alveolar specification. However, they fail to capture the

continued differentiation and maturation of alveolar cells during saccular and alveolar stages of development when fetal tissue cannot be obtained.

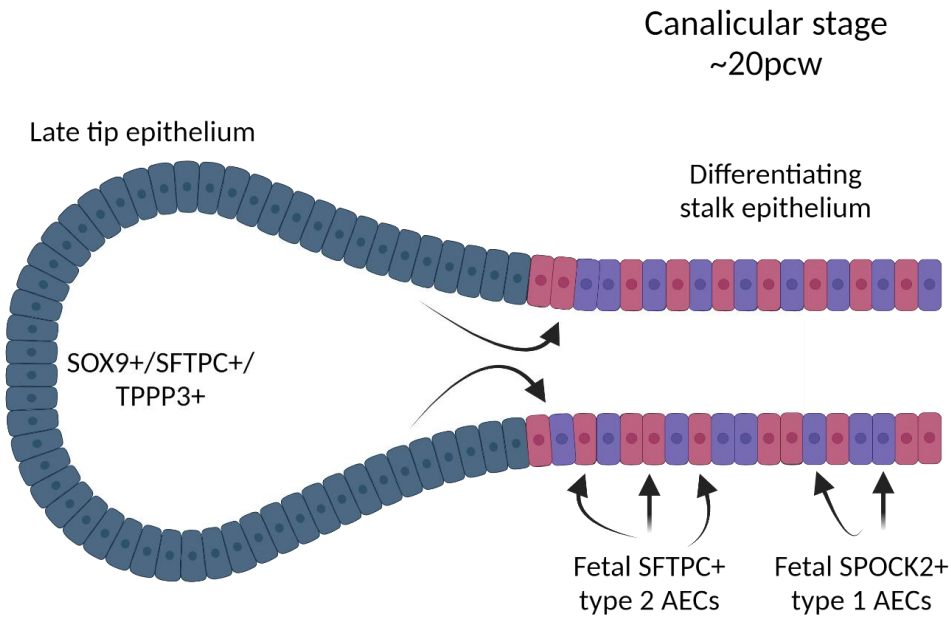


Figure 1-2. Schematic of current hypothesis of alveolar cell differentiation during human embryonic development.

During the late canalicular stage of development the proliferating late tip epithelium of the branching lung co-expresses tip epithelial markers such as SOX9 and TPPP3 and type 2 AEC markers like SFTPC. As the tip continues to proliferate cells leave the tip and enter what is referred to as a 'stalk' state prior to downregulating tip markers and upregulating type 2 AEC markers to acquire type 2 AEC identity. At a later timepoint in development following differentiation to type 2 AECs, other SFTPC-ve stalk cells differentiate to type 1 AECs expressing SPOCK2 through spatiotemporally controlled WNT and FGF signalling in part mediated by alveolar fibroblasts. Figure adapted from (He et al., 2022). Figure generated using Biorender.

1.3. Lung structure and function

1.3.1. Structure of the lung

The right and left lung in humans is separated into 3 and 2 lobes, respectively, and is covered by a membranous lining called the pleura (Tomic et al., 2013). The trachea, bronchi and bronchioles form the conducting airways of the respiratory system and are lined with columnar, pseudostratified epithelium, and conduct inhaled air to the respiratory airways. The terminal bronchioles narrow and give rise to the respiratory bronchioles which terminate into clusters of alveoli that are lined with squamous epithelium for gas-exchange (Karra et al., 2019; Rackley & Stripp, 2012). The trachea and bronchi are supported by cartilage and the bronchioles are surrounded with airway smooth muscle as the airways narrow. Unlike the human, the mouse lacks respiratory bronchioles. Instead, the terminal bronchioles open into the alveoli at a region called the bronchioalveolar duct junction (BADJ) where a specialised population of bronchioalveolar stem cells (BASCs) reside (Jones-Freeman & Starkey, 2020).

1.3.2. Proximal airway cell types and function

The pseudostratified epithelium of the proximal airways consists of several specialised cell types, including basal, secretory (club), goblet, ciliated, neuroendocrine and tuft cells, and pulmonary ionocytes (**Fig. 1-3**) (Meng et al., 2023). Airway basal cells are found primarily in the epithelium of the trachea in mouse, but in humans they extend into the terminal bronchioles (Wu et al., 2022). They are the resident stem cell population of the airway and facilitate epithelial regeneration by differentiating into all airway cell types upon injury (Davis & Wypych, 2021). Club cells are identified by expression of secretoglobin family 1A, member 1 (Scgb1a1) and can be found throughout the mouse lung but are found in the narrower, terminal bronchioles of human airways. Like basal cells, they also have regenerative capacity and can differentiate into ciliated and goblet cells, and in extreme cases, basal cells, in response to injury (Basil et al., 2020; Dean & Snelgrove, 2018). Goblet cells are the primary mucus-producing cells of the conducting airways marked by the expression of gel-forming glycoproteins MUC5AC and MUC5B. They work in conjunction with airway ciliated cells to mediate mucociliary clearance (Whitsett, 2018). Pulmonary neuroendocrine cells are a rare airway epithelial cell type but are critical for translating

the physiological state of the airways and regulating immune responses (Noguchi et al., 2020). Tuft cells have similar sensory properties and are also important for immunity of the lung (Li et al., 2022). Recently, a novel lung epithelial cell type rich in cystic fibrosis transmembrane conductance regulator (CFTR), termed the pulmonary ionocyte, was also discovered, and has been suggested to play a role in cystic fibrosis pathogenesis (Montoro et al., 2018; Plasschaert et al., 2018). Pulmonary ionocytes facilitate the flow of chloride (Cl^-) ions through CFTR and Cl^- channels and mediate the absorption of liquid from the airway epithelial surface (Lei et al., 2023) In mouse, BASCs at the BADJ express markers of club cells and type 2 AECs and can differentiate into both airway and alveolar cells upon injury, but a human counterpart is yet to be described (Liu et al., 2019). It is important to note that scRNA-seq technology is beginning to uncover the complexity of the airway epithelium, identifying multiple molecular cell subtypes (Sikkema et al., 2023; Travaglini et al., 2020). Further investigation is required to understand how such molecular heterogeneity contributes to normal lung epithelial homeostasis, response to injury and disease.

1.3.3. Alveolar cell types and function

In the alveoli reside two major epithelial cell populations, the type 1 and type 2 AECs. Type 1 AECs are squamous cells that cover 95% of the alveolar epithelial surface and form the blood-air barrier with the capillary endothelium, facilitating gas exchange (Basil et al., 2020). The primary role of type 2 AECs is to synthesise and secrete pulmonary surfactant, but they also function as the facultative stem cells of the adult alveoli (**Fig. 1-3**) (Olajuyin et al., 2019). Under homeostatic conditions and following lung injury in mice, type 2 AECs can self-renew and differentiate into type 1 AECs (Barkauskas et al., 2013). Following injury to the mouse lung, it has been shown that a sub-fraction of type 2 AECs become 'primed' in response to macrophage-derived inflammatory signals and give rise to cells termed damage-associated transient progenitors, before committing to type 1 AEC fate (Choi et al., 2020). A subpopulation of *Axin2*+ type 2 AECs highly enriched for targets of Wnt signalling have also recently been identified. This population display normal type 2 AEC markers such as surfactant protein C (*Sftpc*) and maintain a stem-like state in response to Wnt signals in the niche. Following alveolar injury, these cells can proliferate and readily differentiate into type 1 AECs through modulation of Wnt signalling and activation of Fgfr2 via Fgf ligands to

restore epithelial integrity, and have therefore been referred to as the progenitors of the alveoli (Nabhan et al., 2018; Zacharias et al., 2018). Furthermore, a distinct subset of Wnt-responsive type 2 AECs has recently been identified in a single-cell atlas of the human lung, suggesting there might exist a homologous population to the *Axin2*+ progenitors of the mouse (Travaglini et al., 2020). Further work is needed to determine the significance of these cells in alveolar epithelial maintenance and regeneration.

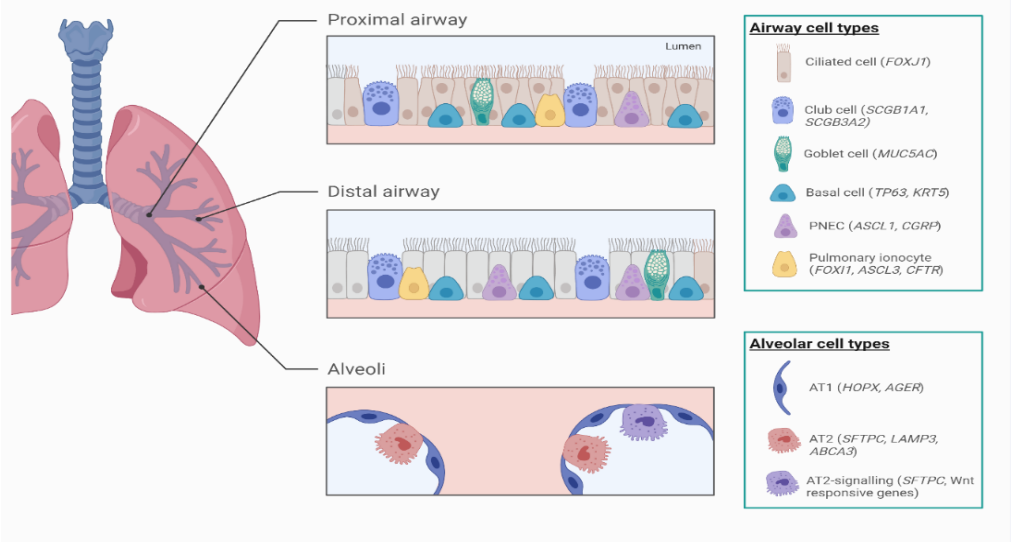


Figure 1-3. Epithelial cell types of the human lung and their respective marker genes.

The proximal airway is lined by pseudostratified ciliated epithelia and transitions to simple columnar epithelium in the terminal and respiratory bronchioles of the distal lung. It is interspersed with specialised airway cell types. The majority of the alveolar surface is lined by type 1 AECs and is interspersed with type 2 AECs of which a subfraction exhibit more stem-like features such as increased responsiveness to Wnt signals, and are therefore referred to as type 2 AEC signalling cells (Sun et al., 2022; Travaglini et al., 2020). Figure generated using Biorender.

1.3.4. Pulmonary surfactant

Pulmonary surfactant is a lipoprotein complex comprising four surfactant proteins and several phospholipid species, which is synthesised and secreted by the type 2 AECs of the lung. It serves to maintain low surface tension to prevent alveolar collapse at periods of low lung volume during respiration (Han & Mallampalli, 2015). Phosphatidylcholine, in the form of dipalmitoyl phosphatidylcholine (DPPC),

constitutes the majority of the phospholipid content of surfactant (Chakraborty & Kotecha, 2013). In conjunction DPPC, the surfactant proteins promote the surface tension reducing properties of surfactant and play critical roles in the innate immunity of the lung (Han & Mallampalli, 2015). Surfactant protein B (SFTPB) and SFTPC are hydrophobic proteins that are synthesised as larger, precursor proteins in the endoplasmic reticulum (ER). They undergo a series of post-translational proteolytic cleavages as they are trafficked from the Golgi to lamellar bodies (LBs), the lysozyme-related storage organelles of type 2 AECs, where the mature proteins reside (Weaver & Conkright, 2001). SFTPB is essential for LB biosynthesis (Sever et al., 2021; Stahlman et al., 2000). Similarly, the ATP-binding cassette sub-family member 3 (ABCA3) transporter, located on the membrane of LBs, is important for LB biosynthesis and is responsible for phospholipid transportation (Cheong et al., 2007). Following fusion of LBs with the membrane of type 2 AECs, pulmonary surfactant is secreted into the alveolar lumen and SFTPB and SFTPC promote formation of the lipid film by enhancing phospholipid adsorption and spreading (Weaver & Conkright, 2001). Surfactant proteins A (SFTPA) and D (SFTPD) are large, hydrophilic proteins and are mainly secreted into the alveolar space independent of LB-mediated exocytosis (Chakraborty & Kotecha, 2013). SFTPA is necessary for surfactant to construct into tubular myelin prior to lipid film formation at the air-liquid interface (KHUBCHANDANI & SNYDER, 2001). Additionally, SFTPA has important host defence and immunomodulatory functions alongside SFTPD (reviewed in Han and Mallampalli (2015); Pastva et al. (2007)). Type 2 AECs re-uptake most components of pulmonary surfactant and recycle them to the LBs where they are re-incorporated into surfactant (Whitsett et al., 2010). Of the four surfactant proteins, SFTPC is expressed exclusively in type 2 AECs and is therefore used as a definitive marker of these cells (Beers & Mulugeta, 2005).

1.3.5. Signalling in the adult alveolar niche

In the adult lung, the alveolar niche is maintained by dynamic signalling between multiple cell types including type 1 and 2 AECs, fibroblasts, immune cells, endothelial cells and pericytes (Hogan, 2020). Wnt signals originating from mesenchymal/fibroblast populations play an important role in maintaining the stemness of the subpopulation of *Axin2*+ type 2 AECs and regulating differentiation to type 1 AECs, as described earlier (Nabhan et al., 2018; Zepp et al., 2017). Findings from

both *in vitro* and *in vivo* mouse studies have suggested that stromal cell derived Bmp signalling might negatively regulate type 2 AEC stemness but has diverse roles in self-renewal and differentiation after lung injury (Chung et al., 2018). Furthermore, roles for Tgf-β, Fgf, Egf and Notch signalling, among others, have been implicated in maintenance of the alveolar niche (Hogan, 2020; Shiraishi et al., 2019; Zepp et al., 2017).

1.3.6. Human primary alveolar cells and organoids

Historically, immortalised cell lines have been used as a source of alveolar epithelial cells for research, but they fail to recapitulate many of the major features of their *in vivo* counterparts. As a result, there have been increasing efforts to establish human primary alveolar cells *in vitro* to address the lack of human adult models of the alveolar niche. This has been challenging, as primary type 2 AECs have a propensity to rapidly differentiate to type 1 AECs in culture (Beers & Moodley, 2017). Improvements in maintaining type 2 AEC phenotype have been reported in self-assembling 3D primary organoids compared to traditional 2D monolayer culture, but this was dependent on co-culture with fibroblasts and long-term maintenance was not possible (Evans & Lee, 2020). Just recently, several groups have successfully developed feeder-free, defined conditions for the expansion of primary type 2 AECs as 3D 'alveolospheres' using combinations of FGF, WNT, Noggin and EGF ligands with inhibitors of BMP, TGF-β, and GSK3β signalling to potentiate WNT activity (Katsura et al., 2020; Salahudeen et al., 2020; Youk et al., 2020). Primary type 2 AECs in these cultures can be clonally expanded for several months, dependent on continued WNT signalling, and retain the capacity to differentiate to type 1 AECs. These developments provide another source of alveolar epithelial cells for experimental purposes and gives further insight into the signalling requirements of type 2 AECs in the adult. However, there are limitations to these studies. Although culture medium was defined, expansion of organoids relied on the undefined mouse derived basement membrane matrix Matrigel™ and contained other animal-derived materials. In one study, organoids could be maintained for up to 10 months, but it was reported that markers of type 2 AECs, such as pro-SFTPC, were lost in some cells (Youk et al., 2020). Furthermore, it was not stated whether organoids could be cryopreserved and thawed for further use, meaning limited access to donor tissue for repeated cell isolation could still hinder this

approach. Moreover, a major consideration with primary cells is donor variability, and this was not investigated in these studies. Confounding variables such as heterogeneous genetic backgrounds need to be controlled when using primary cells to investigate disease mechanisms and response to viral infections.

1.4. Major diseases of the lung

1.4.1. Idiopathic pulmonary fibrosis

IPF is a chronic, age-related interstitial lung disease characterised by progressive, irreversible lung scarring (Maher et al., 2007). Excessive deposition of extracellular matrix (ECM) and destruction of normal alveolar architecture leads to increased lung stiffness and reduced capacity for gas exchange (Wuyts et al., 2013). This manifests clinically in patients as dyspnoea and a decline in lung function performance (Nakamura & Suda, 2015). The prognosis of IPF is considerably poor, with a mean survival rate of 3-5 years from the point of diagnosis. In the UK it is estimated that 32,500 individuals have IPF and approximately 6000 new cases are diagnosed every year (Snell et al., 2016). In most instances, IPF is a sporadic disease arising in adults over 60 years of age (Raghu et al., 2018). However, a subset of IPF cases cluster in families and are associated with earlier onset of disease with diagnosis at a mean age of 55 (Krauss et al., 2019; Steele et al., 2005). To date, only two therapeutic agents are licensed for use in the treatment of IPF: the tyrosine kinase inhibitor Nintedanib, and the anti-fibrotic Pirfenidone (Raghu et al., 2015). These compounds delay the rate of disease progression but do not stop the fibrotic response. Therefore, more predictive models of human IPF are needed to understand underlying disease mechanisms and identify new therapeutics.

Both genetic and environmental factors contribute to IPF pathogenesis and a model for disease progression has been proposed whereby repeated damage to genetically susceptible alveolar epithelium leads to onset of the fibrotic response (Mei et al., 2021). Several genetic risk factors have been identified for IPF, including variants in genes responsible for surfactant protein processing (SFTPC, SFTPA1), telomere maintenance, lung epithelial barrier integrity and cell cycle regulation, as well as single nucleotide polymorphisms (SNPs) in *MUC5B* (Michalski & Schwartz, 2020). A number of rare pathogenic mutations in the *SFTPC* gene have been discovered. Several of these

variants are localised to the highly conserved C-terminal BRICHOS domain of the *SFTPC* pro-protein and appear to cause mis-trafficking and aggregation of misprocessed SFTPC in the ER, preceding ER stress (Nogee et al., 2001; Ono et al., 2011; Thomas et al., 2002). Markers of ER stress and the unfolded protein response (UPR) in type 2 AECs are a frequent characteristic of IPF lungs (Korfei et al., 2008; Lawson et al., 2008). As a result, type 2 AECs adopt a hyper-functional state and release several pro-fibrotic growth factors, such as TGF-β1, resulting in myofibroblast differentiation. Myofibroblasts secrete excessive ECM leading to increased lung stiffness (Betensley et al., 2016). It is the complex interplay between damaged epithelial cells, immune cells, myofibroblasts and active components of the ECM which perpetuates fibrosis in IPF (Huang & Horowitz, 2014).

1.4.2. Models of lung fibrosis

IPF is a disease of unknown cause and late onset, primarily affecting individuals with a mean age of 65 years (Collard, 2010). Recapitulating the progressive nature of IPF and developing age-appropriate models of human respiratory fibrosis has been challenging. Modelling IPF has been met with challenges given its late-onset and relatively unknown cause. Animal models have been established to try and elucidate the mechanisms mediating lung fibrosis. Bleomycin-induced injury is the most common experimental animal model of fibrosis (Tashiro et al., 2017). Bleomycin is a chemotherapeutic antibiotic, and when injected intratracheally, can induce damage specifically in alveolar epithelial cells which initiates inflammation and fibrosis. Although some histopathological features between this model and human IPF are conserved, fibrosis resolves with single bleomycin treatment, inconsistent with the irreversible nature of human fibrosis (Moore et al., 2013). Such models fail to replicate the heterogeneity of human IPF and poorly predict therapeutic outcomes (Jenkins et al., 2017). A detailed review of animal models of fibrosis is provided in Moore et al. (2013). Multitude of cell-based models have also been developed to study lung fibrosis but there are several limitations, such as only focussing on the fibroblast response, lack of 3D tissue architecture, limited access human tissue, use of animal-derived materials and substrates and failure to induce injury in alveolar epithelial cells which is the prevailing hypothesis of IPF pathogenesis. These models have been extensively

reviewed in Sundarakrishnan et al. (2018) and their strengths and weaknesses are discussed.

1.4.3. Cystic fibrosis

Cystic fibrosis is an autosomal recessive disease caused by pathogenic mutations in the CFTR gene affecting approximately 1/3000 – 1/6000 individuals of European descent (Scotet et al., 2020) . The CFTR protein is a member of the adenosine triphosphate (ATP)- binding cassette (ABC) transporter family which functions as a transmembrane chloride (Cl-) channel on epithelial cells of the airways, pancreatic ducts and several other organs with an epithelial lining (Chen et al., 2021). CFTR regulates the movement of Cl- across the apical membrane of airway epithelial cells and limits sodium (Na+) reabsorption by influencing activity of the epithelial sodium channel (ENaC). This ensures the composition of the mucin-containing airway surface liquid (ASL) is maintained (Saint-Criq & Gray, 2017). In CF, ion transport mediated by CFTR is disrupted leading to enhanced NA+ reabsorption and dehydration of the ASL, viscous mucous secretions and reduced mucociliary clearance. As a result, recurrent bacterial infections are frequent in CF patients and cause chronic inflammation of the airways (Hanssens et al., 2021). Respiratory failure resulting from progressive lung disease is the most common cause of death in CF (Zolin et al., 2018). Over 2000 gene variants in CF have been identified and can be classified depending on the functional impact of the mutation. Some variants cause loss of the CFTR protein altogether, whereas others disrupt CFTR trafficking or impair ion conductance and channel stability (Elborn, 2016). Improvements in the treatment approach has drastically improved the life expectancy of CF patients and the median age of survival is now approximately 50 years, but there remains no cure, although advances in gene therapy show promise (McBennett et al., 2022; Scotet et al., 2020).

1.4.4. Chronic obstructive pulmonary dysfunction

COPD is a disease of the lung caused primarily, but not exclusively, by exposure to cigarette smoke resulting in airway inflammation, airflow obstruction and impaired respiratory function (Alfahad et al., 2021). COPD is the third leading cause of death worldwide, so is a significant global healthcare concern. Respiratory infections can trigger acute exacerbations in COPD which are associated with greater mortality

(Sapey & Stockley, 2006). Repetitive injury to the respiratory epithelium causes multitude of cellular processes such as immune cell activation/infiltration and cellular senescence/apoptosis. As a consequence of chronic inflammation, there is small airway narrowing and mucus hypersecretion (chronic bronchitis), and destruction of the alveolar architecture (emphysema), reducing the area for efficient gas exchange (Bourdin et al., 2009). COPD patients experience persistent coughing, excessive sputum production and dyspnoea (Devine, 2008). Given a large percentage of smokers do not develop COPD, genetic and other environmental risk factors, such as occupational exposure and particulate matter, play a role in its pathogenesis (Duan et al., 2020); "Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015" (2017). Cessation of smoking is the most effective intervention for COPD, but it does not prevent the inflammatory response once established (Rovina et al., 2013). Currently, treatments for COPD focus on managing symptoms and reducing acute exacerbations, such as the use of bronchodilators and anti-inflammatory drugs (Beeh, 2016). Further understanding of the mechanisms linking epithelial damage to chronic inflammation will deliver better therapies for COPD.

1.5. Human pluripotent stem cells and *in vitro* differentiation to type 2 AECs of the lung

1.5.1. Introduction to human pluripotent stem cells

In 1998, James A. Thompson isolated and established the first human embryonic stem cell (hESC) lines derived from the inner cell mass of blastocyst-stage human embryos (Thomson et al., 1998). Pluripotency is a transient state during early development, but the capacity of a single hESC to give rise to all cell types of the three embryonic germ layers can be retained *in vitro*. In culture, hESCs self-renew meaning they can be propagated indefinitely, providing an unlimited source of human stem cells with pluripotent potential (Romito & Cobellis, 2016). The pluripotent state is governed by a core transcriptional network consisting of *OCT4*, *SOX2* and *NANOG*. OCT4 and SOX2 form heterodimers, and in cooperation with NANOG, bind to several genomic targets, including their own promoter regions, to control expression of pluripotency genes and repress lineage differentiation (Chew et al., 2005; Li & Belmonte, 2017). This

transcriptional network is established by integration of extrinsic FGF2 and TGFβ/Activin/Nodal signalling which maintain self-renewal of human pluripotent stem cells (hPSCs) (Fathi et al., 2017; Vallier et al., 2005). Culture methods have been developed to routinely maintain hPSCs on the basement membrane matrix Matrigel™ or Geltrex™, both derived from Engelbreth-Holm-Swarm mouse tumours, in in the presence of recombinant FGF2 and TGFβ proteins (Viswanathan et al., 2014). The expression of specific cell surface markers has been used to further characterise hPSCs, including the glycosphingolipids SSEA3 and SSEA4, and the keratan sulphate antigens TRA-1-60 and TRA-1-81 (Adewumi et al., 2007; Goodwin et al., 2020).

The ethical concerns related to use of human embryos to derive hESCs were circumvented with the discovery of cellular reprogramming; the ability to reprogramme human somatic cells to a pluripotent state via over-expression of SOX2, OCT4, Krüppel-like factor 4 (KLF4) and c-MYC. These cells were termed human induced pluripotent stem cells (hiPSCs) (Takahashi et al., 2007). Like hESCs, hiPSCs are self-renewing and can be propagated in culture, express core pluripotency factors and can give rise to all three germ layers both *in vitro* and *in vivo* (Ye et al., 2013). Cellular reprogramming technology has provided a unique opportunity to access an unlimited reservoir of hiPSCs with little ethical considerations (Singh et al., 2015). Given their pluripotent potential, there has been a significant effort to develop methods to direct the differentiation of hPSCs to multitude of adult cell types of the body *in vitro*. This meant turning to our current knowledge of embryogenesis to recapitulate signalling events to drive the development of cell types of interest (Zakrzewski et al., 2019). One major advantage of hiPSCs is they can be derived from patients with known pathogenic mutations and subsequent differentiation to disease-relevant cell types can be a powerful tool for modelling disease. Moreover, given hiPSCs are amenable to genetic editing, isogenic cell lines from the same donor can be generated to control for confounding factors like heterogeneous genetic background (Soldner et al., 2011). Although protocols for directed differentiation of hPSCs to cells of the lung have been developed, they are still at relatively infant stages compared to other tissue types such as the liver and intestine. Given the stem cell properties and involvement in disease of the lung, development of a robust platform to generate type 2 AECs from hPSCs would be a highly valuable tool for regenerative medicine.

1.5.2. hPSC differentiation to DE and patterning of the gut tube *in vitro*

Like most directed differentiation protocols, attempts to specify type 2 AECs from hPSCs have aimed to recapitulate major stages of embryonic lung development. This begins with differentiation to embryonic endoderm, the origin of all epithelial cell types of the lung. High levels of embryonic Nodal signalling can be mimicked *in vitro* with recombinant activin A to activate downstream Smad2/3 signalling to drive DE induction (Brown et al., 2011; D'Amour et al., 2005). Formation of the DE layer is followed by patterning along its anterior-posterior axis into embryonic hindgut, midgut, and foregut domains. It is from the anterior foregut region where the lung buds first emerge. Continued activin A signalling in hPSC-derived DE is able to induce the expression of the foregut marker *SOX2* and this population can be maintained and expanded *in vitro*. These cells exhibit multipotency and retain the capacity to differentiate into cells expressing genes related to liver, pancreatic and lung lineages, and can give rise to distal airway epithelium (Hannan, Fordham, et al., 2013; Hannan et al., 2015). Green et al. (2011) screened 24 combinations of recombinant growth factors and small molecule inhibitors and demonstrated that only inhibition of BMP and TGFβ via NOGGIN and SB431542, respectively, following activin A induced-endoderm differentiation, could permit expression of the foregut marker *SOX2*, repress *CDX2* hindgut fate and maintain *FOXA2* (**Fig 1-3**). Upon further differentiation these cells expressed *NKX2.1* and lacked expression of the thyroid marker *PAX8*, indicative of lung identity. There is also a brief period during development where anterior foregut endoderm cells are exposed to the Wnt inhibitor DKK1, and improvements in lung differentiation efficiency have been reported by sequential inhibition of BMP, TGFβ and Wnt signalling during foregut patterning (Huang et al., 2014). Nonetheless, AFE cells remain poorly characterised, and no specific markers have yet been identified to isolate this transient multipotent population.

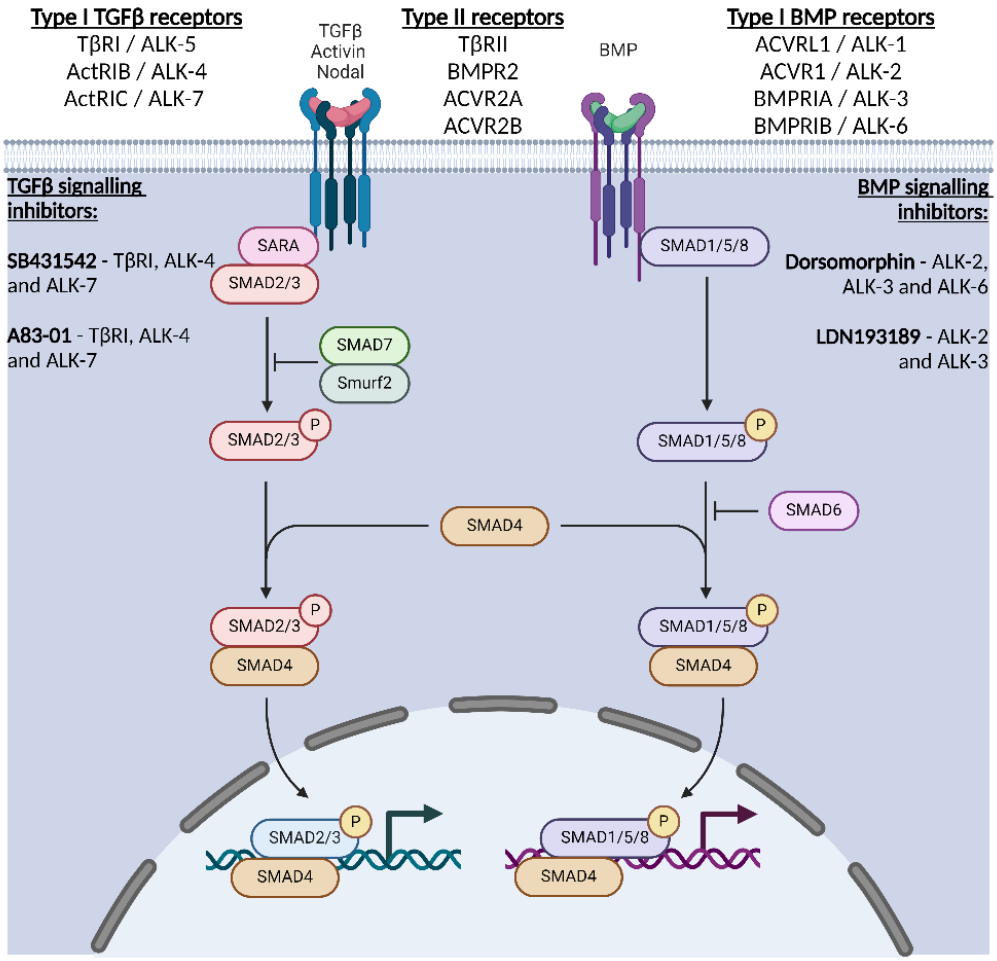

Type I TGFβ receptors
TβRI / ALK-5
ActRIB / ALK-4
ActRIC / ALK-7
TGFβ
Activin
Nodal
Type II receptors
TβRII
BMPR2
ACVR2A
ACVR2B
BMP
Type I BMP receptors
ACVRL1 / ALK-1
ACVR1 / ALK-2
BMPRIA / ALK-3
BMPRIB / ALK-6
TGFβ signalling inhibitors:
SB431542 - TβRI, ALK-4 and ALK-7
A83-01 - TβRI, ALK-4 and ALK-7
BMP signalling inhibitors:
Dorsomorphin - ALK-2, ALK-3 and ALK-6
LDN193189 - ALK-2 and ALK-3
SARA
SMAD2/3
SMAD7
Smurf2
SMAD1/5/8
SMAD4
SMAD6
P

Figure 1-4. Schematic of TGFβ and BMP signalling pathways.

The TGFβ superfamily comprises 33 genes encoding for several cytokines including 3 TGFβ isoforms, BMPs, activins and Nodal among several other members. TGFβ family receptors are characterised as both serine/threonine and tyrosine kinase receptors. There are 7 type I receptors, known as anaplastic lymphoma kinase (ALK) receptors, and 5 type II receptors that form specific heterotetrameric complexes upon binding of TGFβ family members. Activated type II receptors phosphorylate type I receptors that phosphorylate SMAD proteins for signal transduction. SMAD2/3 is phosphorylated primarily following binding of TGFβ isoforms, activin members and Nodal, and SMAD1/5/8/ is activated by binding of BMP signalling molecules. SMAD4 forms a complex with both sets of SMAD proteins enabling translocation to the nucleus and regulation of gene transcription. Several pharmacological inhibitors have been developed to inhibit TGFβ family signalling at the level of the receptors including SB431542, A83-01, dorsomorphin and LDN193189 (Heldin & Moustakas, 2016). Figure generated using Biorender templates.

1.5.3. Specification of NKX2.1 lung epithelial progenitor cells

The gut tube is further restricted to organ-specific domains and the prospective lung field emerges from a population of ventral AFE cells determined by the onset of *NKX2.1* expression (Zorn & Wells, 2009). It has been demonstrated that WNT, BMP and RA are the minimum signalling factors required to induce ventral identity in AFE cells and specify NKX2.1+ lung epithelial progenitors *in vitro* (Gotoh et al., 2014; Serra et al., 2017). Canonical WNT signalling is usually stimulated *in vitro* using a pharmacological inhibitor of glycogen synthase kinase-3 (GSK3), CHIR99021 (CHIR). GSK3 phosphorylates β-catenin leading to its destruction under steady state, but when inhibited β-catenin can translocate to the nucleus of the cell and form a complex T-cell factor/lymphoid enhancer-binding factor (TCF/LEF) to regulate gene transcription (figure) (Wu & Pan, 2010). To enable AFE to respond to WNT and BMP signals to induce lung fate, treatment with RA is critical (Rankin et al., 2016). Furthermore, FGF signalling appears to be dispensable for induction of lung fate but is required for differentiation to prospective thyroid cells (Serra et al., 2017). Given the lung and thyroid emerge

from the same ventral AFE region and both lineages express *NKX2.1*, it is necessary to confirm the absence of thyroid markers like *PAX8* during differentiation (Serra et al., 2017). NKX2.1 reporter cell lines have been generated to track and purify lung epithelial progenitors in culture. However, this limits the reproducibility of differentiation protocols and there are some considerations when using reporter lines. It was shown by scRNA-seq that airway and alveolar cells differentiated from an enriched *NKX2.1*-GFP population contained a proportion of non-lung endodermal cell types with low *NKX2.1* expression. This suggests *NKX2.1* could be expressed at low levels in these cells, or the reporter system did not faithfully select NKX2.1+ cells only (McCauley et al., 2018).

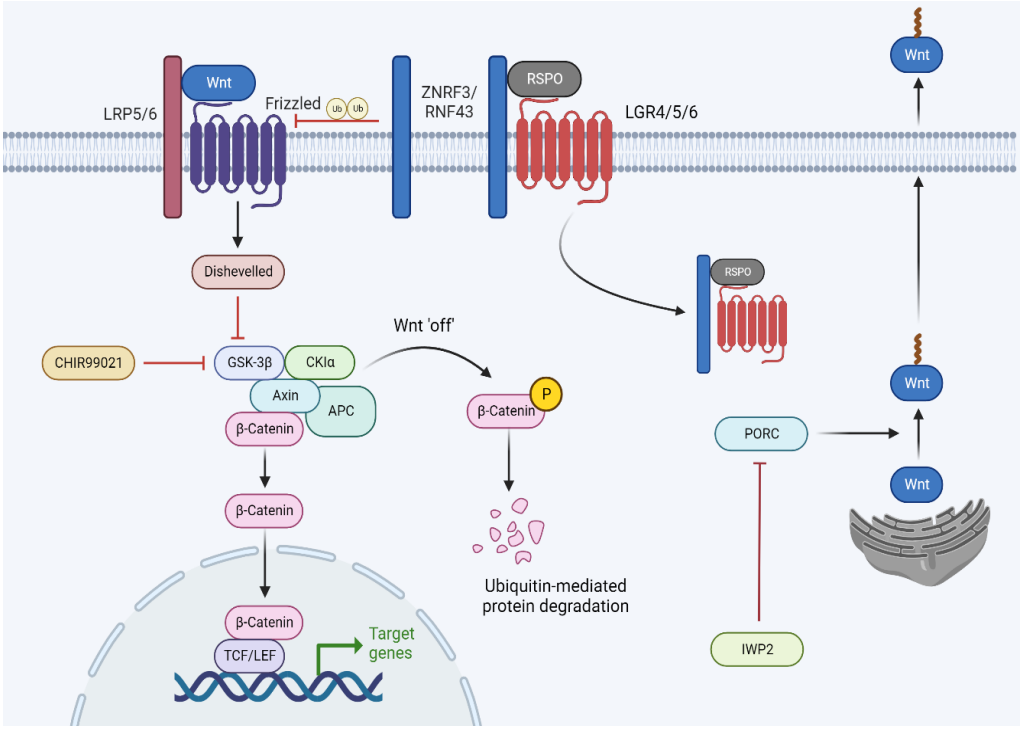


Figure 1-5. Wnt/β-catenin signalling pathway.

In the absence of Wnt receptor binding, a destruction complex formed by GSK-3β, Axin, APC and CKIα phosphorylates cytoplasmic β-catenin leading to its subsequent ubiquitination and degradation. Upon binding of extracellular Wnt ligands to the transmembrane receptor Frizzled, the LRP5/6 co-receptor is phosphorylated and forms a complex with Frizzled leading to activation of Dishevelled. This disrupts the destruction complex and prevents proteasomal degradation of β-catenin leading to accumulation of stabilised β-catenin in the cytoplasm. As a result, β-catenin translocates to the nucleus and binds to TCF/LEF proteins to regulate Wnt target gene expression. Pharmacological inhibition of GSK-3β by the small molecule CHIR99021 inhibits phosphorylation of β-catenin leading to its cytoplasmic stabilisation. Another layer of Wnt signalling regulation is mediated by the ubiquitin ligases ZNRF3 and RNF43 which degrade Frizzled and LRP5/6. Binding of the R-spondin (RSPO) family of signalling proteins to ZNRF3/RNF43 and the receptors LGR4/5/6 clear the ubiquitin ligases from the membrane and limit Wnt receptor degradation. Porcupine (PORC), a membrane-bound O-acyltransferase, adds palmitoyl groups to Wnt ligands to enable their secretion and binding to Frizzled. IWP2 is a pharmacological inhibitor of PORC which inhibits the Wnt signalling pathway (ter Steege & Bakker, 2021; Yang et al., 2016). Figure generated using Biorender templates.

1.5.4. Induction of type 2 AECs from hPSC-derived lung progenitor cells

To induce the maturation of lung epithelial progenitor cells to type 2 AECs *in vitro*, groups have isolated NKX2.1+ cells and embedded them in Matrigel™ to form three-dimensional (3D) alveolar organoids, or 'alveolospheres' (Gotoh et al., 2014; Jacob et al., 2017). Organoids are self-assembling 3D structures comprising either single or multiple cell types and they provide a more physiologically relevant architecture to promote specification of hPSC derived cells. Jacob et al. (2017) developed a *SFTPC*-tdTomato reporter cell line and cultured NKX2.1+ cells in the presence of CHIR, FGF7, the glucocorticoid dexamethasone (Dex) and activators of cAMP-dependent protein kinase A, including cAMP and 3-isobutyl-1-methylxanthine (IBMX), to stimulate maturation of the surfactant processing system of type 2 AECs (Gonzales et al., 2002). They also suggested removal of WNT signalling promoted upregulation of transcripts related to type 2 AEC maturation, but this remains to be rigorously validated by other groups. Another study used a similar method but co-cultured lung progenitor cells with human fetal lung fibroblasts supplemented with FGF7, Dex, cAMP and IBMX. They also developed feeder-independent conditions but supplementation with CHIR and the TGFβ inhibitor SB431542 was required (Gotoh et al., 2014; Yamamoto et al., 2017). These alveolar organoids could be expanded long-term in culture and contained type 2 AEC-like cells that express markers such as *SFTPC* and *SFTPA2* and have lamellar body-like inclusions as identified by electron microscopy (Jacob et al., 2017; Yamamoto et al., 2017). Functional characteristics such as surfactant processing and secretion are also reported. However, the population of SFTPC+ cells from these organoids are heterogeneous as they express varying levels of type 1 AEC markers, indicative of bipotent alveolar progenitors (Yamamoto et al., 2017). Moreover, sorted SFTPC+ cells appear to reflect more proliferative, immature type 2 AECs and share features with fetal-like alveolar cells (Jacob et al., 2017; Sun et al., 2021). This highlights a potential drawback of using *SFTPC* reporter lines to isolate type 2 AECs given SFTPC is expressed in the lung bud and alveolar epithelial progenitor cells during development.

1.5.5. Human lung organoids and modelling branching morphogenesis

Branching morphogenesis establishes the respiratory tree during development and hPSC derived lung models that aim to capture this process have been developed. In contrast to adherent monolayer culture described previously, Dye et al. (2015) induced the formation of self-aggregating foregut spheroids expressing NKX2.1 by treating hPSC derived DE with CHIR, FGF4, Noggin, SB431542 and an agonist of SHH signalling. After embedding in Matrigel™ and exposure to FGF10, these structures developed into 'human lung organoids' (HLOs) with SOX2+ proximal and SOX9+ distal domains distributed throughout. Markers of airway cell types including basal, club and ciliated cells were observed and were surrounded by mesenchyme. In addition, alveolar structures were identified with cells expressing markers of type 1 and 2 AECs., but often co-expressed with SOX9 suggesting a more immature progenitor-like phenotype. Building on this work, it has been shown that treatment of 3D foregut spheroids with CHIR, FGF7 and RA induced branching and gave rise to NKX2.1+ patterned organoids ('patterned lung organoids') (PLOs) with SOX2+ only proximal regions and SOX2+/SOX9+ distal budding regions that co-stained with ID2 and pro-SFTPC. When PLOs were passaged via needle shearing, they formed NKX2.1+ epithelial cyst-like structures with an increased proportion of SOX9+ proliferative domains. These were therefore termed bud tip organoids (BTOs). When treated with FGF7, BTOs generated cells expressing markers of goblet, club, neuroendocrine and type 1 and 2 AECs, indicating BTOs retained multipotent potential (Miller et al., 2019; Miller et al., 2018). Further optimisation led to improved induction of NKX2.1 expression in foregut spheroids which could develop into branched organoids with NKX2.1+ regions that co-expressed other tip progenitor markers SOX2, SOX9 and CPM. This population expands with extended culture and the induced bud tip progenitor cells can be isolated and cultured long-term to give rise to organoids (induced bud tip progenitor organoids) (iBTOs) which maintain NKX2.1 expression and are highly enriched for cells expressing bud tip progenitor markers. These iBTOs could give rise to both airway and alveolar cells under previously established conditions for differentiation. iBTOs are also transcriptionally similar to primary fetal bud tip progenitors as shown by scRNA-seq. This shows the bud tip progenitor state can be captured *in vitro*. However, even following enrichment of BTP cells, hindgut lineages persisted in culture highlighting the plasticity of these cells (Hein et al., 2022).

A similar approach has also been reported where NKX2.1+ lung bud organoids can give rise to highly branched structures in Matrigel™ in the presence of CHIR, FGF7, FGF10, BMP4 and RA. These organoids contained SOX9+ tip regions, expressed markers of type 2 AECs and rare goblet cells were observed, but other airway cells were absent. A small population of mesenchymal cells was also reported (Chen et al., 2017). This protocol was further optimised to include a maturation stage by treatment of lung bud organoids in rat tail collagen I gels with FGF10, FGF7, Dex, IBMX and cAMP. This led to branched organoids containing proximal airway basal, club, ciliated and goblet cells, and type 1 and 2 AECs in distal regions (de Carvalho et al., 2019; Rodrigues Toste de Carvalho et al., 2021). Such protocols can generate mixed cell organoids and are useful tools to study branching morphogenesis. However, they take extended periods in culture to be generated, there is a lack of control over the proportion of each cell type specified, some airway cell types are missing, and alveolar cells are rare. Furthermore, the cells generated lack maturity and reflect fetal stages of development. As such, these organoids might be more suited to studying mesenchymal-epithelial interactions in development and disease or as platforms to investigate the effect of viral infections on different lung epithelial cell types.

1.6. hiPSC-derived type 2 AECs for disease modelling

1.6.1. Modelling Hermansky-Pudlak syndrome using hiPSC-type 2 AECs

Hermansky-Pudlak syndrome (HPS) is a rare genetic disease with autosomal recessive inheritance and is associated with early-onset interstitial lung disease (Huizing et al., 2000). Mutations in HPS genes disrupt the protein assembly of biogenesis of lysosome-related organelle complexes (BLOCs) and lead to disrupted biogenesis and functioning of lysosome-related organelles (LROs) (Vicary et al., 2016). As the constituent LRO of type 2 AECs, LB dysfunction leads to impaired surfactant turnover and secretion, and precedes ER stress and type 2 AEC apoptosis, implicating epithelial damage in the initiation of HPS-associated ILD (Mahavadi et al., 2010). Of the ten currently identified HPS subtypes, pulmonary fibrosis manifests in HPS-1, HPS-2 and HPS-4 variants only (Vicary et al., 2016).

Strikoudis et al. (2019) introduced frameshift mutations in HPS-1, HPS-2, and HPS-4 genes in human embryonic stem cells (hESCs) using CRISPR/Cas9 and recapitulated

some features of lung fibrosis upon differentiation to lung organoids. Mutant organoids displayed greater deposition of ECM components fibronectin and collagen and enhanced expression of markers associated with mesenchymal cell activation. Expression of the cytokine Interleukin-11 (IL-11) was heightened in HPS-1 and HPS-2 mutant organoids and could directly induce upregulation of markers associated with myofibroblast activation. Markers of fibrosis were lost in HPS-$4^{-/-}$/IL-$11^{-/-}$ double-knockout lung organoids. A further study demonstrated HPS-2-mutant alveolar organoids differentiated from patient-derived hiPSCs could also reproduce some clinical features of HPS, such as enlarged LBs and defective surfactant secretion, but were not used to model the associated fibrosis. This is likely because of the lack of mesenchyme present in this system (Korogi et al., 2019).

1.6.2. Modelling SFTPB deficiency with hiPSC-type 2 AECs

SFTPB deficiency is an autosomal recessive disease causing fatal respiratory disease in newborn infants (Kurath-Koller et al., 2015). It is most commonly caused by a homozygous frameshift mutation in *SFTPB* resulting in loss of the SFTPB protein (NOGEE et al., 2000). Two groups have independently reprogrammed fibroblasts from patients with this *SFTPB* frameshift mutation to hiPSCs and differentiated them to lung organoids, albeit using different methods (Jacob et al., 2017; Leibel et al., 2019). Jacob et al. (2017) showed *SFTPB* mutant alveolar organoids contain type 2 AECs lacking SFTPB with mis-folded pro-SFTPC owing to lack of protein cleavage in absence of LBs. By correcting both *SFTPB* mutant alleles in hiPSCs via CRISPR/Cas9, these functions in alveolospheres were restored. Similar findings were reported in a second study where lentiviral infection of mutant hiPSCs with the *SFTPB* gene could restore SFTPB and SFTPC expression and surfactant processing in differentiated lung organoids (Leibel et al., 2019). These proof-of-principle studies demonstrate the feasibility of using hiPSCs to improve our understanding of respiratory disease and to validate gene therapy.

1.7. Aims and objectives of the project

Despite improvement in developing human models of lung development and disease, there are still several challenges that need to be addressed. Limited access to tissue and donor variability limits the application of primary type 2 AECs, and immortalised cell lines do not faithfully recapitulate major features of their *in vivo* counterparts. Developments in technologies like scRNA-seq have permitted more extensive investigation of human lung development in fetal samples, but limited access to tissue from later developmental stages means critical periods of alveolar development are yet to be modelled. Furthermore, current animal and cell models of lung disease such as IPF fail to capture all the features of human disease and therapeutics have routinely failed in clinical trials due to poor pre-clinical models. Directed differentiation of hiPSCs to type 2 AECs has emerged as a highly valuable tool for respiratory research that can circumvent many of the limitations discussed. However, current protocols rely on feeder cells, undefined basement membrane matrices and the presence of animal-derived materials. This is problematic for several reasons; findings of disease mechanisms are likely to be confounded by these factors, it hinders the reproducibility of protocols, it limits the therapeutic application of hiPSC-derived cells and raises several ethical and economic considerations. Furthermore, contaminating cell types of non-lung lineages can persist in culture suggesting refinement of the signalling pathways for differentiation is needed. Therefore, there remains an unmet need for fully defined, animal-free, human-relevant models of lung development and disease. Therefore, the objective of this project was to develop a novel, xeno-free *in vitro* platform for the directed differentiation of hiPSCs to type 2 AECs by recapitulating major stages of embryonic lung development that have the potential to be used for disease modelling purposes. During this project we aimed to do the following:

- **i.)** Establish a platform to differentiate normal and disease-relevant hiPSCs to type 2 AECs of the lung using cell surface markers for stage-specific cell enrichment in chemically defined, xeno-free conditions
- **ii.)** Explore and optimise the signalling pathways to improve the efficiency of lung specification
- **iii.)** Benchmark our cells derived under xeno-free conditions against published fetal lung data using transcriptomics.

2. Methods and Materials

2.1. Cell culture

All cells were maintained at 37°C and 5% CO_2. Basal media was made and stored at 4°C and growth factors/small molecules were added on the day of cell feeding. All medium was warmed to 37°C before feeding and all cell culture was performed under aseptic conditions in a type II biological safety cabinet.

2.1.1. Coating cell culture plasticware with Matrigel™

Growth factor reduced Matrigel™ (Corning, #354230) was diluted in cold DMEM at a concentration of 0.0347mg/cm^2 and was left at 37°C for at least 1 h before use for hiPSC maintenance, or overnight before seeding hiPSCs for differentiation. Prior to use, Matrigel coated plasticware was washed 3 times with Dulbecco's Phosphate Buffered Solution (DPBS) (Gibco™, #14190144).

2.1.2. Coating cell culture plasticware with Biolamina-521

Human recombinant laminin 521 (BioLamina, #LN521) was diluted in cold DPBS with Ca^{2+} and Mg^{2+} (Gibco™, #14040-133) at a concentration of 5 µg/ml for hiPSC maintenance in T25 flasks and 10 µg/ml for differentiation in 12-well plates. Plates were wrapped in parafilm and always left at 4°C overnight before seeding cells. Prior to use, plates were washed once with DPBS without Ca^{2+} and Mg^{2+}.

Table 2. 1. Composition of cell culture basal media

Basal Medium	Component	Final concentration	Supplier, identifier
Homebrew E8 basal medium (HB E8)	DMEM/F-12 50/50 Mix (1X)		Corning, #10-092-CM
	L-ascorbic acid 2-phosphate tri sodium salt	64 µg/ml	Sigma-Aldrich, #49752
	Heparin sodium salt	100 ng/ml	Sigma-Aldrich, #H3149
	Sodium selenite	14 ng/ml	Sigma-Aldrich, #S5261
	Transferrin	5 µg/ml	Sigma-Aldrich, #T3705

	Human recombinant insulin	20 µg/ml	Sigma-Aldrich, #11376497001
	Penicillin/Streptomycin (P/S) (5,000 U/ml)	1%	Gibco, #15070-063
RPMI differentiation (RPD) medium	RPMI-1640 Medium		Sigma-Aldrich, #R8758
	B-27™ supplement (50X)	2%	Gibco, #17504044
	Non-essential amino acids (NEAA) (100X)	1%	Gibco, #11140050
	P/S (5,000 U/ml)	1%	Gibco, #15070-063
Serum-free lung differentiation (SFL) medium	Iscove's Modified Dulbecco's Medium (1X)		Gibco, #12440053
	Ham's F-12 Medium (1X)		Corning, #10-080-CV
	N-2 supplement (100X)	0.5%	Gibco, #17502048
	B-27™ supplement (50X)	1%	Gibco, #17504044
	GlutaMAX™	1%	Gibco, #35050038
	Bovine serum albumin Fraction V (BSA) (7.5%)	0.6%	Gibco, #15260037
	α-monothioglycerol	4.5×10^{-4}M	Sigma-Aldrich, #M6145
	Ascorbic acid	50 µg/ml	Sigma-Aldrich, #A4544
	P/S (5,000 U/ml)	1%	Gibco, #15070-063
Xeno-free lung differentiation (XFL) medium	Iscove's Modified Dulbecco's Medium (1X)		Gibco, #12440053
	Ham's F-12 Medium (1X)		Corning, #10-080-CV
	Polyvinyl alcohol (PVA) (5%)	1 mg/ml	Sigma-Aldrich, #P8136
	Insulin-transferrin-selenium-X (ITS-X)	2%	Gibco, #51500-056
	CD lipid concentrate (100X)	0.1%	Gibco, #11905-031
	α-monothioglycerol (1.3%)	4.5×10^{-4}M	Sigma-Aldrich, #M6145
	Ascorbic acid 2-phosphate	64 µg/ml	Sigma-Aldrich, #A8960

	GlutaMAX™ (100X)	1%	Gibco, #35050038
	NEAA (100X)	1%	Gibco, #11140050
	P/S (5,000 U/ml)	1%	Gibco, #15070-063

2.1.3. Basal medium composition

2.1.3.1. RPMI differentiation (RPD) medium

To make 50 ml of RPD medium, supplement 48 ml RPMI 1640 with 1 ml of B27™ supplement, 500 µl NEAA and 500 µl P/S.

2.1.3.2. Serum-free lung differentiation (SFL) medium

For ~500 ml of SFL medium, combine 375 ml IMDM with 125 ml of Ham's F-12 and supplement with 5 ml of B27™ supplement, 5 ml of GlutaMAX™, 2.5 ml of N-2 supplement, 3.33 ml of 7.5% BSA Fraction V, 500 µl of ascorbic, 1.5 ml of α-monothioglycerol (in DMEM) and 2.5 ml of P/S.

2.1.3.3. Xeno-free lung differentiation (XFL) medium

For ~500 ml of animal-free differentiation medium, combine 351.7 ml IMDM with 117.2 ml Ham's F-12 and supplement with 10 ml PVA (5%), 10 ml ITS-X, 500 µl CD lipid concentrate, 1.5 ml α-monothioglycerol (in DMEM), 6.4 ml of ascorbic acid 2-phosphate, 5 ml of GlutaMAX™, 5 ml of NEAA and 5 ml of P/S.

Table 2. 2. List of growth factors and small molecules used for hiPSC maintenance and differentiation

Growth factor/small molecule	Final concentration	Supplier, identifier
Recombinant human FGF2	100 ng/ml	Peprotech, #100-18B
Recombinant human TGF-β	2 ng/ml	Peprotech, #100-21
Recombinant human Activin A	100 ng/ml	Peprotech, #120-14P
Recombinant human Wnt-3a	50 ng/ml, 100 ng/ml	R&D Systems, #5036-WN
Recombinant human KGF (FGF-7)	10 ng/ml	Peprotech, #AF-100-19
Recombinant human FGF-10	10 ng/ml	Peprotech, #100-26
LY294002	10 µM	Selleckchem, #S1105
CHIR99021	1 µM 3 µM	Sigma Aldrich, #SML1046
SB431542	10 µM	Tocris, #1614
Retinoic acid	100 nM	Sigma-Aldrich, #R2625
Dorsomorphin	2 µM	Tocris, #3093
Recombinant human BMP4	10 ng/ml	Peprotech, #120-05
Dexamethasone	50 nM	Sigma Aldrich, #D4902
3-Isobutyl-1-methylxanthine (IBMX)	100 µM	Sigma Aldrich, #I5879
Recombinant human R-spondin-1	250 ng/ml	Peprotech, #120-38
8-Bromoadenosine 3',5'-cyclic monophosphate sodium salt (cAMP)	100 µM	Sigma Aldrich, #B7880
IWP2	1 µM	Tocris, #3533
LDN 193189 dihydrochloride	250 nM	Tocris, #6053
Recombinant human EGF	50 ng/ml	Peprotech, #AF-100-15
A 83-01	1 µM	Tocris, #2939
Y-27632 dihydrochloride	10 µM	Tocris, #1254

All growth factors and small molecules are commercially available and were reconstituted as per the manufacturer's instructions before supplementing cell culture medium.

2.1.4. Derivation of hiPSC lines

hiPSCs were derived from skin fibroblasts of a female patient diagnosed with familial IPF. Skin biopsies were donated by Dr. Caroline Van Moorsel from the Division of Heart and Lungs, University Medical Centre Utrecht. Reprogramming was performed with non-integrating Sendai virus vectors carrying the Yamanaka factors OCT4, SOX2, KLF4 and c-MYC using the CytoTune™-iPS 2.0 Sendai Reprogramming kit (Invitrogen,

#A16517) as per the manufacturer's instructions. Briefly, dermal fibroblasts were seeded into 6-well plates at a density of 20,000 cells/well and after 24 h were transduced with Sendai virus vectors in human dermal fibroblast medium consisting of DMEM (Gibco, #21969-035) supplemented with Antibiotic-Antimycotic (100X) (2%) (Gibco, #15240-062), Foetal bovine serum (10%) (Gibco, #10270106), GlutaMAX™ supplement (1%) (Gibco, #35050-038), non-essential amino acids (100X) (1%) (Gibco, #11140-050) and Penicillin/streptomycin (10,000 U/ml) (1%) (Gibco, #15140122). Medium was replenished after 24h and cells were cultured for 3 – 7 days following transduction. Cultures at 80-90% confluency were dissociated with 0.05% trypsin-EDTA (Gibco, #25300062) and seeded into Matrigel™-coated 6-well plates in commercially available TeSR-E7 medium (StemCell Technologies, #05914). Medium was changed daily. When colonies started to form between days 13 – 22 after transduction, medium was changed to commercially available TeSR-E8 (StemCell Technologies, #05990). Colonies with a diameter of 500 – 1000 µm were isolated using EDTA (Invitrogen, #15575020) or ReLeSR (StemCell Technologies, #05872) and were picked and transferred into one well of a Matrigel™-coated 48-well plate in TeSR-E8 medium supplemented with Y-27632 (10 µM). hiPSC colonies were maintained in TeSR-E8 medium with daily media changes. This was performed by Dr Nick Hannan and Dr Peggy Lo at the University of Nottingham.

2.1.5. hiPSC cell lines

For initial experiments to characterise a pre-established lung differentiation protocol (as stated in the results section), the REBL-PAT hiPSC line was used. This cell line was reprogrammed from human dermal fibroblasts using Sendai virus vectors by Dr Gary Duncan at the University of Nottingham. For all subsequent experiments, the hiPSC line IPF1 was derived from a patient carrying a heterozygous mutation c.338A>G (Y113C) in exon 4 of the BRICHOS domain of *SFTPC* (*SFTPC*$^{WT/MUT}$). Based on prior characterisation for pluripotency markers and screening for differentiation capacity by Dr Ana Serna Valverde, clone 12 of the IPF1 hiPSC line (IPF1.12) was expanded and cryopreserved for experimental work. Isogenic pairs of cell lines were generated by genetically modifying the IPF1.12 line to generate homozygous-corrected wild-type BE.31 and BE.32 cell lines (G>A) (*SFTPC*$^{WT/WT}$) and the homozygous-mutant ABE.27

(SFTPC$^{MUT/MUT}$) cell line (A>G) (*SFTPC*$^{MUT/MUT}$). All gene editing was performed by Dr Sara Cuevas Ocaña as detailed below.

2.1.6. Genetic modification of hiPSCs using base editing

Base editing was performed to introduce single nucleotide variations in the heterozygous IPF1.12$^{WT/MUT}$ hiPSC line harbouring a *SFTPC* mutation c.338A>G to generate isogenic pairs of cell lines without inducing double stranded DNA breaks. Correction of the mutation (G>A) (*SFTPC*$^{WT/WT}$) was performed using a plasmid expressing a CMV promoter and spCas9 BEmax base editor with P2A-GFP marker and an ampicillin resistance cassette (pCMV_AncBEmax-P2A-GFP) (Addgene, #112094). For introduction of the mutation into the second allele (A>G) (*SFTPC*$^{MUT/MUT}$) a plasmid expressing a CMV promoter and spCas9 adenine BEmax base editor with P2A-GFP marker and an ampicillin resistance cassette was used (pCMV_ABEmax-P2A-GFP) (Addgene, #112095). Guide RNAs designed to target the DNA sequence of interest were annealed and cloned into vectors with a gRNA backbone with U6 promoter and ampicillin and puromycin resistance cassettes (Addgene, #51133).

Transformation of plasmids into NEB® 5-alpha Competent *E. coli* (NEB, #C2987H) was performed by incubating 50 ng of plasmid DNA with *E. coli* on ice for 30 min followed incubation at 42°C for 90 sec and cold shock on ice for 5 min. Lysogeny broth was added for *E. coli* growth and incubated in a shaking incubator at 37°C for 1 h at 225 rpm. The solution was transferred to LB agar with ampicillin (100 µg/ml) and incubated at 37°C overnight. Ampicillin resistant colonies were picked, and plasmid DNA was extracted.

Base editor and gRNA plasmids were transfected into IPF1.12 (SFTPC$^{WT/MUT}$) hiPSCs using the Amaxa™ 4D-Nucleofactor and P3 Primary Cell 4D-Nucleofector™ X Kit S (Lonza, #V4XP-3024) as previously optimised (Cuevas Ocana et al., 2022). IPF1.12 (SFTPC$^{WT/MUT}$) hiPSCs were dissociated with TrypLE™ and centrifuged for 4 min at 300 x *g*. Per transfection, 0.6 x 10^6 cells were subject to nucleofection. After, cells were resuspended in E8 medium supplemented with Y-27632 and plated on Matrigel™-coated 24-well plates at 37°C for 24 h. After 24 h, E8 medium was replaced without Y-27632 and cells were incubated with puromycin dihydrochloride (0.7 µg/ml) (Gibco, #A1113802) for transient puromycin selection. Transfection efficiency was assessed

by GFP expression. Puromycin-resistant cells were seeded at a density for clonal isolation. After colony formation, manual dissection was performed to isolate cells for clonal expansion and direct genomic DNA (gDNA) extraction for genotyping.

For direct gDNA extraction, the Phire Tissue Direct PCR Master (ThermoFisher, #F170S) was used and the protocol was previously optimised (Cuevas-Ocaña et al., 2023). The PCR reaction mix was as follows: 1 µl gDNA, 0.5 µl of each primer, 5 µl Phire master mix and 3 µl nuclease-free ddH_2O. The PCR reaction was performed with the following parameters: initial denaturation at 98°C for 5 min; 98°C for 5 sec, primer annealing at 60-71.4°C (primer dependent) for 15 sec and extension at 72°C for 1 min for 40 cycles; final extension at 72°C for 1 min and hold at 10°C. DNA products were analysed by 1% agarose gel electrophoresis using the LAS-4000 imaging system (Fujifilm Life Science). Restriction digestion of PCR products was performed using the HpyCH4v restriction enzyme (New England Biolabs, #R0620L) to confirm base editing at the population level. Restriction digestion was performed with 0.5 µl enzyme in rCutSmart™ Buffer (New England Biolabs, #B6004S) and 5 µl of the amplified PCR product at 37°C for 3 h. Digested products were analysed as described above. PCR products (1 ng/µl per 100 bp as per the manufacturer's instructions) were sent to SourceBioScience, Nottingham for Sanger sequencing and analysis was performed using SnapGene Viewer 6.2.2. Dr Sara Cuevas Ocaña performed all gene-editing in this study as previously described (Cuevas-Ocaña et al., 2023).

2.1.7. Karyotype analysis of hiPSCs

hiPSCs were seeded into T25 flasks and at after 48 h (~80% confluency) they were incubated with KaryoMAX™ Colcemid™ solution (100 ng/ml) (Gibco, #15212012) for 1 h at 37°C. Medium was collected and cells were dissociated with TrypLE and centrifuge for 4 min at 160 x *g*. The cell pellet was carefully resuspended in 0.6% tri-sodium citrate (Fisher Scientific, #S/3320/53) and incubated for 30 min at RT followed by centrifugation for 5 min at 300 x *g*. For fixation, the cell pellet was carefully resuspended in 16.7% acetic acid glacial (Fisher Scientific, #A/0400/PB17) in methanol and stored at -20°C. Sample preparation was prepared by Dr Peggy Lo and Dr Sara Cuevas Ocaña at the University of Nottingham before karyotyping was performed at the Cytogenetics department of Nottingham City hospital according to the International System for Human Cytogenetic Nomenclature.

2.1.8. Maintenance of human induced pluripotent stem cells

Human induced pluripotent stem cells were maintained in HB E8 basal medium supplemented with FGF2 (100 ng/ml) and TGF-β (2 ng/ml) (HB E8) on growth factor reduced Matrigel™-coated T25 flasks. hiPSCs were passaged every 3-4 days when ~80% confluent. Briefly, hiPSCs were washed once with DPBS and incubated with TrypLE™ (Gibco™, #12604021) at room temperature (RT) for 2-4 min. TrypLE™ was removed and hiPSCs were collected in HB E8 basal medium and seeded at a ratio of 1:12-1:20 (dependent on cell line and growth rate) in HB E8 supplemented with Y-27632 (10 µM). After 24 h, HB E8 minus Y-27632 was replenished and medium was replaced daily.

2.1.9. Cryopreservation of hiPSCs

For cryopreservation, hiPSCs were dissociated as described above (section **2.1.4**) and centrifuged for 4 min at 300 x *g*. Cells were resuspended in 10% DMSO (Sigma Aldrich, #D2650) in 90% heat inactivated fetal bovine serum (HI FBS) (Gibco, #10500-064) and 500 µl of suspension was placed in each cryovial. Cryovials were placed in a Mr. Frosty™ Freezing container at -80° C for 24-48 h before being transferred to liquid nitrogen for long-term storage.

2.1.10. Pre-optimised differentiation of hiPSCs to lung organoids

Before optimisation, hiPSCs were differentiated to lung organoids using a pre-established protocol to recapitulate the major stages of embryonic lung development. To prepare for differentiation, hiPSCs at ~80% confluency were washed once with DPBS and incubated with TrypLE™ for 2-4 min at RT. TrypLE™ was removed and hiPSCs were collected in HB E8 basal medium and centrifuged for 4 min at 300 x *g*. hiPSCs were resuspended in HB E8 supplemented with Y-27632 (10 µM) and seeded into Matrigel™-coated 12-well plates at a density of 10,000 – 40,000 cells/cm^2 (dependent on cell line; during optimisation experiments the density is specified in text). hiPSCs were replenished with HB E8 minus Y-27632 after 24h. After 48 h from seeding (D0), differentiation was induced by incubating hiPSCs with RPD medium supplemented with Wnt3a (50 ng/ml) and Activin A (100 ng/ml) and medium was replaced daily (D0-D3). After 3 days of DE differentiation, cells were cultured with RPD medium

supplemented with activin A (50 ng/ml) (D3-D5) with daily medium changes to induce foregut endoderm. For differentiation to lung epithelial progenitor cells, on D5 medium was replaced with RPD medium supplemented with CHIR99021 (1 µM), BMP4 (5 ng/ml) and retinoic acid (250 nM) with daily medium changes for 12 days (D5-D17). On day 17, cells were washed once with DPBS and collected by mechanical dissociation. Cell aggregates were centrifuged for 30 sec at 200 x *g* and transferred into growth factor reduced phenol red-free Matrigel™ (Corning, #356231) and plated as 50 µl droplets on a Nunc™ treated 48-well plate (approximately 1 well of a 12-well plate generated 8 droplets). After allowing the Matrigel™ to solidify at 37°C for 10 min, the droplets were cultured with RPD supplemented with CHIR99021 (3 µM), FGF7 (10 ng/ml), FGF10 (10 ng/ml), dexamethasone (50 nM), IBMX (100 nM), cAMP (100 nM) and Y-27632 (10 µM). After 48 h, medium was replaced to remove Y-27632 and was replenished every 2 days to generate self-assembling organoids. Organoids were passaged every ~6-9 days (depending on growth rate). Briefly, organoids were mechanically dissociated with the tip of a 5 ml stripette and collected in cold DPBS and centrifuged for 1 min at 200 x *g*. This process was repeated until the Matrigel™ had fully dissolved. Disaggregated organoids were resuspended in medium and re-plated as described above.

2.1.11. Directed differentiation of hiPSCs to DE

The following steps describe the optimised chemically defined, xeno-free differentiation of hiPSCs to type 2 AECs.

To induce DE, hiPSCs were seeded into Laminin-521- or Matrigel™-coated 12-well plates in HB E8 supplemented with Y-27632 at a seeding density of between 20,000 – 40,000 cells/cm^2. HB E8 medium was replenished after 24 h and differentiation was initiated after 48 h from seeding on day 0 with XFL medium supplemented with CHIR99021 (3 µM) (D0-D1), activin A (100 ng/ml) (D0-D4) and LY294002 (10 µM) (D0-D2). Medium was replaced daily until differentiation day 4. At this stage, DE cells were assessed for CXCR4 and CKIT expression via flow cytometry (section **2.2.2**) and cultures with a differentiation efficiency >90% were continued for anterior foregut endoderm specification.

For further optimisation experiments, RPD or XFL medium was supplemented with combinations of the following growth factors: CHIR99021 (3 µM), activin A (100 ng/ml), LY294002 (10 µM), wnt3a (50 ng/ml), BMP4 (10 ng/ml) and FGF2 (16 ng/ml) and was replenished daily until differentiation D3, 4 or 5. This is specified in the results section.

2.1.12. Patterning of DE to anterior foregut endoderm

Day 4 DE cells were washed once with DPBS and incubated with gentle cell dissociation reagent at RT for 2-3 min. GCDR was removed and DE cells were collected as small aggregates and passaged at a ratio between 1:2 – 1:21 (dependent on cell line and hiPSC seeding density; for optimisation experiments the ratio is specified in results section) into Laminin-521- or Matrigel™-coated Nunc™-treated 12-well plates in XFL medium supplemented with SB431542 (10 µM) and dorsomorphin (2 µM) (D4-D7).

For further optimisation experiments, XFL medium was supplemented with combinations of SB431542 (10 µM), dorsomorphin (2 µM), IWP2 (1 µM), LDN193189 (250 nM) and A83-01 (1 µM). This is specified in the results section.

2.1.13. Specification of lung epithelial progenitor cells

Differentiation to lung epithelial progenitor cells was induced on day 7 of differentiation with XFL medium supplemented with CHIR (3 µM), or Wnt3a (100 ng/ml) and R-spondin (250 ng/ml) (specified in results section text), BMP4 (10 ng/ml) and Retinoic acid (100 nM) for a total of 6-8 days, and medium was changed daily.

For further optimisation experiments, RPD, SFL or XFL medium was supplemented with combinations of the following growth factors: CHIR (3 µM), Wnt3a (50 ng/ml; 100 ng/ml), R-spondin (100 ng/ml, 200 ng/ml, 250 ng/ml), BMP4 (10 ng/ml), Retinoic acid (100 nM, 250nM), FGF2 (10 ng/ml), FGF7 (10 ng/ml) and FGF10 (10 ng/ml). This is specified in the results section.

2.1.14. Differentiation to type 2 AEC organoids

Lung progenitor cells were washed once with DPBS and incubated with 500 µl/well 0.05% trypsin-EDTA (Gibco™, #25300062) at 37°C for 10-15 min. After 10 min cells were visualised under the microscope to check for dissociation to single cells. If not

fully dissociated, cells were left at 37°C for another 5 min. 1 ml trypsin-quenching medium was added to the wells and cell were collected and centrifuged for 5 min at 300 x *g*. Cells were washed once with DMEM and centrifuged for 5 min at 300 x g. At this stage, cells were either sorted based on the expression of CPM (refer to section **2.2.2**) or were resuspended in phenol red-free Matrigel™ at a density of 20,000-40,000 cells/droplet and transferred as 50 µl droplets onto the bottom of a 48-well plate. After allowing Matrigel™ to solidify at 37 °C for 10 min, droplets were incubated with XFL medium supplemented with CHIR99021 (3 µM), FGF7 (10 ng/ml), FGF10 (10 ng/ml), dexamethasone (50 nM), IBMX (100 nM), cAMP (100 nM). CloneR™2 (10%) was added for the first 48 h and medium was replaced every 2 days.

For optimisation experiments, XFL medium was supplemented with

2.1.15. Passaging type 2 AEC organoids

Organoids were passaged every ~9-18 days (depending on growth rate). To removed organoids from the Matrigel™ droplets, they were incubated with dispase® at 37°C for 1 h. After the Matrigel™ had dissolved, organoids were collected in DMEM and centrifuged for 4 min at 200 x *g*. Organoids were washed once in DMEM, centrifuged for 4 min at 200 x *g*, and resuspended in 0.05% trypsin-EDTA at 37°C for 10-15 min. A small suspension was taken after 10 min to check for single cell dissociation. If not fully dissociated, cells were left at 37°C for another 5 min. Cells were gently resuspended in trypsin-quenching medium and centrifuged for 5 min at 300 x *g*. Cells were washed once in DMEM and centrifuge for 5 min at 300 x g. AT this stage, single cells were either sorted based on the expression of SLC34A2 (NaPi2b) (refer to section on FACS) or re-plated at a density of 20,000 cells/droplet as described above (section **2.1.7**). NaPi2b+ cells were cultured with XFL medium supplemented with CHIR99021 (3 µM), FGF7 (10 ng/ml), FGF10 (10 ng/ml), dexamethasone (50 nM), IBMX (100 nM) and cAMP (100 nM), and cloneR™2 for the first 48 h only. Medium was changed every 2 days.

2.2. Fluorescence activated cell sorting (FACS) and flow cytometry

2.2.1. Flow cytometry reagents

- **Trypsin quenching medium:** 5 mL heat-inactivated (HI) FBS + 45 mL DMEM.
- **FACS buffer (2%):** 1 mL HI FBS + 49 mL DPBS. Filter and store at 4°C
- **Wash buffer (FBS/PBS (1%)):** 500 µL HI FBS + 50 mL DPBS.
- **Sorting buffer (2%):** 200 µL HI FBS + 10 mL HBSS+ (Gibco, #14025092) (on the day of sorting, add Y-27632 (10 µM) and propidium iodide (PI) (10 µM) (Sigma Aldrich, #P4170). Filter and store at 4°C.
- **Dispase:** Add dispase II powder (Gibco™, #17105041) to DMEM at a concentration of 2 mg/ml. Filter and store at 4°C.

Table 2. 3. Antibodies for flow cytometry and FACS

Antibody	Description	Dilution	Supplier, identifier
Primary antibodies			
CXCR4-PE	Mouse monoclonal IgG2a antibody (12G5) to human CD184 (CXCR4), phycoerythrin (PE) conjugated	1:50	Invitrogen, #MHCXCR404
CKIT-APC	Mouse monoclonal antibody (104D2) to human CKIT (CD117), allophycocyanin (APC) conjugated	1:50	Invitrogen, #CD11705
CPM	Mouse monoclonal antibody to human carboxypeptidase M (CPM)	1:200	Wako/Fujifilm, #014-27501
NaPi2b	Mouse monoclonal antibody to human NaPi2b (SLC34A2) (MX35)	0.8µl/10^6 cells	Ludwig Institute for Cancer Research, NY
EpCAM	Goat polyclonal antibody to human EpCAM	1.25µl/10^6 cells	R&D systems, #AF960
Secondary antibodies/isotype controls			
IgG1-PE	Mouse IgG1, κ Isotype control, PE	1:50	BioLegend, #400113

IgG1-APC	Mouse IgG1 Isotype control, APC	1:50	Invitrogen, #MA5-18093
Donkey anti-mouse-647	Alexa Fluor™ 647 AffiniPure Donkey Anti-Mouse IgG (H+L)	1:300	Jackson Immunoresearch, #715-605-150
Donkey anti-goat-488	Alexa Fluor™ 488 Donkey anti-Goat IgG (H+L) Cross-Adsorbed	1:300	Invitrogen, #A-11055

2.2.2. Flow cytometry analysis of DE cells

On day 4 of differentiation, two wells of a 12-well plate were washed once with DPBS and incubated with 500 µl Gentle-cell dissociation reagent (GCDR) (Stemcell Technologies, #100-0485) at 37°C for 5 min. Cells were collected in DMEM and centrifuged for 5 min at 300 x *g*. This step was repeated once, and cells were resuspended in 300 µl FACS buffer and divided into three 1.5 ml microcentrifuge tubes. 2µl of conjugated antibodies or isotype controls were co-stained as follows: unstained; IgG1 isotype-PE/IgG1 isotype-APC; CXCR4-PE/CKIT-APC (**table 2.3**) and incubated on ice for 30 min. After 15 min, tubes were gently flicked and placed back on ice for an additional 15 min. 1 ml of wash buffer was added to each tube and cells were gently resuspended and centrifuged for 5 min at 300 x *g*. This step was repeated once. Cells were resuspended in 350 µl FACS buffer and transferred to 5 ml FACS tubes. Flow cytometry was performed using a FC500 flow cytometer (Beckman Coulter) or an ID7000 Spectral Cell Analyser (Sony) (specified in results section). Gates were set according to the isotype control samples.

2.2.3. FACS of 2D lung progenitor cells

To sort lung epithelial progenitors, cells were dissociated as described previously for transfer to 3D culture (refer to section **2.1.7**). After dissociation, cells were resuspended in DMEM and counted. Cells were centrifuged for 5 min at 300 x *g* and were resuspended in an appropriate volume of FACS buffer to achieve a concentration of 10^6 cell/100 µl. For the unstained (no antibody) and secondary antibody only (no primary antibody) controls, 10^5 cells were each transferred to 1.5 ml microcentrifuge tubes and made to a volume of 100 µl with FACS buffer. For primary antibody staining,

cells were incubated with mouse anti-human CPM in FACS buffer (0.5 µl/10^6 cells) at 4°C for 15 min. After 15 min, cells were resuspended in wash buffer and centrifuged for 5 min at 300 x *g*. For secondary antibody staining, cells were resuspended in FACS buffer at a concentration of 10^6 cells/100 µl and were incubated with donkey anti-mouse-647 (0.33 µl/10^6 cells) at 4° C for 15 min. After 15 min, cells were resuspended in wash buffer and centrifuged for 5 min at 300 x *g*. This step was repeated once. All conditions were resuspended in sorting buffer supplemented with Propidium Iodide (PI) and Y-27632 (10 µM) and cells were passed through a 70 µm nylon cell strainer pre-coated with DMEM into sterile 5 ml FACS tubes. FACS was performed using an Astrios EQ cell sorter (Beckman Coulter). Gates were set according to control cells. Sorted cells were collected in 500 µl XFL basal medium supplemented with Y-27632 (10 µM) and were seeded into phenol red-free Matrigel™ droplets at a concentration of 20,000 cells/droplet as described above (refer to section **2.1.7**).

2.2.4. FACS of 3D type 2 alveolar organoids

Organoids were dissociated as previously described for single cell passaging (refer to section **2.1.8**) to prepare for sorting. The same protocol for staining cells was used as above and 10^5 cells each were taken for unstained (no antibody) and secondary antibody only (no primary antibody) controls (section **2.2.2**). For primary antibody staining, 10^6 cells per 100 µl FACS buffer were incubated with mouse anti-human NaPi2b (0.8 µl/10^6 cells) and goat anti-human EpCAM (1.25 µl/10^6 cells) at 4°C for 15 min. Cells were resuspended in wash buffer and centrifuged for 5 min at 300 x *g*. Cells were resuspended in FACS buffer at a concentration of 10^6 cells/100 µl and incubated with donkey anti-mouse-647 (0.33 µl/10^6 cells) and donkey anti-goat-488 (0.33 µl/10^6 cells) at 4°C for 15 min. Cells were resuspended in wash buffer and centrifuged for 5 min at 300 x *g*. This was repeated once. The same steps were used for staining secondary antibody only controls. All conditions were resuspended in sorting buffer supplemented with PI and Y-27632 (10 µM) and cells were passed through a 70 µm nylon cell strainer pre-coated with DMEM into sterile 5 ml FACS tubes. FACS was performed using an Astrios EQ cell sorter (Beckman Coulter). Gates were set according to control cells. Sorted cells were collected in 500 µl XFL basal medium supplemented with Y-27632 (10 µM) and were seeded into phenol red-free Matrigel™ droplets at a concentration of 20,000 cells/droplet as described above (refer to section **2.1.7**).

2.2.5. Intracellular flow cytometry

For intracellular flow cytometry of pluripotency markers, hiPSCs were dissociated as previously described (**2.1.8**) and centrifuged for 4 min at 300 x *g*. Cells were fixed in 4% paraformaldehyde (PFA) (ThermoFisher, #28908; diluted in DPBS) for 20 min at 4°C and washed once with 10 volumes of DPBS. Cells were then resuspended in 1 ml of ice-cold methanol and incubated overnight at -20°C for permeabilization. Samples were washed once with 10 volumes of DPBS and centrifuged for 3 min at 300 x *g*. Cells were resuspended in FACS buffer (2% FBS) and incubated with primary antibodies for 2 h at RT. Samples were washed with 1% FBS and incubated with secondary antibodies in FACS buffer (2% FBS) for 30 min at RT. The following primary antibodies were used: Goat anti-human NANOG (1:100; R&D Systems, #AF1997), mouse anti-human SOX2 (1:50; R&D Systems, #MAB2018) and mouse anti-human OCT4 (1:50; R&D Systems, #MAB17591). The following secondary antibodies were used: AlexaFluor™ 488 Donkey anti-goat (1:400; ThermoFisher, #A11055) and AlexaFluor™ 647 Donkey anti-mouse (1:400; ThermoFisher, #A31571).

2.3. Molecular biology

2.3.1. RNA extraction and cDNA synthesis

Total RNA was extracted from cells using the RNeasy® Mini Kit (Qiagen, #74106) as per the manufacturer's instructions. Briefly, cells were lysed with 300 µl RLT buffer and mixed with one volume of 70% ethanol. The lysate was transferred to a RNeasy spin column and centrifuged to bind RNA. The column was washed with buffer RW1. A stock solution of DNase I (Sigma-Aldrich, #DNASE70) was prepared according to the manufacturer's instructions and on-column DNase digestion was performed at RT for 15 min. The column was washed with RW1 buffer and centrifuged. This was followed by two washes with RPE buffer. The column was dried by centrifugation for 1 min. RNA was eluted from the silica-gel-membrane in 20 µl nuclease-free water (NFW) (Invitrogen, #AM9932) and RNA concentration and purity was measured using a NanoDrop™ 2000 spectrophotometer (Thermo Scientific™, #ND-2000). For cDNA synthesis, a total of 500 ng RNA per sample was reverse transcribed in a total reaction volume of 20µl. The reaction was as follows: 500ng RNA was made up to a total volume of 11.875 µl NFW and was combined with 0.5 µl random primers (Promega, #C1181)

and 1 µl dNTPs (Promega, #U1511) per reaction, and was incubated at 65° for 5 min to denature secondary structures and then snap-cooled on ice. A master mix with a total volume of 6.625 µl per reaction containing 4 µl 5x 1st strand buffer, 2 µl 0.1M DTT, 0.5 µl RNase OUT (Invitrogen, #10777-019) and 0.125 µl SuperScript™ II reverse transcriptase (Invitrogen, #18064071) was added to each sample and incubated in a thermocycler with the following settings: 25°C for 10 min, 42°C for 50 min and 70 °C for 15 min. Each cDNA sample was diluted to a final volume of 600 µl with nuclease-free water and was stored at -20°C.

2.3.2. Quantitative real-time polymerase chain reaction (qRT-PCR)

RT-qPCR was performed in 96-well PCR plates using the SensiMix™ SYBR® Low-ROX kit (Bioline, #QT625-05). Each reaction was made up to a total volume of 15 µl as follows: 5 µl cDNA, 7.5 µl SYBR green SensiMix (2x), 1.3 µl nuclease-free water, 0.6 µl forward primer (5 µM) and 0.6 µl reverse primer (5 µM). qRT-PCR was performed on a 7500 real-time PCR system (Applied Biosystems™) using the following programme: initial denaturation for one cycle at 95°C for 5 min and 40 cycles at 95°C for 15 sec, 60°C for 30 sec and 72°C for 30 sec. The melt curve stage was performed at 95° C for 15 sec, 60°C for 60 sec, 95°C for 30 sec and 60°C for 15 sec. Relative gene expression was calculated using the comparative CT method and based on the average CT value of the technical triplicate, normalised to an internal reference gene, porphobilinogen deaminase (PBGD), and displayed either as relative expression to PBGD ($\Delta\Delta$CT) or as fold change in expression over control cells using 2-$\Delta\Delta$CT (Livak & Schmittgen, 2001).

Table 2. 4. Primer sequences for qRT-PCR

Gene	Forward primer	Reverse primer
SCGB1A1	TTCAGCGTGTCATCGAAACCCC	ACAGTGAGCTTTGGGCTATTTTT
SOX2	GCACATGAAGGAGCACCCGGATTA	CGGGCAGCGTGTACTTATCCTTCTT
FOXJ1	GAGCGGCGCTTTCAAGAAG	GGCCTCGGTATTCACCGTC
KRT5	GGAGTTGGACCAGTCAACATC	TGGAGTAGTAGCTTCCACTGC
TP63	ACTTCACGGTGTGCCACCCT	GAGCTGGGGTTTCTACGAAACGCT
MUC5AC	CCATTGCTATTATGCCCTGTGT	TGGTGGACGGACAGTCACT
CFTR	CTATGACCCGGATAACAAGGAGG	CAAAAATGGCTGGGTGTAGGA
AQP5	GCTCACTGGGTTTTCTGGGTA	CCTCGTCAGGCTCATACGTG
MUC5B	ACAATGGCACCTTCTACGGG	TACTCAAAGCCCTGGGGACA

SFTPB	CAGCACTTTAAAGGACGGTGT	GGGTGTGTGGGACCATGT
SFTPC	AGCAAAGAGGTCCTGATGGA	CGATAAGAAGGCGTTTCAGG
NKX2.1	CTCATGTTCATGCCGCTC	GACACCATGAGGAACAGCG
SCGB3A2	GAGGGGCTAAGGAAGTGTGT	CCATCCACCTCCGCTCTTTA
CALCA	GGGGTGTGGTGAAGAACAAC	CCATGGAGCCTTTCCTACAA
SYP	TTTGTGAAGGTGCTGCAATG	ACCTCGATGCTGAGGTCACT
ABCA3	GCCCTCTTTACACTCAGTTTTCA	GACGAGCAGTTGTCGTACCTAAT
LAMP3	ACTTCAACATCGACCCCAAC	CACTCACGCACTTGAAGGAA
SLC34A2	TCGCCACTGTCATCAAGAAG	CTCTGTACGATGAAGGTCATGC
PDPN (T1α)	TCCAGGAACCAGCGAAGAC	CGTGGACTGTGCTTTCTGA
HOPX	GCCTTTCCGAGGAGGAGAC	TCTGTGACGGATCTGCACTC
AGER	GCCACTGGTGCTGAAGTGTA	TGGTCTCCTTTCCATTCCTG
CPM	GCGCTGGATTTCAACTACCAC	TCCCGCCCAACAGTCTCAT
AFP	AAACTATTGGCCTGTGGCGA	TTTTGTCCCTCTTCAGCAAAGC
ALB	CTCGGCTTATTCCAGGGGTG	AAAGGCAATCAACACCAAGGC
CXCR4	CACCGCATCTGGAGAACCA	GCCCATTTCCTCGGTGTAGTT
FOXA2	GGGAGCGGTGAAGATGGA	TCATGTTGCTCACGGAGGAGTA
GATA4	TCCCTCTTCCCTCCTCAAAT	TCAGCGTGTAAAGGCATCTG
HHEX	TGCATAAAAGGAAAGGCGGC	TTGCTTTGAGGGTTCTCCTGT
NANOG	CATGAGTGTGGATCCAGCTTG	CCTGAATAAGCAGATCCATGG
PBGD	GGAGCCATGTCTGGTAACGG	CCACGCGAATCACTCTCATCT
SOX17	CGCACGGAATTTGAACAGTA	GGATCAGGGACCTGTCACAC
TTR	ACCGGTGAATCCAAGTGTCC	GGTTTTCCCAGAGGCAAATGG
PDX1	CCCATGGATGAAGTCTACC	GTCCTCCTCCTTTTTCCAC
PAX9	TGGTTATGTTGCTGGACATGGGTG	GGAAGCCGTGACAGAATGACTACCT
FOXP2	AATCTGCGACAGAGACAATAAGC	TCCACTTGTTTGCTGCTGTAAA
SOX9	CTCTGGAGACTTCTGAACGAGAG	CCTTGAAGATGGCGTTGGGG
CDX2	GGCAGCCAAGTAAAACCAG	TTCCTCTCCTTTGCTCTGCG
HNF1β	GCACCCCTATGAAGACCCAG	GGACTGTCTGGTTGAATTGTCG
NGN3	GCTCATCGCTCTCTATTCTTTTGC	GGTTGAGGCGTCATCCTTTCT
Villin1	CTGAGCGCCCAAGTCAAAG	AGCAGTCACCATCGAAGAAGC
CKIT	ATTCCCAGAGCCCACAATAG	ACCACTAGCTTTCCAAACGG
SFTPA1	TGTGTGTGGGTCGCTGATTT	GGCTTCAACACAAACGTCC
KRT17	ATCCTGCTGGATGTGAAGACGC	TCCACAATGGTACGCACCTGAC

CAV1	CCAAGGAGATCGACCTGGTCAA	GCCGTCAAAACTGTGTGTCCCT

2.3.3. Immunocytochemistry in 2D

For immunostaining, cells cultured in 2D were fixed in 4% PFA at 4°C for 20 min and washed twice with DPBS. Cells were blocked and permeabilised by incubation with 0.1% Triton X-100 and 10% donkey-serum in DPBS at RT for 30 min. Samples were incubated with primary antibody diluted in 0.01% Triton X-100 and 1% donkey-serum in PBS at 4° overnight. Samples were washed 3 times for 15 min in 1% donkey-serum solution. Samples were incubated in secondary antibody diluted in 1% donkey-serum in solution at RT for 1 h. The wash steps were repeated to remove excess secondary antibody. Samples were incubated in DAPI (1 µg/ml) at RT for 5 min and washed twice for 5 min in 1% donkey-serum. Samples were incubated with PBS to prevent the cells from drying out and were imaged using the Operetta High Content Imaging System (PerkinElmer) or Leica TCS SPE confocal microscope. A list of antibodies is provided.

2.3.4. Immunocytochemistry in 3D

For immunofluorescence imaging of 3D organoids, processing was completed using a published method (Dekkers et al., 2019). Briefly, organoids were removed from Matrigel by incubation with cell recovery solution (Corning, #354253) on a horizontal shaker at 4° for 60 min. Organoids were centrifuged in PBS at 70 x *g* for 3 min and fixed with 4% PFA. Organoids were resuspended in cold 0.1% PBS-Tween (Sigma, #P7949) and incubated at 4°C for 10 min. Organoids were blocked in Triton X-100, BSA and PBS (organoid was buffer; OWB). Staining was completed in low-adherence 24-well plates (ThermoFisher). Primary antibodies (2x concentration) were incubated at 4°C overnight on a horizontal shaker. Samples were washed three times in OWB and incubated with secondary antibody (2x concentration) at 4°C overnight on a horizontal shaker. Wash steps were repeated. Nuclei were sained with DAPI (1:500) for at RT for 20 min. Organoids were cleared by resuspension in fructose-glycerol solution (60% (vol/vol)) and incubated at RT for 20 min. Imaging spacers (SecureSeal™) were put onto glass slides and organoids were transferred to the spacers, and cover slips were place over them. Samples were imaged using a Zeiss LSM880 confocal laser scanning

microscope at the SLIM facility at the University of Nottingham by Tim Self and Dr Robert Markus. Images were analysed using ZEN 3.1 and ImageJ.

Table 2. 5. Antibodies for immunocytochemistry

Antibody	Dilution	Supplier / identifier
Primary antibodies		
Goat anti-human NANOG	1:100	R&D Systems, #AF1997
Mouse anti-human SOX2	1:50	R&D Systems, #MAB2018
Mouse anti-human OCT4	1:50	R&D Systems, #MAB17591
Mouse anti-human SOX17	1:50	R&D Systems, #MAB1924
Goat anti-human FOXA2	1:20	R&D Systems, #AF2400
Rabbit anti-human NKX2.1	1:250	Abcam, #ab76013
Secondary antibodies		
AlexaFluor™ 488 Donkey anti-goat (H+L) Cross-Adsorbed	1:400	ThermoFisher, #A11055
Alexa Fluor™ 488 Donkey anti-mouse (H+L) Cross-Adsorbed	1:400	ThermoFisher, #A21202
AlexaFluor™ 488 Donkey anti-rabbit (H+L) Cross-Adsorbed	1:400	ThermoFisher, #A21206
AlexaFluor™ 647 Donkey anti-mouse (H+L) Cross-Adsorbed	1:400	ThermoFisher, #A31571
AlexaFluor™ 647 Donkey anti-goat (H+L) Cross-Adsorbed	1:400	ThermoFisher, # A21447
DAPI	1:500	Sigma, #D9542

2.3.5. RNA sequencing

RNA was extracted from differentiation samples as previously described (section **2.3.1**) and stored at -80°C. Bulk RNA sequencing was outsourced to GeneWiz NGS Laboratory and performed using the Ilumina HISeq platform. 20 million reads were done as base paired ends. Samples were mapped to reference genome GRCh38 p7. Bioinformatic analysis was completed using RStudio. The limma package from Bioconductor was used to statistically analyse bulk mRNA expression dataset (Ritchie et al., 2015). Data was normalised to $[\log]_2$ (1+Dataset). Gene hit count dataset was tested for differential gene expression using limma to fit linear models. Empirical Bayes moderation of the standards errors towards a common value were done for each gene between samples for differential expression analysis, using a threshold p-value of 0.05. Benjamini-Hochberg procedure (False Discovery Rate) was used for multiple hypothesis test

correction. PCA plots, volcano plots, bar graphs, and heatmaps were generated in R. All analysis was performed in consultation with Carlos Sainz Zuniga.

2.3.6. Statistical analysis

All data in this work are displayed as the mean with the standard error of the mean (mean ± SEM) unless stated otherwise. For qRT-PCR analysis, comparison between groups was assessed using unpaired two-tailed Student's t test or one-way analysis of variance (ANOVA) with either Tukey's multiple comparison or Dunnett's multiple comparison post hoc test as stated in the figure legend. *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001 was used to determine statistically significant differences between groups. All statistical analysis was performed in GraphPad Prism version 9 software.

3. Development of a protocol for differentiation of patient-derived hiPSCs to type 2 AECs of the lung

3.1. Introduction

The ability to generate patient-specific type 2 AECs from hiPSCs harbouring pathogenic mutations is an attractive prospect for modelling disease of the lung *in vitro*. Combining this approach with genome-editing of hiPSCs can permit the correction of genetic variations to generate isogenic control cell lines that have the same genomic background, providing a powerful tool to study causal gene mutations. The discovery and development of gene-editing tools such as transcription activator-like effector nucleases (TALENs) and the Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas9 system of prokaryotes has made this possible (Jinek et al., 2012; Joung & Sander, 2013). The CRISPR-Cas9 system has been engineered to mediate double-stranded DNA breaks (DSBs) by the Cas9 endonuclease directed to almost any target of interest by small RNA sequences known as guide RNAs (gRNAs). This initiates DNA repair processes such as homology directed repair (HDR) and providing donor templates with homologous DNA sequences can enable precise genomic modifications such as single base substitutions, insertions or deletions (Adli, 2018). Although CRISPR/Cas9 is most commonly used for genomic-editing, DSBs initiate other repair mechanisms like non-homologous end joining (NHEJ) which can limit efficiency of HDR and can also introduce undesired insertions and deletions of nucleotides in the DNA sequence (Bennett et al., 2020).

Recent development of base editing technology for genomic editing has circumvented some of the problems associated with CRISPR/Cas9. Base editors consist of a CRISPR-Cas nuclease with catalytically inactivated endonuclease activity and a DNA deaminase enzyme to facilitate single nucleotide variations without introducing DSBs or the need for donor templates. The conversion of C•G to T•A base pairs can be mediated by cytosine base editors (CBEs), and adenine base editors (ABEs) can mediate A•T to G•C changes (Anzalone et al., 2020).

These new genetic tools are highly useful for interrogating diseases such as IPF which is associated with a number of single nucleotide polymorphisms (SNPs) (Allen et al., 2017). Rare genetic variations in type 2 AEC genes like *SFTPC* have been identified in

familial IPF cases (Nogee et al., 2001). The mature SFTPC peptide is first synthesised as a larger, 197 amino acid-long precursor integral membrane protein. It has four distinct domains: a cytoplasmic N-terminal domain important for trafficking, a transmembrane region where the mature SFTPC peptide resides, a highly conserved BRICHOS domain which facilitates insertion of SFTPC into the membrane and acts as a chaperone to prevent formation of amyloid fibrils, and the C-terminal linker region which anchors the BRICHOS domain to the mature SFTPC transmembrane domain (Mulugeta et al., 2015). An SNP in the linker domain of *SFTPC*, g.1286 T > C, leads to the replacement of threonine with isoleucine at codon 73 (I73T) and is the most common *SFTPC* mutation that has been identified (Salerno et al., 2016). Several substitution mutations in the BRICHOS domain have also been identified, such as c.298 G > A (G100S) (Mulugeta et al., 2005; Ono et al., 2011). To modify genetic variants such as those described, base editing is highly applicable. It is therefore essential to develop robust platforms to differentiate hiPSCs derived from patients harbouring these mutations to relevant cell types such as type 2 AECs to model disease.

Directed differentiation protocols attempt to recapitulate the normal signalling events during embryonic development to drive the specification of cells of interest. To drive hPSC differentiation to DE *in vitro*, high concentrations of activin A are required to mimic high embryonic Nodal signalling but this does not fully suppress pluripotency genes (D'Amour et al., 2005). This requires the synergistic action of BMP, FGF, and WNT signalling for efficient DE specification (Teo et al., 2014; Vallier et al., 2009). This effect is significantly enhanced with WNT stimulation through GSK3β inhibition via CHIR99021 (Teo et al., 2014). Activation of the PI3K signalling pathway by insulin has been shown to be refractory to DE development and has led to the removal or replacement of serum which contains high concentrations of insulin or the need for PI3K inhibitors such as LY294002 to improve differentiation (D'Amour et al., 2005; McLean et al., 2007; Yu et al., 2015). A well characterised set of markers have been identified that are routinely used to validate and quantify differentiation to DE including the transcription factors *SOX17*, *FOXA2*, *GATA4* and *GATA6*, and the cell surface receptors CD184 (CXCR4) and CD117 (CKIT) (Cheng et al., 2012; D'Amour et al., 2005; Green et al., 2011; McLean et al., 2007).

Following differentiation to endoderm, inhibition of the signalling transduction SMAD proteins of the TGFβ superfamily via inhibitors of TGFβ and BMP signalling, such as

SB431542 and dorsomorphin (referred to as dual-SMAD inhibition), respectively, can induce an anterior foregut identity in DE cells (Green et al., 2011). This permits cells to further differentiate to the lung through combinations of WNT, FGF, BMP, SHH and RA signalling molecules to promote emergence of NKX2.1+ lung epithelial progenitor cells *in vitro* (Dye et al., 2015; Gotoh et al., 2014; Green et al., 2011; Huang et al., 2015; Mou et al., 2012). Protocols have been established for the differentiation of human pluripotent stem cells to NKX2.1+ lung-competent cells in 2D monolayer culture, but rely on reporter cell lines or feeder layers for growth (Firth et al., 2014; Gotoh et al., 2014; Hawkins et al., 2017; Huang et al., 2014; Huang et al., 2015; Jacob et al., 2017; Jacob et al., 2019; Konishi et al., 2016; Longmire et al., 2012; Rodrigues Toste de Carvalho et al., 2021; Serra et al., 2017; Yamamoto et al., 2017). As such, there have been attempts to develop feeder-free protocols and identify cell surface markers to enrich for NKX2.1+ lung-competent cells. Gotoh et al. (2014) found *carboxypeptidase M* (*CPM*), a membrane-bound peptidase, was significantly upregulated with *NKX2.1* after 'ventralisation' of hPSC derived-AFE, and more than 90% of sorted CPM cells were NKX2.1+ (Gotoh et al., 2014; Yamamoto et al., 2017). This has been corroborated in a further study which further identified the upregulation of *CD47* with *NKX2.1*, and cell sorting based on CD47+/CD26- could enrich for NKX2.1+ cells (~90%) (Hawkins et al., 2017).

Lung epithelial progenitor cells have the capacity to differentiate into both airway and alveolar fates *in vitro* (Chen et al., 2017; Hein et al., 2022; Miller et al., 2019; Rodrigues Toste de Carvalho et al., 2021). The maturation of NKX2.1+ foregut spheroids can generate organoids containing cells of both airway and alveolar identity, including a fraction of SFTPC+ type 2 AECs (Chen et al., 2017; Huang et al., 2014; Miller et al., 2019; Rodrigues Toste de Carvalho et al., 2021). Alternatively, lung progenitor cells isolated via NKX2.1-fluorescent reporter lines or cell surface proteins like CPM and embedded in 3D Matrigel™ matrices can give rise to alveolar epithelial-only organoids containing a more homogeneous population of SFTPC+ type 2 AECs (Gotoh et al., 2014; Hawkins et al., 2017; Jacob et al., 2017; Yamamoto et al., 2017). 3D 'alveolospheres' maintaining a self-renewing population of SFTPC+ type 2 AECs have now been established by repeated purification of SFTPC+ cells via fluorescent reporters and continued culture (Jacob et al., 2019). The gene SLC34A2, encoding a multi-transmembrane sodium dependent phosphate transporter (NaPi2B), was recently

identified in adult and fetal SFTPC+ cells, and it was found to enrich for hPSC derived SFTPC+ cells (>80%) (Korogi et al., 2019). A commercially available monoclonal antibody targeting the extracellular domain of NaPi2B has been developed, offering a useful method to isolate and purify type 2 AECs without reporter cell lines (Korogi et al., 2019; Yin et al., 2008).

Our aim was to establish a directed differentiation protocol to generate type 2 AECs from patient-derived hiPSC lines with a mutation in the BRICHOS domain of the *SFTPC* gene. We sought to develop a strategy to generate and purify lung epithelial progenitor cells in 2D culture and type 2 AECs in 3D organoids using cell surface markers to overcome the need for reporter cell lines, providing a reproducible protocol to derive an unlimited source of alveolar cells for future disease modelling applications.

3.2. Results

3.2.1. Overview of the strategy to generate type 2 AECs

To develop a platform for the generation of type 2 alveolar epithelial cells (AECs) for disease modelling and therapeutic applications, a hiPSC line was established from a patient diagnosed with familial IPF harbouring a heterozygous mutation in the *SFTPC* gene. The approach was to genetically modify this hiPSC line to generate isogenic homozygous-corrected and homozygous-mutant cell lines to provide a continuum of genetic disease. Establishing a protocol to differentiate hiPSCs to type 2 AECs would permit investigation of *SFTPC* mutations on type 2 AEC phenotype and function while controlling for heterogeneous genetic background (**Fig. 3-1**). All gene-editing was carried out by Dr Sara Cuevas Ocaña at the University of Nottingham.

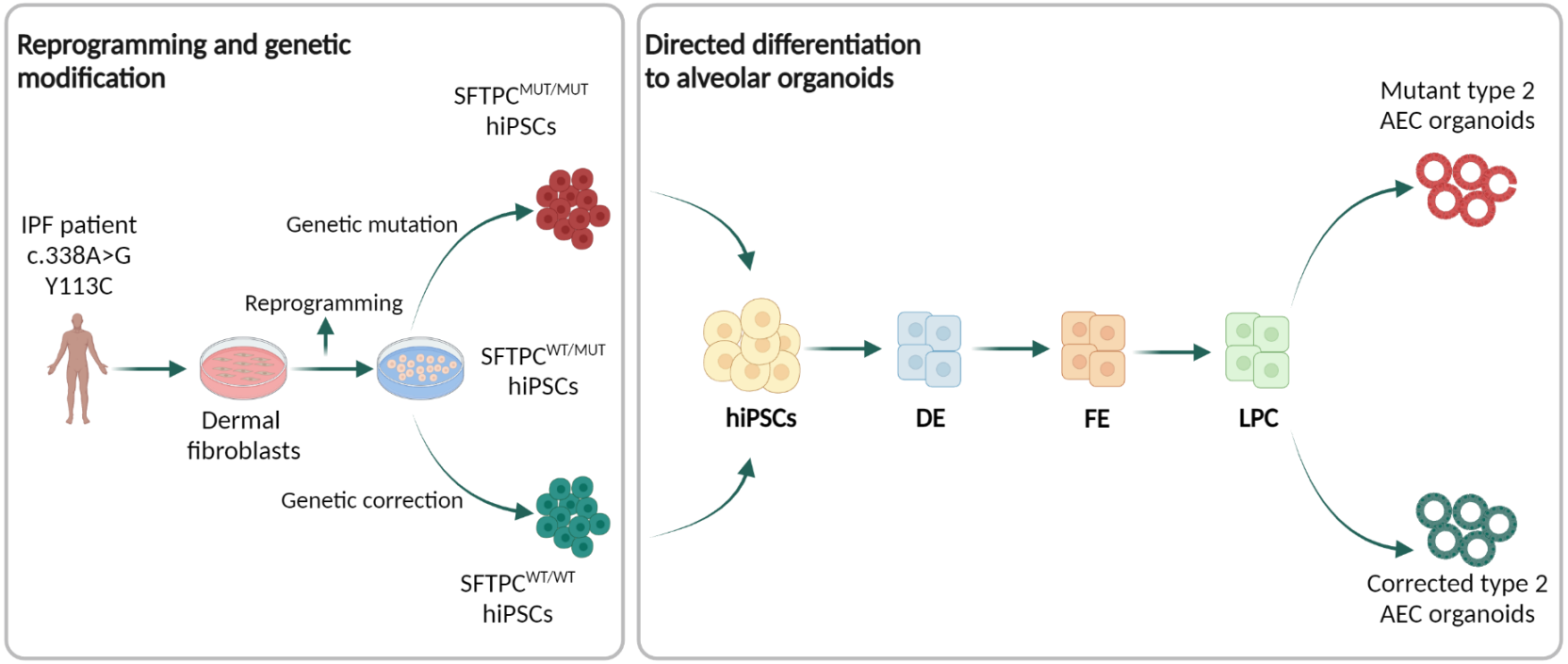


Figure 3-1. Overview of the project strategy to generate type 2 AECs for disease modelling and therapeutic purposes.

3.2.2. Gene-editing patient-derived hiPSCs

hiPSC lines were established via non-integrating Sendai virus reprogramming of patient-derived dermal fibroblasts that harboured a heterozygous point mutation c338.A>G causing a base pair change (A>G) in exon 4 of the highly conserved BRICHOS domain of *SFTPC*. This resulted in the substitution of tyrosine by cysteine at codon 113 (p.Y113C). The patient was diagnosed with the familial form of IPF. To generate isogenic cell lines, gene editing was performed utilising newly developed base editing technology which can introduce single nucleotide variations without inducing double stranded DNA breaks unlike CRISPR-Cas9 (Gaudelli et al., 2017; Komor et al., 2016). In this approach two base editing systems were used; cytosine base editors (CBEs) to mediate C•G-to-T•A base pair conversion to correct the mutation (G>A), and adenine base editors (ABEs) to facilitate A•T-to-G•C base pair conversion to introduce the mutation (A>G). Firstly, transfection of the patient-derived parental hiPSC line, referred to as IPF1.12, was optimised by using a plasmid vector expressing enhanced GFP (eGFP) (**Fig. 3-2a**). hiPSCs were nucleofected with 419 ng of eGFP plasmid in either commercially available P3 or Ingenio® transfection buffer and 7 Nucleofector programs were tested. Expression of GFP was assessed by flow cytometry 48 h after transfection. The P3 buffer was superior to Ingenio® reagent and combined with the DN-100 program led to the highest transfection efficiency (**Fig. 3-2b**). Using this optimised condition, the mean ± SEM transfection efficiency was 77.4 ± 7.6% (N=26) (**Fig. 3-2c**).

Following optimisation of transfection, hiPSCs were transfected with plasmid vectors for base editing. This consisted of pCMV_AncBE4max_P2A_GFP (**Fig. 3-3a**) for correction of the mutation (G>A), pCMV_ABEmax_P2A_GFP (**Fig. 3-3b**) for introduction of the mutation (A>G), and a plasmid vector for the gRNA containing the specific target sequence (**Fig. 3-3c**). These plasmids expressed GFP, which was used to assess transfection efficiency, and a puromycin resistance cassette for enrichment. Following transfection, hiPSCs were seeded into a 24-well plate and underwent a transient puromycin selection after 24h. Puromycin-resistant colonies were dissociated to single cells and seeded at a density to isolate single cell clones. At this

stage, the restriction digestion HpyCH4V enzyme assay was performed to determine the editing efficiency at the population level (**Fig. 3-4a, b**).

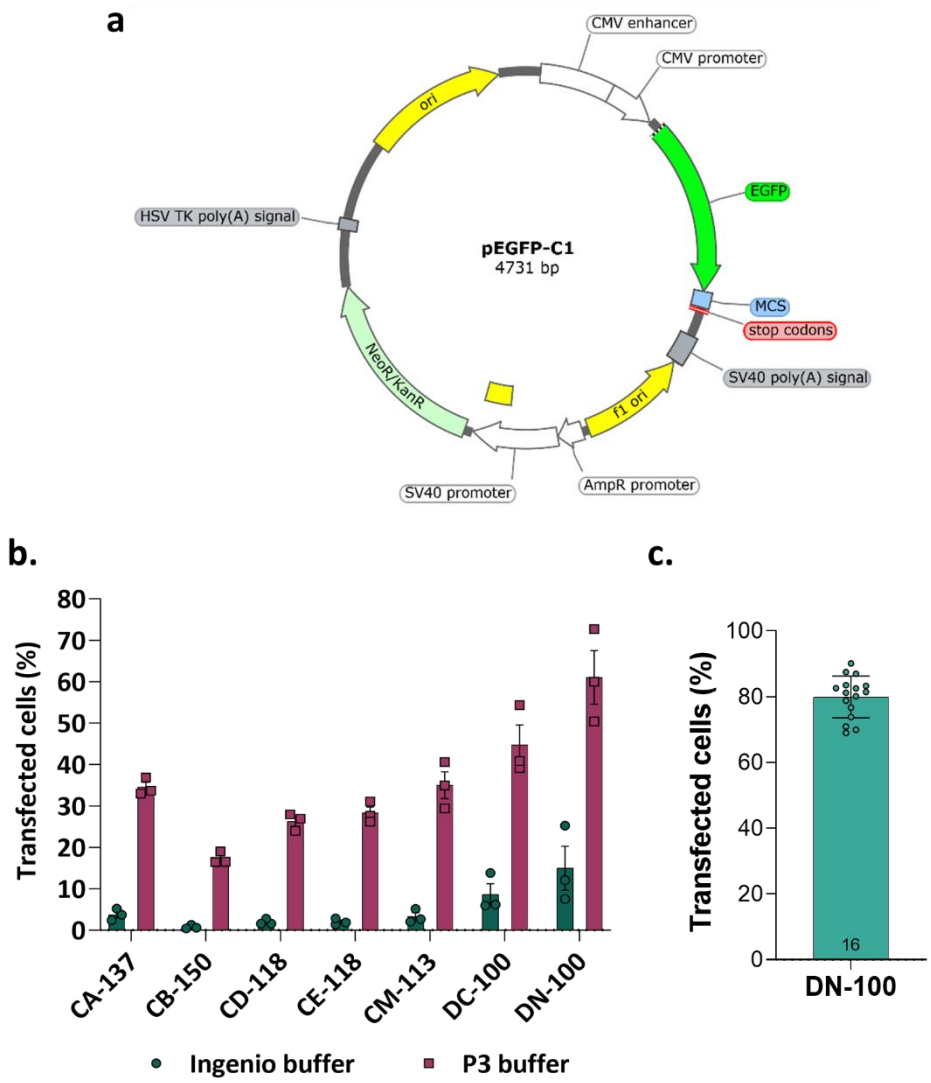


Figure 3-2. Optimisation of nucleofection efficiency using an eGFP plasmid vector.

(a) Plasmid map of eGFP-C1 used for assessing transfection efficiency. **(b)** Quantification of GFP+ cells via flow cytometry after transfection with P3 buffer or Ingenio electroporation buffer using different Nucleofector programs (N=3). **(c)** Quantification of GFP+ cells via flow cytometry after transfection in P3 buffer using the DN-100 Nucleofector program (N=16).

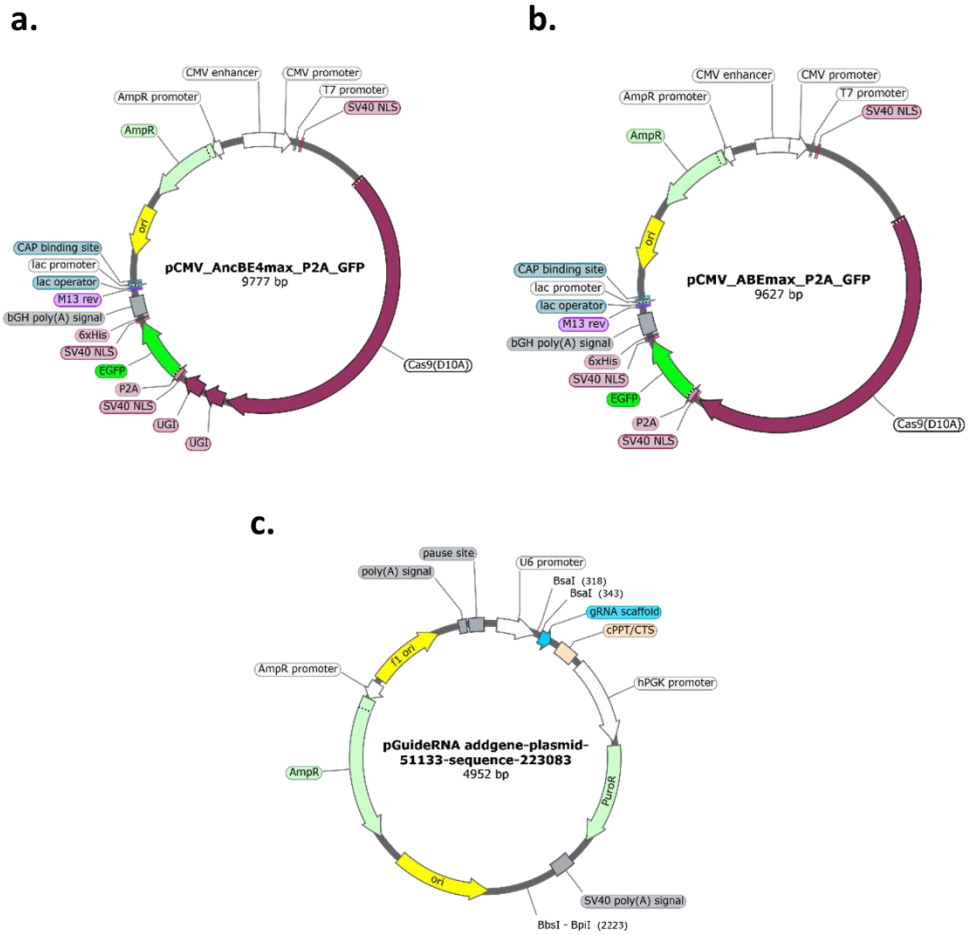


Figure 3-3. Plasmid maps for base editing.

Plasmid maps showing the features of the vectors used for **(a)** correction of the mutation (G>A) pCMV_AncBE4max_P2A_GFP and **(b)** introduction of the mutation (A>G) pCMV_ABEmax_P2A_GFP and of the **(c)** guide RNA pGuideRNA addgene-plasmid-51133-sequence-223083.

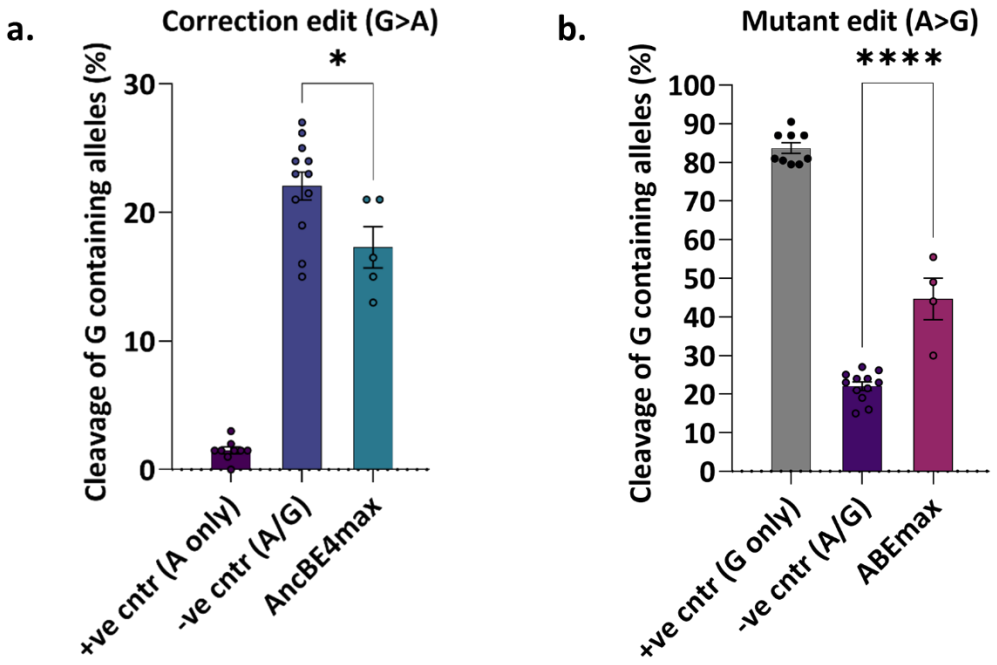


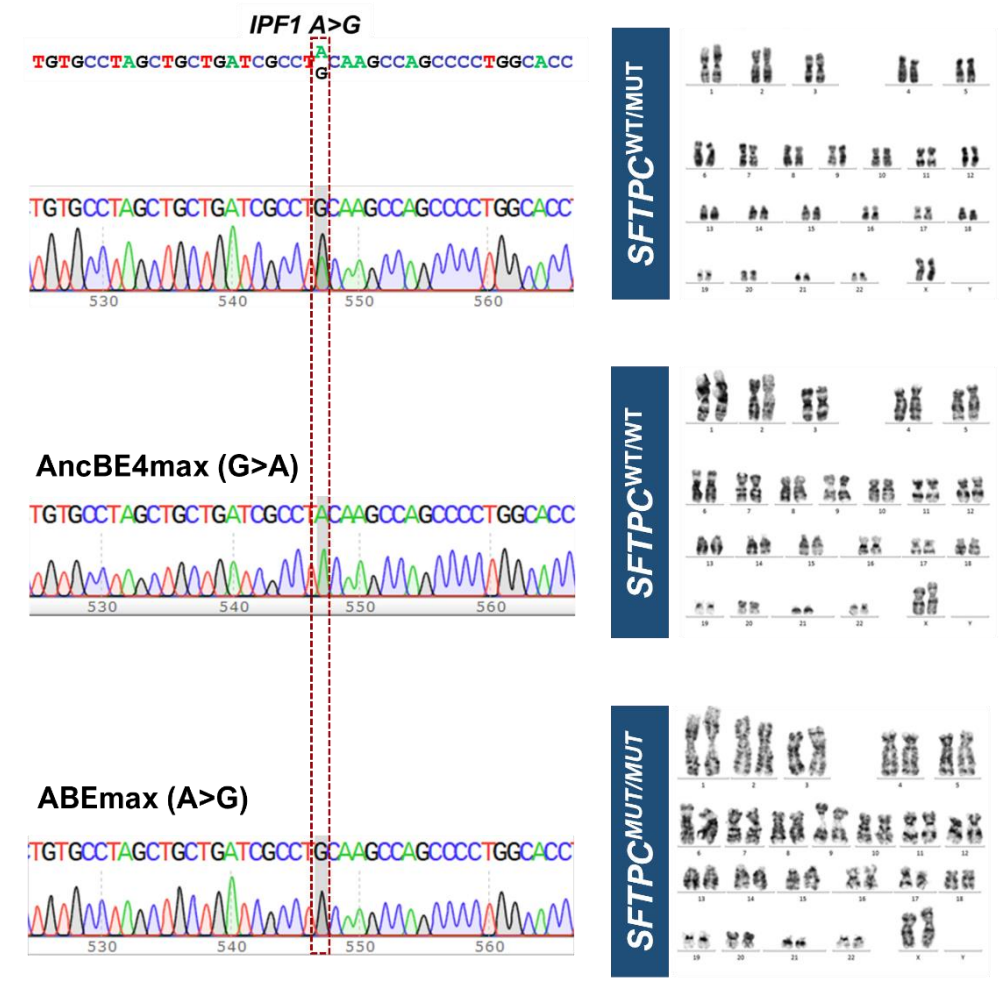
Figure 3-4. Base edits can be detected at the population level using the restriction digestion HpyCH4V assay.

Quantification of the HpyCH4V assay to detect the number of clones correctly edited to **(a)** correct the mutation (G>A) and **(b)** introduce the mutation (A>G) into IPF1.12 hiPSCs. Bars represent mean ± SEM. *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001, one-way ANOVA with Dunnet's multiple comparisons test.

Following clonal expansion, a small fragment of the puromycin-resistant colonies was mechanically isolated for direct gDNA extraction and genotyping by PCR and the restriction digestion HpyCH4V assay at the clonal level. The remainder colony was transferred to a 48-well plate for expansion. From the clones identified as correctly modified by PCR screening, sanger sequencing was performed for confirmation of the modified sequence. This confirmed the AncBE4max system successfully induced G>A base pair conversion and the ABEmax system induced A>G base pair conversion at the site of the mutation (**Fig. 3-5a**). Karyotype analysis confirmed there were no chromosomal abnormalities in the IPF1.12 parental hiPSC line or following genetic modification with base editors (**Fig. 3-5b**). The cell lines used in this study were defined as follows: parental IPF1.12 (*SFTPC*$^{WT/MUT}$), homozygous-corrected BE.31 and BE.32 (SFTPC$^{WT/WT}$) and homozygous mutant ABE.27 (SFTPC$^{MUT/MUT}$) (**Fig. 3-5c**).

a.

b.



c.

IPF.12	BE.31 / BE.32	ABE.27
SFTPC$^{WT/MUT}$	SFTPC$^{WT/WT}$	SFTPC$^{MUT/MUT}$
A>G	G>A	A>G

Figure 3-5. Karyotype analysis and validation of base-editing of IPF1.12 hiPSCs.

(a) DNA sequencing chromatograms showing the single nucleotide variation in the IPF1.12 parental hiPSC line (A>G) and confirmation of the correct modification by AncBE4max (G>A) (*SFTPC*$^{WT/WT}$) and ABEmax (A>G) (*SFTPC*$^{MUT/MUT}$). **(b)** Chromosome analysis showing normal karyotypes of the IPF1.12 parental cell line and the wildtype and mutant gene-edited cell lines. **(c)** Table summarising the cell line names and the corresponding genetic modification.

3.2.3. Maintenance and characterisation of hiPSC lines

All cell lines were maintained in HB E8 medium and displayed normal morphology with formation of compacted colonies of cells with a high nuclear-to-cytoplasmic ratio and were routinely split as single cells every 3 to 4 days (**Fig. 3-6a**). BE.31 ($SFTPC^{WT/WT}$) cells displayed protein expression of the pluripotency markers OCT4, SOX2 and NANOG as determined by immunocytochemistry (**Fig. 3-7a**), and the BE.31 ($SFTPC^{WT/WT}$) and ABE.27 ($SFTPC^{MUT/MUT}$) hiPSCs showed 100% expression of the pluripotency SSEA4 (**Fig 3-7b**). The parental and modified cell lines also showed >95% SOX2 expression >98% of cells were positive for OCT4 as assessed by intracellular flow cytometry (**Fig. 3-7c**). Expression of NANOG was more variable across cell lines, with co-expression of NANOG and OCT4 in nearly 60% and 70% of BE.32 ($SFTPC^{WT/WT}$) and IPF1.12 ($SFTPC^{WT/MUT}$) hiPSCs, respectively, but only 22% in ABE.27 ($SFTPC^{MUT/MUT}$) hiPSCs (**Fig. 7d**). NANOG is essential for maintenance of pluripotency but has been widely reported to display heterogeneous expression patterns in embryonic and induced PSCs that can impact capacity for differentiation (Torres-Padilla & Chambers, 2014).

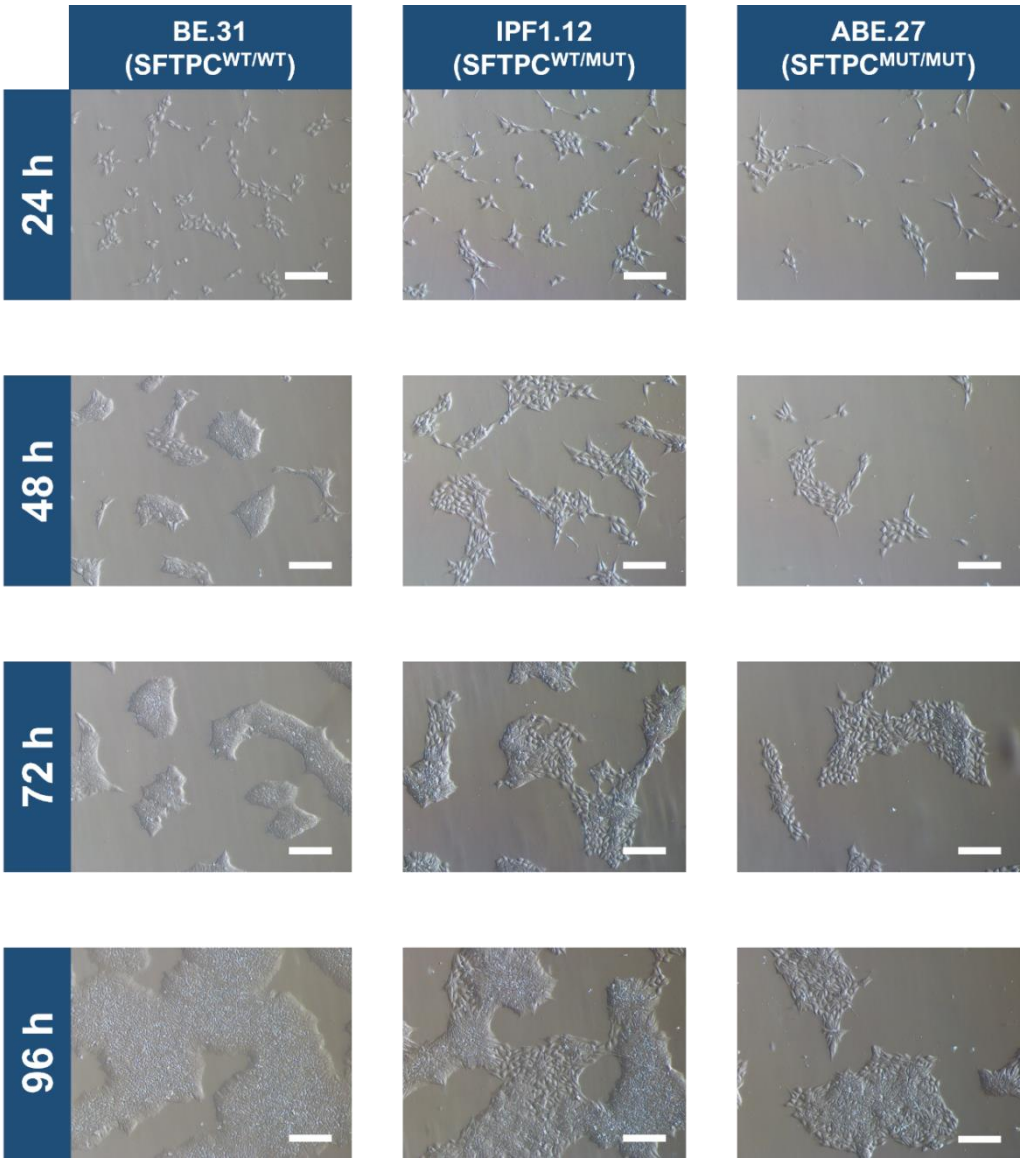


Figure 3-6. Morphology of human induced pluripotent stem cells.

Representative light microscope images of BE.31 ($SFTPC^{WT/WT}$) (left), IPF1.12 ($SFTPC^{WT/MUT}$) (middle) and ABE.27 ($SFTPC^{MUT/MUT}$) (right) hiPSCs maintained in Essential 8 medium over 4 days. Scale bars, 200 µm.

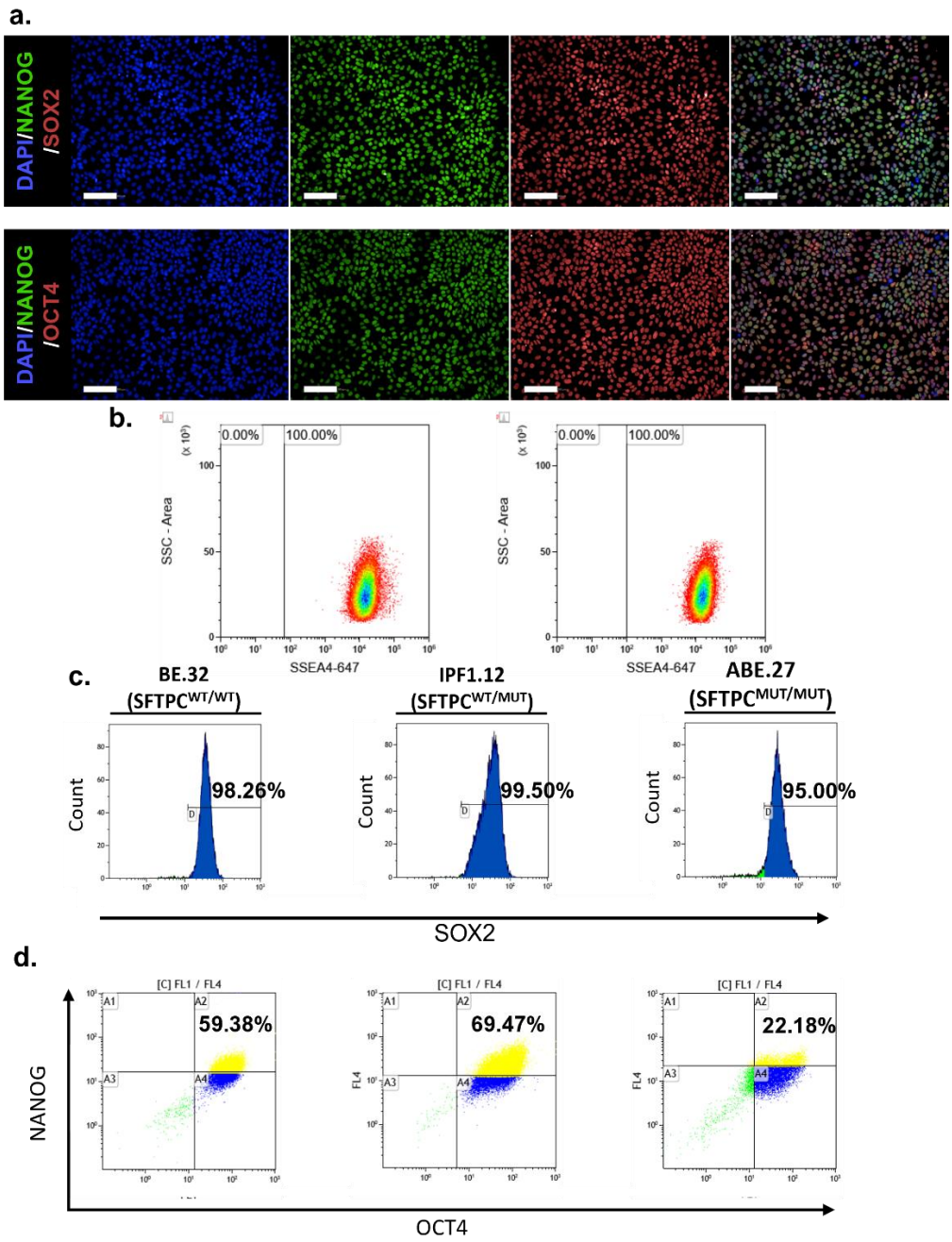


Figure 3-7. Expression of pluripotency markers in hiPSCs.

(a) Representative immunofluorescent images of the pluripotency markers SOX2, NANOG and OCT4 in BE.31 ($SFTPC^{WT/WT}$) hiPSCs maintained in HB E8 medium (Nuclei stained with DAPI). Scale bar, 100 µm. **(b)** Flow cytometry analysis of SSEA4 protein expression in BE.31 ($SFTPC^{WT/WT}$) (left) and ABE.27 ($SFTPC^{MUT/MUT}$) (right) hiPSCs maintained in HB E8 medium (N=1). **(c, d)** Intracellular flow cytometry analysis of pluripotency markers for parental and genetically modified hiPSCs (N=1).

3.2.4. Differentiation of REBL-PAT control hiPSCs to lung progenitor cells using a previously established protocol

To determine the potential of hiPSCs to differentiate into type 2 alveolar epithelial cells (AECs), we used a previously established stepwise differentiation protocol that mimics distinct stages of embryonic lung development (**Fig. 3-8a**). This begins with specification of definitive endoderm via activation of wnt and activin signalling, and patterning of the gut tube to a foregut fate by continued activin signalling. Differentiation to NKX2.1-expressing lung epithelial progenitor cells was driven by activation of Wnt signalling via CHIR99021, and BMP and retinoic acid signalling (**Fig. 3-8a**). As genetically modified hiPSCs were still being generated and validated, we first used an established in-house healthy hiPSC line with no known respiratory pathology, REBL-PAT, as a benchmark for the differentiation protocol (Mosqueira et al., 2018). REBL-PAT cells readily expanded 24 h after endoderm induction to form a 'cobblestone'-like morphology and a compact monolayer was maintained during the entire protocol (**Fig. 3-8b**). Gene expression analysis at major stages of the lung differentiation showed the pluripotency gene *NANOG* was highly expressed in hiPSCs and downregulated upon differentiation. Definitive endoderm genes *SOX17, FOXA2* and *GATA4* were upregulated on endoderm day 3, but for *SOX17* there was large variation between biological repeats. As *SOX17* expression gradually decreased, *FOXA2* and *GATA4* expression peaked on day 4 of lung specification but was maintained throughout lung progenitor cell differentiation (**Fig. 3-9a**). Expression of *NKX2.1*, a master regulator of normal lung development, was expressed highest during foregut patterning, but its expression was diminished during lung specification when it should be upregulated. Although *SFTPC* expression was shown to peak on endoderm day 3 and gradually increased from lung day 4 to 12, all CT values were >35 cycles of amplification and the majority of samples were undetermined (>40 cycles), apart from one technical triplicate CT value of 32 on DE D3 (**Fig. 3-9a**). This suggests there was no *SFTPC* expression and non-specific PCR product amplification was responsible for the false positive result observed, as *SFTPC* is not expressed at early stages of germ layer specification during development (Wambach et al., 2010).

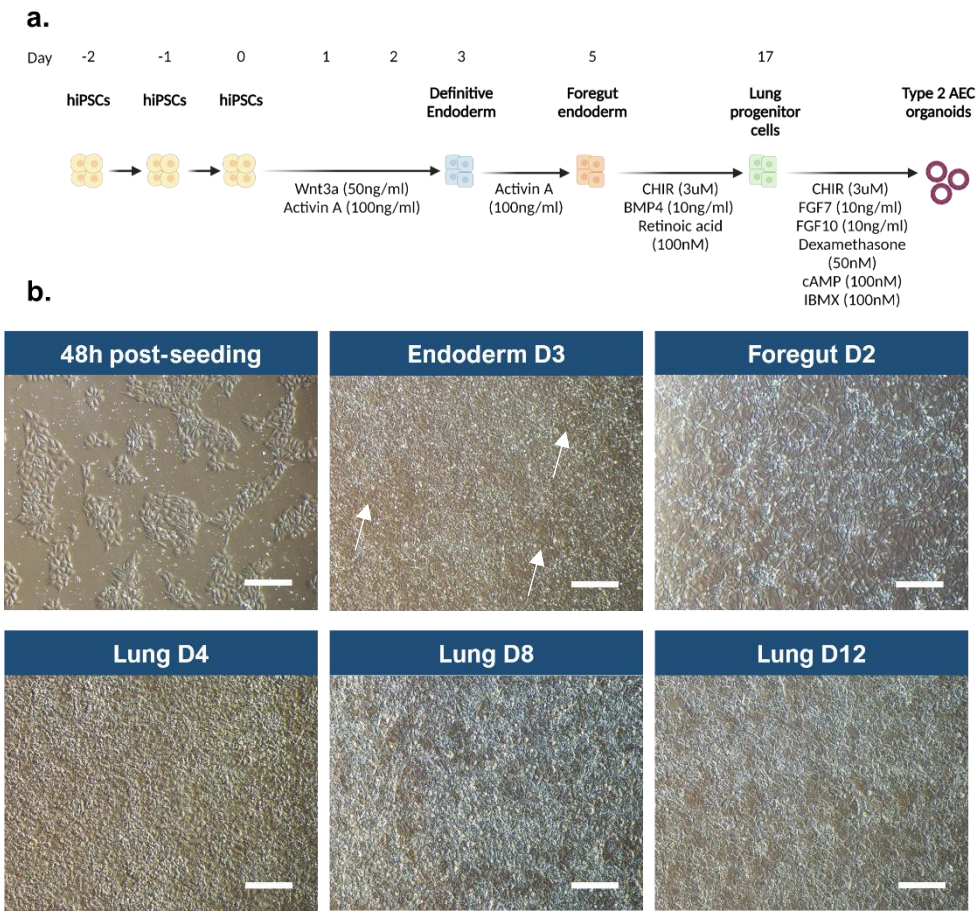


Figure 3-8. Differentiation of REBL-PAT hiPSCs to alveolar epithelial type 2 cells.

(a) Schematic of the stepwise, directed differentiation of hiPSCs to type 2 AECs with stage-specific growth factor combinations. **(b)** Representative light microscope images of cells at major stages of differentiation. Arrows indicate areas of DE cells with a 'cobblestone' morphology. Scale bars, 200 µm.

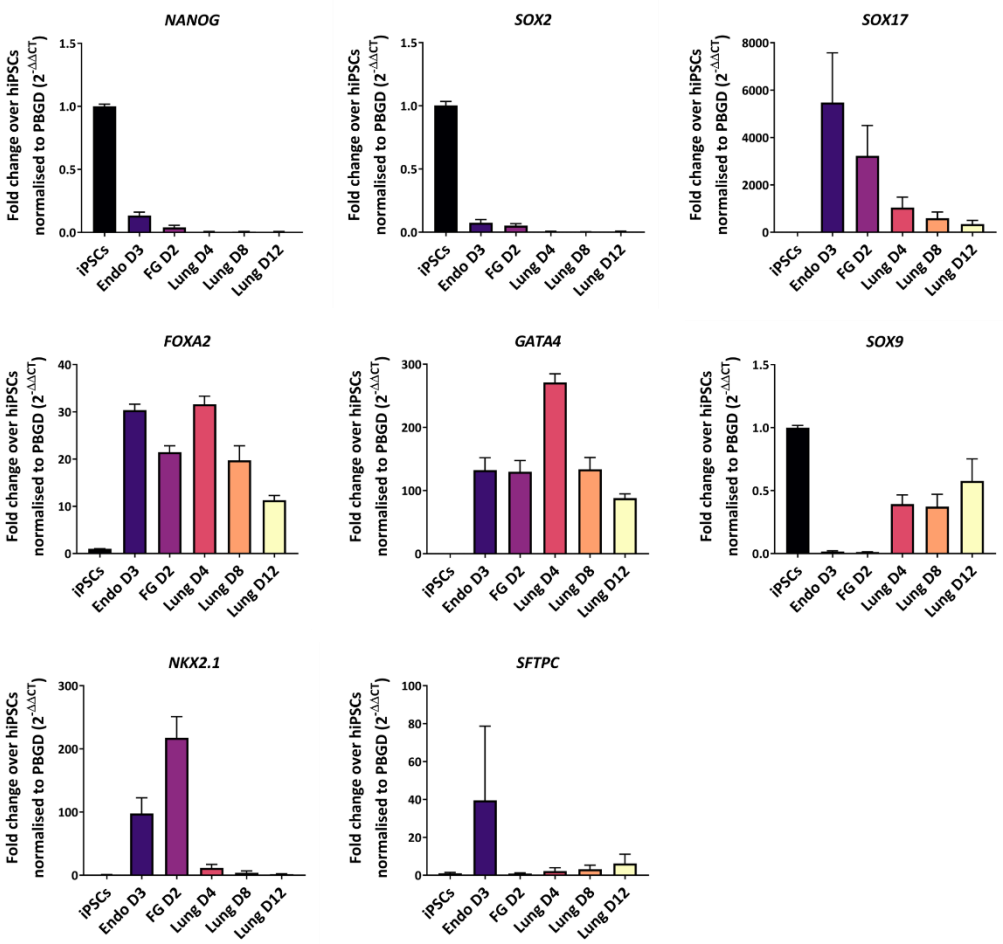


Figure 3-9. Expression of stage-specific genes during lung differentiation of REBL-PAT hiPSCs

Gene expression analysis via RT-qPCR of distinct stages of differentiation to lung epithelial progenitor cells with bars representing mean normalised expression to the housekeeping gene PBGD relative to hiPSCs ($2^{-\Delta\Delta Ct}$) ± SEM. (N=2, 3 technical replicates per biological replicate).

3.2.5. Differentiation of IPF1.12 hiPSCs to lung organoids using a previously established protocol

We applied this protocol to our patient derived IPF1.12 ($SFTPC^{WT/MUT}$) hiPSCs to assess the capacity to differentiate to lung progenitor cells. In comparison to REBL-PAT hiPSCs, on day 3 of endoderm differentiation, large undifferentiated cell colonies had persisted and were still visible during foregut specification. These gradually developed into large 3D aggregates during lung specification (**Fig. 3-10a**). By analysing the gene expression profiles during differentiation, we showed the pluripotency marker *SOX2* was expressed in hiPSCs but was upregulated in DE cells. *SOX2* expression should re-emerge during foregut specification, as we showed, but it should be downregulated at the DE stage, indicating the persistence of hiPSC colonies and incorrect foregut patterning. Although DE markers (*SOX17*, *FOXA2*, *GATA4*) were upregulated on day 3, their expression was maintained or increased during lung differentiation (**Fig. 3-10b**). Like the REBL-PAT hiPSCs, *NKX2.1* expression peaked on foregut day 2, but decreased during specification of lung progenitors and compared to total lung RNA, which was used as a benchmark, was much lower. Similarly, *SFTPC* expression was not upregulated across lung differentiation, and was drastically lower than in total lung (**Fig. 3-10b**), suggesting cells were not responding correctly to induction signals. At this stage, we also assessed the DE induction efficiency by intracellular flow cytometry. We showed that only 55% of IPF1.12 ($SFTPC^{WT/MUT}$) hiPSCs, and <20% of BE.32 ($SFTPC^{WT/WT}$) and ABE.27 ($SFTPC^{MUT/MUT}$) hiPSCs expressed SOX17 after 3 days of endoderm differentiation (**Fig. 3-11a**). The collection and analysis of this data highlighted a problem with our current differentiation protocol showing failure to efficiently induce DE cells and lack of further differentiation to lung progenitor and type 2 AECs in IPF1.12 ($SFTPC^{WT/MUT}$) hiPSCs.

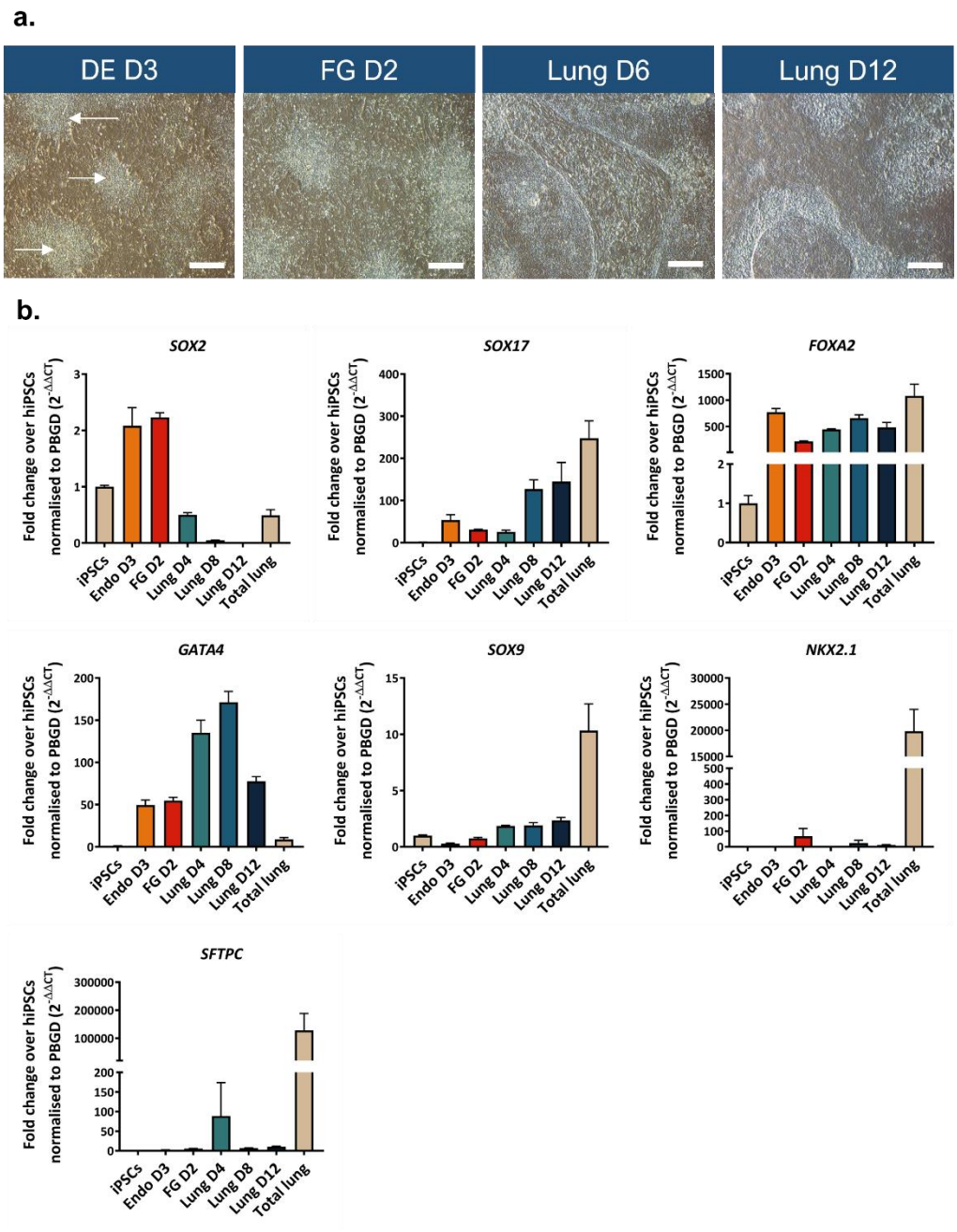


Figure 3-10. Gene expression during lung differentiation of IPF1.12 (SFTPC$^{WT/MUT}$) hiPSCs.

(a) Representative light microscope images of cells at major stages of lung differentiation. Scale bars, 200 µm. Arrows indicate undifferentiated colonies. **(b)** Expression of endoderm and lung-related genes during differentiation to lung progenitor cells, including total adult lung, with bars representing mean normalised expression to the housekeeping gene PBGD relative to hiPSCs ($2^{-\Delta\Delta Ct}$) ± SEM (N=1, 3 technical replicates).

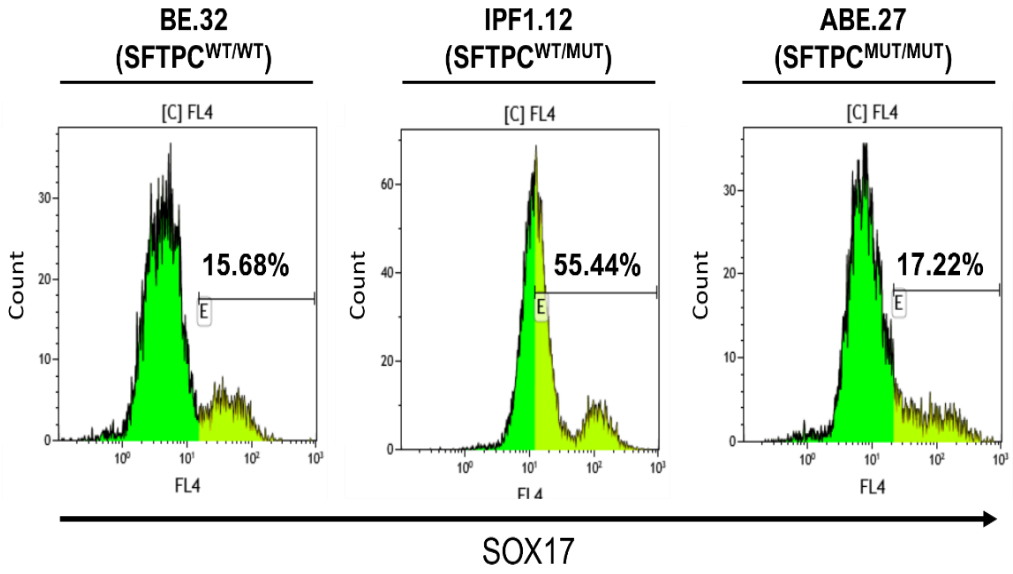


Figure 3-11. Expression of SOX17 varies across hiPSC lines.

Intracellular flow cytometry analysis of SOX17 in IPF1.12 ($SFTPC^{WT/MUT}$) parental and homozygous corrected BE.32 ($SFTPC^{WT/WT}$) and homozygous mutant ABE.27 ($SFTPC^{MUT/MUT}$) hiPSCs following 3 days of DE differentiation with WNT3a (50 ng/ml) and Activin A (100 ng/ml) (N=1).

3.2.6. Generation of three-dimensional organoids

Despite sub-optimal expression of stage-specific genes during lung differentiation, we wanted to determine whether there was any population of progenitor cells present at low frequency in the culture that we could still drive to type 2 AECs expressing SFTPC using specific inductive signals for differentiation. On day 12 of lung progenitor cell specification, we collected cells as clumps and embedded them in three-dimensional (3D) Matrigel™ droplets in the presence of WNT and FGF signalling, and the well-established lung cell maturation factors dexamethasone, cAMP and IBMX (**Fig. 3-8a**) (Gonzales et al., 2002). Cell clumps rapidly self-assembled to form 3D cystic spheres with a thin epithelium and a large lumen (**Fig. 3-12a**). These structures could be disaggregated and reaggregated every 5-7 days and could be maintained for over 7 passages with continued expansion of organoid size (**Fig. 3-12a**). We assessed the expression of lung-specific markers over 4 passages and demonstrated that *NKX2.1* and *SFTPC* were not expressed in our organoids but were upregulated in the total lung positive control sample. Expression of *SOX9*, a marker of distal bud tip progenitor cells in the developing lung, was upregulated across sequential passages compared to total lung expression (**Fig. 3-12b**). This might indicate these organoids express an immature phenotype, but further validation would be necessary. Nonetheless, despite the lack of type 2 AEC gene expression, we wanted to determine whether organoids remained viable after cryopreservation so they could be stored and readily accessed when needed. Following 24 h from thawing, clumps of cells were established in Matrigel™ droplets, and by day 6 they had rapidly expanded and appeared healthy (**Fig. 3-13a**).

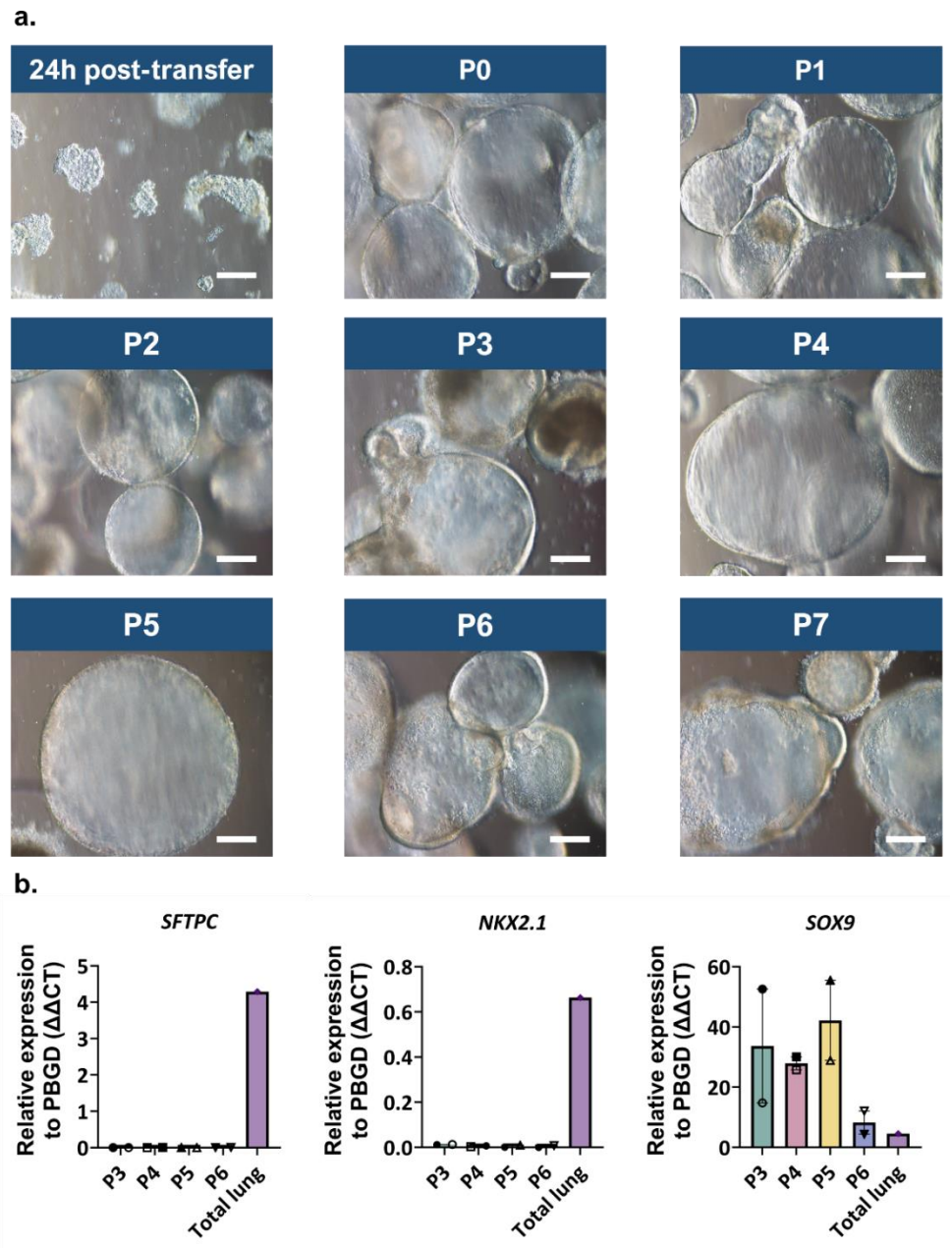


Figure 3-12. Generation of three-dimensional lung organoids.

(a) Representative light microscope images of IPF1.12 (SFTPC$^{WT/MUT}$) 3D Matrigel-embedded lung organoids at successive passages prior to being split and re-seeded as clumps. Scale bars, 200 µm. **(b)** Expression of lung-related genes via RT-qPCR of lung organoids of increasing maturity from passage 3 to 6 with bars representing normalised expression to the housekeeping gene PBGD (ΔΔCt) ± SEM. (N=2, 3 technical replicates per biological replicate; open symbols represent IPF1.12 (SFTPC$^{WT/MUT}$) and closed symbols represent IPF5.4 (SFTPC$^{WT/MUT}$) hiPSCs).

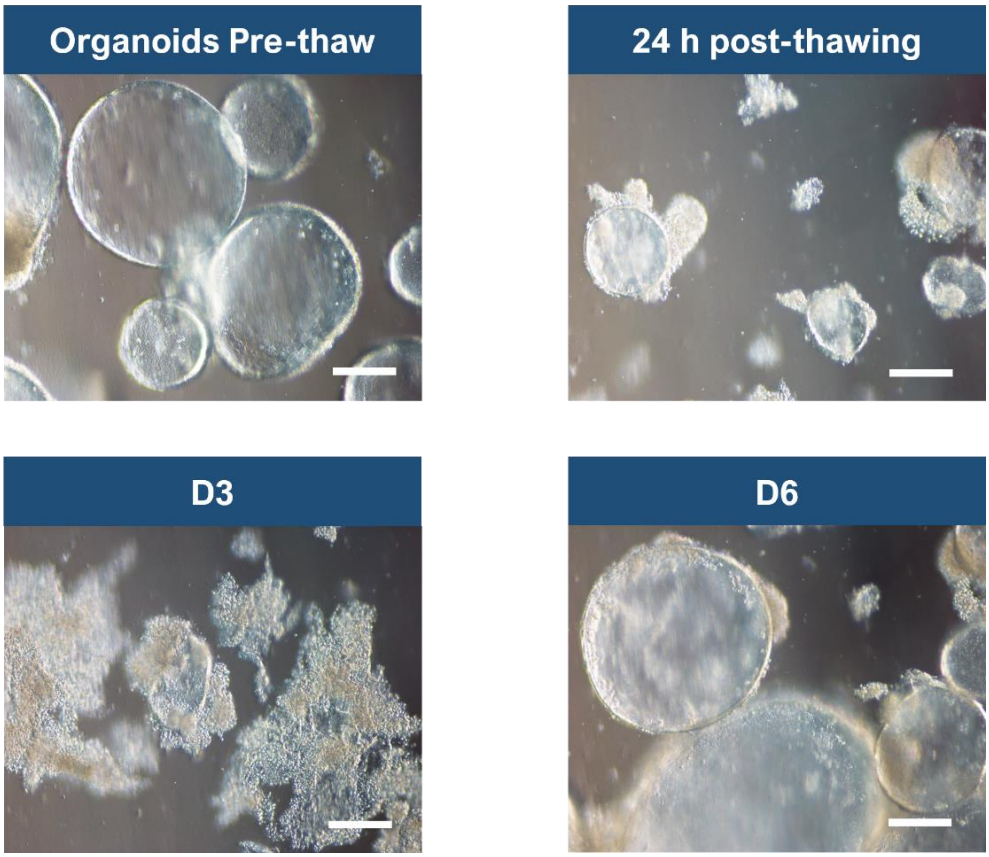


Figure 3-13. Organoids can be recovered after cryopreservation.

Representative light microscope images of IPF1.12 ($SFTPC^{WT/MUT}$) organoids passaged as clumps pre- and post-cryopreservation. Scale bars, 200 µm.

3.2.7. Single cell passaging of organoids and enrichment via FACS for CPM and NaPi2B

Given the poor induction of lung-related genes in our organoid cultures, we reasoned that clump passaging might be promoting the expansion of non-lung endoderm populations. We hypothesised that passaging organoids as single cells and using cell surface markers for enrichment would purify our organoid system. Two cell surface proteins, CPM and NaPi2B, were previously identified to enrich for NKX2.1+ lung progenitor cells and SFTPC+ type 2 AECs, respectively (Gotoh et al., 2014; Korogi et al., 2019). We first wanted to determine whether organoids could be passaged as single cells. Organoids dissociated to single cells and seeded into Matrigel™ droplets self-assembled into 3D organoids and expanded with further culture (**Fig. 3-14a**). Next, we performed FACS on IPF1.12 organoids that had been maintained in 3D culture and showed that at both passage 7 and 9 we were unable to enrich for CPM or NaPi2B cells, suggesting lack of lung-related cells in organoids maintained by single-cell passaging, similar to clump passaging (**Fig. 3-14b**). As a proof-of-principle, after FACS we collected and embedded NaPi2B-negative cells in Matrigel™ droplets and observed the self-assembly into large, cystic organoids with a thin epithelium (**Fig. 3-14c**). This highlighted the capacity for FACS-sorted single cells to be put back into culture and give rise to 3D organoids, despite the lack of lung markers.

3.2.8. Lung progenitor cells in monolayer culture lack CPM expression

As CPM has been shown to enrich for lung progenitor cells, we attempted to isolate a CPM+ population from an earlier timepoint in our differentiation. All three hiPSC lines were differentiated using the pre-established protocol and assessed for CPM expression on day 8 of lung specification. There was a lack of CPM cells across all hiPSC lines, indicating failed induction to a lung progenitor population (**Fig. 3-15a, b**). To determine whether our differentiation basal medium and growth factor concentrations were refractory to lung differentiation, we tested our conditions for specification of lung progenitor cells against a recently published protocol for lung differentiation (Jacob et al., 2017). We used BE.32 ($SFTPC^{WT/WT}$) hiPSCs and performed differentiation in our normal RPD medium with CHIR (1 µM), BMP4 (5 ng/ml) and RA (250 nM) (condition A) and compared it to SFL basal medium with our growth factor

concentrations (condition B) and SFL basal medium with slightly altered growth factor concentrations: CHIR (3 µM), BMP4 (10 ng/ml) and RA (100 nM) (condition C). Again, there was lack of any significant expression of CPM of lung day 8 (**Fig. 3-15c**), indicating the basal medium composition and growth factors were not responsible for the limited differentiation to lung cells.

Although we had demonstrated the feasibility of differentiating hiPSCs, generating 3D organoids, and passaging them as single cells for FACS, unfortunately, we could not induce a lung identity and as a result the enrichment of CPM+ and NaPi2B+ cells was unsuccessful. To address the poor differentiation capacity with our newly established hiPSC lines, we decided it was necessary to optimise the current differentiation platform from the beginning.

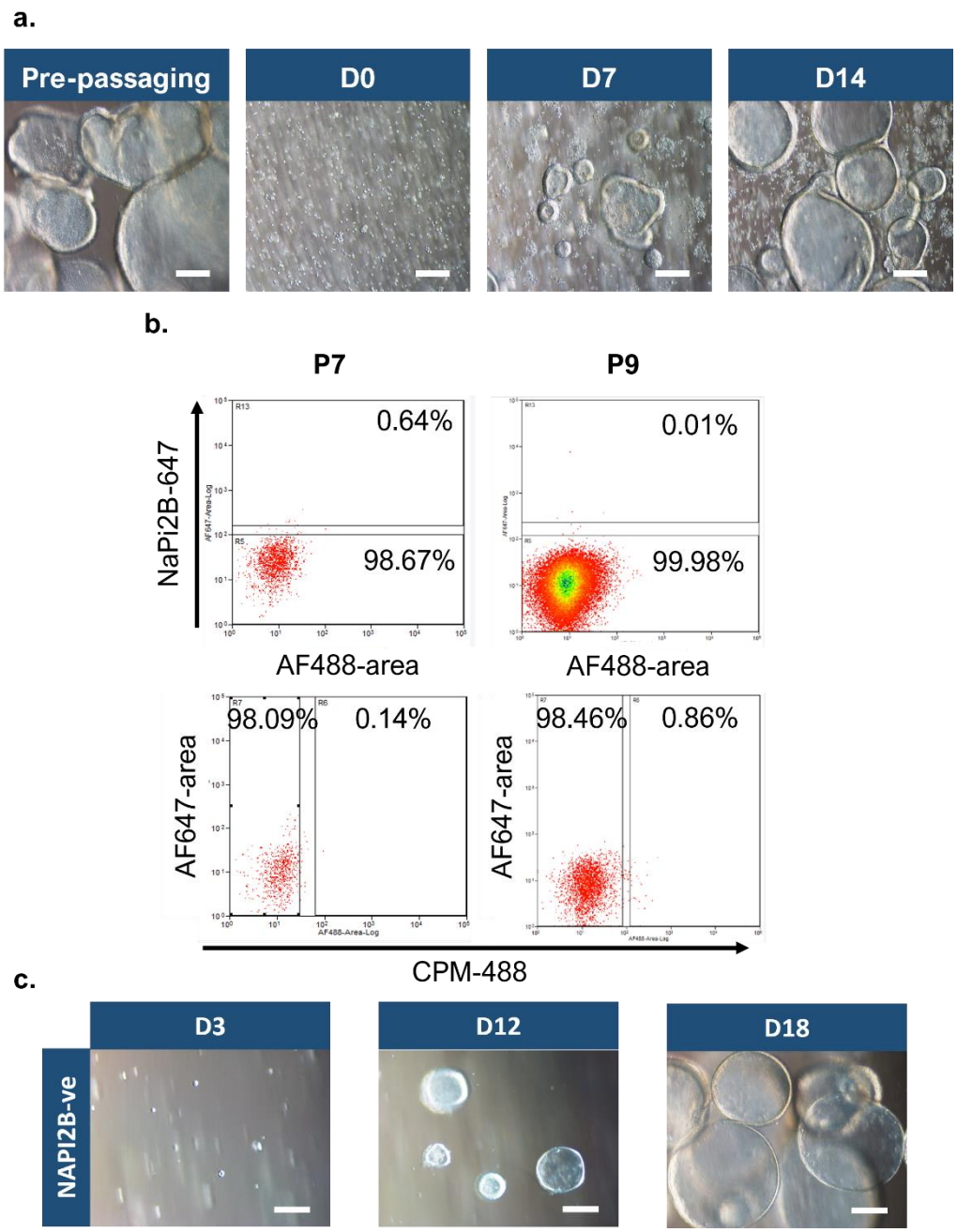


Figure 3-14. Organoids can be dissociated into single cells for FACS.

(a) Representative light microscope images of IPF1.12 (SFTPC$^{WT/MUT}$) organoids passaged and seeded as single cells into 3D Matrigel™ matrix. **(b)** FACS plots of IPF1.12 organoids sorted based on the expression of CPM and NaPi2B for two independent passages. **(c)** Representative light microscope images of NaPi2B-ve cells self-assembling into 3D organoids in Matrigel™ matrix. Scale bars, 200 µm.

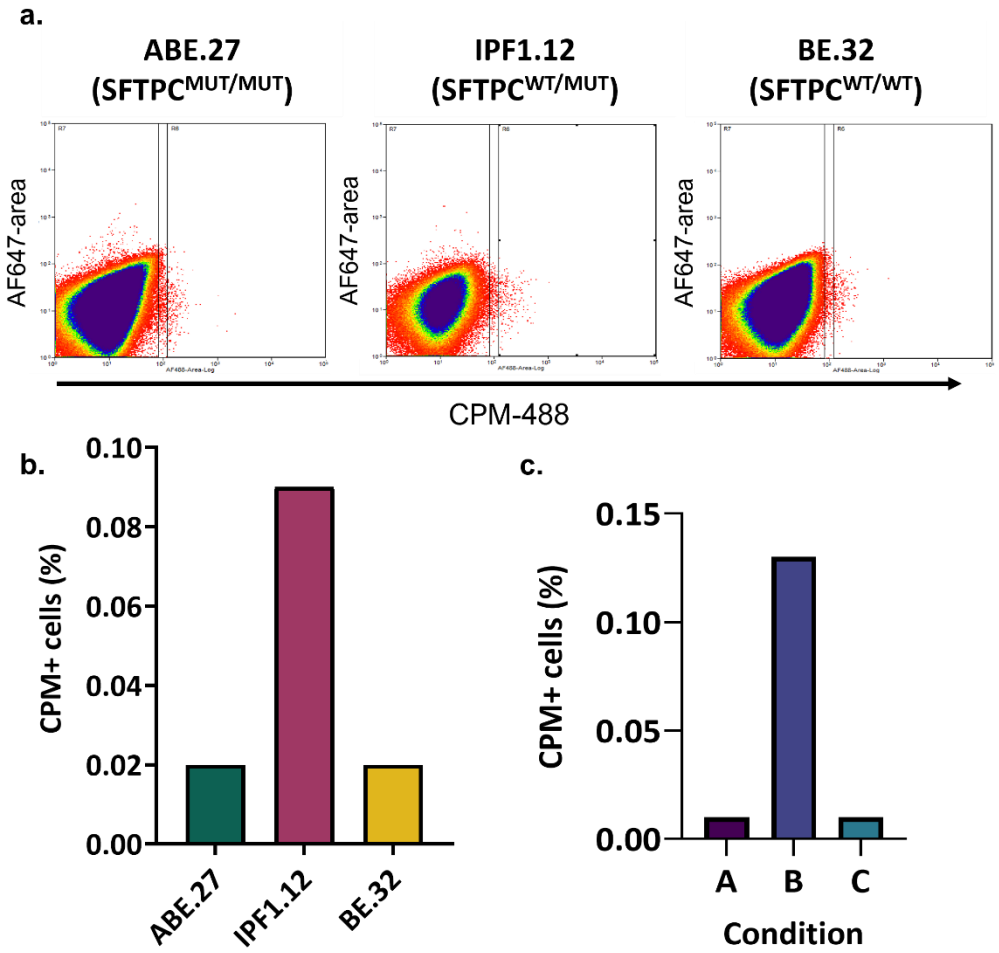


Figure 3-15. CPM is expressed at very low levels after 8 days of lung differentiation.

(a) FACS plots of cells sorted on lung D8 for CPM for ABE.27 (SFTPC$^{MUT/MUT}$), IPF1.12 (SFTPC$^{WT/MUT}$) and BE.32 (SFTPC$^{WT/WT}$) hiPSC lines. **(b)** Quantification of CPM+ cells by flow cytometry on day 8 of lung differentiation. **(c)** Comparison of BE.32 (SFTPC$^{WT/WT}$) CPM+ cells by flow cytometry on lung day 8 of differentiation for three separate lung differentiation growth conditions; Condition A = RPD basal medium supplemented with CHIR (1 µM), BMP4 (5 ng/ml) and RA (250 nM), condition B = SFL basal medium supplemented with CHIR (1 µM), BMP4 (5 ng/ml) and RA (250 nM), condition C = SFL basal medium supplemented with CHIR (3 µM), BMP4 (10 ng/ml) and RA (100 nM).

3.2.9. Growth factor optimisation of definitive endoderm specification

A homogenous starting population of DE cells is critical for the efficient generation of lung progenitor cells. Previously, we had shown limited expression of the endoderm marker SOX17 in our DE cells, so we reasoned that subsequent differentiation might be a result of poor endoderm induction. To assess the differentiation efficiency during optimisation of definitive endoderm specification, we used the expression of two surface markers, CXCR4 (CD184) and CKIT (CD117), known to be expressed in stem cell-derived endoderm. Using flow cytometry, we established a gating strategy to quantify the percentage of endoderm cells expressing CXCR4 and CKIT (**Fig. 3-16a**). We first assessed our original protocol of WNT and activin A signalling alone, and with activation of BMP and FGF signalling, known promoters of DE identity, on the induction of DE on day 3, 4 and 5 of differentiation (**Fig. 3-17a**). For all conditions, cells expanded and formed a monolayer, but hiPSC aggregates persisted across all days and by day 5 cells exposed to WNT and activin alone started to undergo apoptosis (**Fig 3-17b**). WNT and activin A stimulation resulted in very poor numbers of CXCR4 and CKIT cells, and although the activation of BMP4 and FGF2 signalling had a synergistic effect on the induction of DE, CXCR4+ cells were <40% (**Fig. 3-18a, b**) and CKIT+ cells <50% at the peak of expression on day 3 (**Fig. 3-19a, b**). Prolonged culture had a negative impact on the number of DE cells. Gene expression analysis revealed that the expression of *NANOG* was progressively downregulated in a temporal manner and potentiated with the subsequent addition of BMP4 and FGF2 but persisted strongly in the presence of WNT and activin signalling only, expressed nearly 2-fold more than in hiPSCs on day 3, and was still expressed in cultures with addition of BMP4 (**Fig. 3-20a**). This indicates WNT, activin and BMP4 signalling in combination is insufficient to drive hiPSC differentiation to DE resulting in persistence of hiPSCs in culture. The DE markers were upregulated across all conditions, with *GATA4* and *FOXA2* expression peaking on day 3. Expression of *SOX17* was highest on day 5 in the presence of WNT, activin, BMP4 and FGF2 signalling, but there was significant variation in *SOX17* expression between hiPSC lines (**Fig. 3-20a**). The hindgut marker *CDX2* displayed variable expression across all conditions and days but given the large variation in technical repeats suggests this was background amplification during the qRT-PCR reaction (**Fig. 3-20a**). These findings

show an incomplete differentiation to endodermal identity with contamination of persisting pluripotent stem cell aggregates.

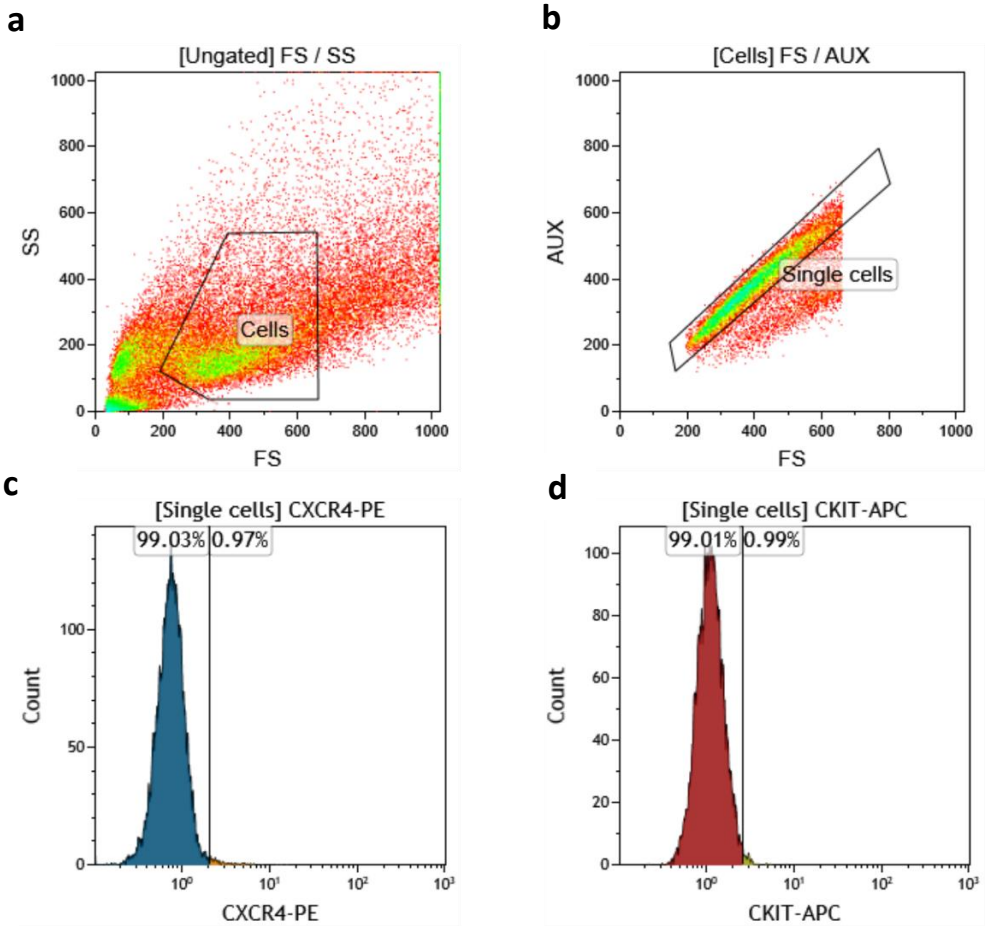


Figure 3-16. Gating strategy for flow cytometry analysis of CXCR4 and CKIT.

(a) Flow cytometry plot of forward scatter (FS) vs. side scatter (SS) to identify cells based on size and gating to exclude debris. **(b)** Flow cytometry plot of FS area vs FS height to identify single cells and exclude doublet cells observed under the gate. Flow cytometry plots of isotype controls for **(c)** CXCR4 and **(d)** CKIT used as a negative control to determine non-specific binding of the antibody.

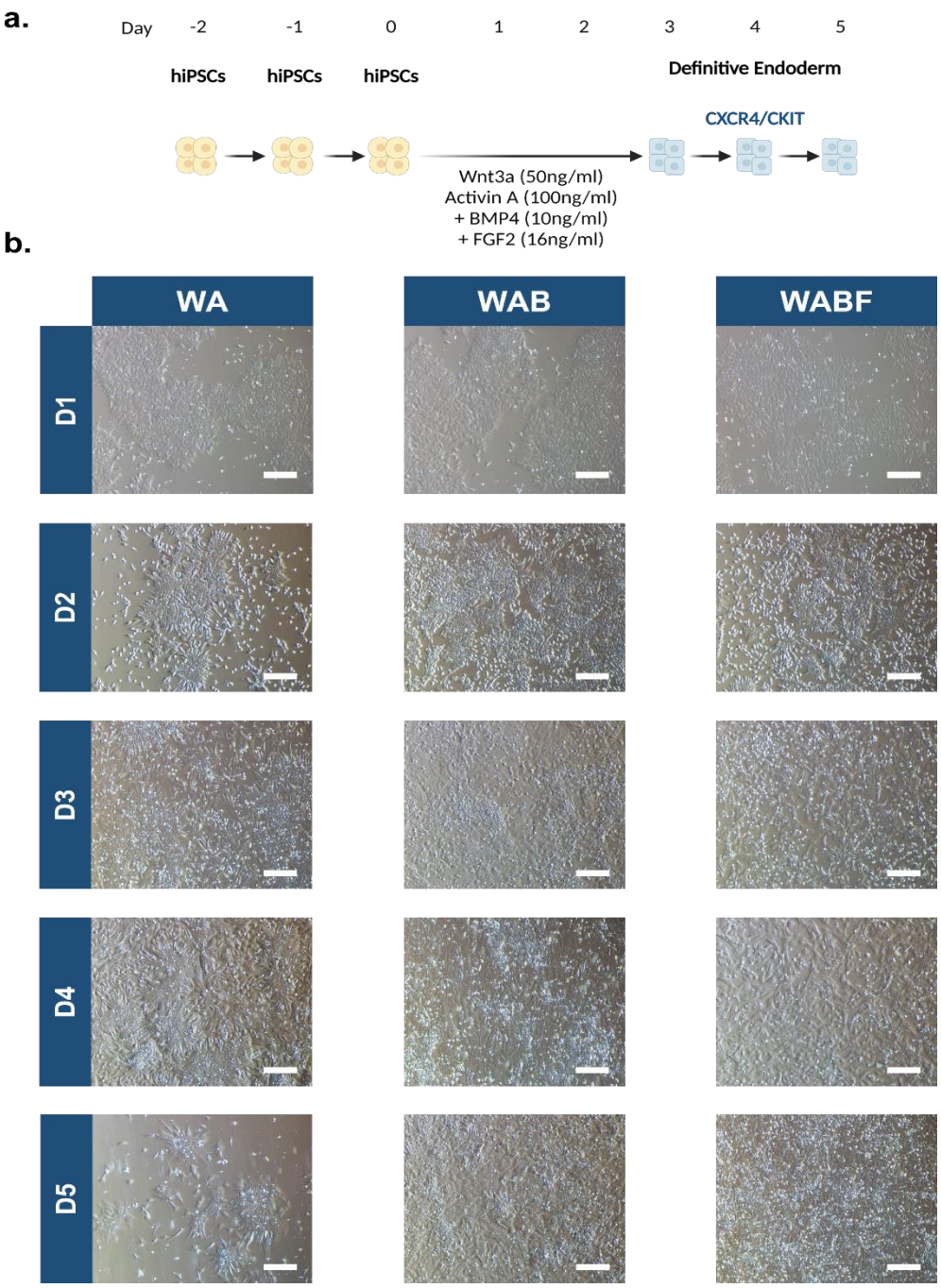


Figure 3-17. Optimisation of definitive endoderm specification using recombinant proteins.

(a) Schematic of the growth factor combinations for the differentiation of hiPSCs to definitive endoderm with analysis for CXCR4 and CKIT expression on days 3, 4 and 5. **(b)** Representative light microscope images of BE.32 ($SFTPC^{WT/WT}$) cells at stages of endoderm differentiation for each optimisation condition. Scale bars, 200 µm WA = WNT3a + Activin A; WAB = WNT3a + Activin A + BMP4; WABF = WNT3a + Activin A + BMP4 + FGF2.

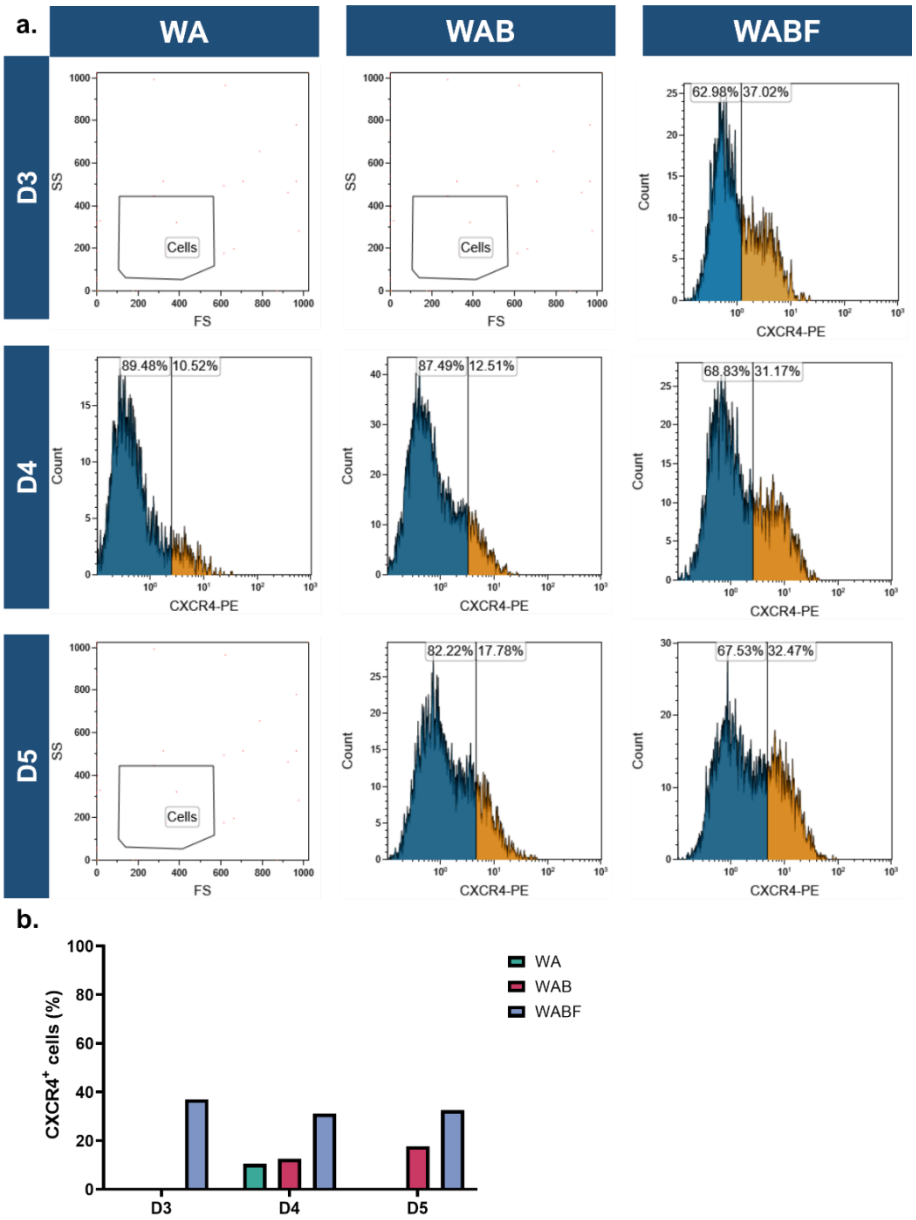


Figure 3-18. Induction of BE.32 (SFTPC$^{WT/WT}$) CXCR4-positive endoderm cells is sub-optimal using recombinant proteins.

(a) Flow cytometry analysis of CXCR4 expression on day 3, 4 and 5 of endoderm differentiation with combinations of growth factors and **(b)** quantification of CXCR4-positive cells (N=1). Blank plots reflect too few cells for analysis or conditions that had crashed.WA = WNT3a + Activin A; WAB = WNT3a + Activin A + BMP4; WABF = WNT3a + Activin A + BMP4 + FGF2. The following concentrations were used; WNT3a (50 ng/ml), Activin A (100 ng/ml), B = BMP4 (10 ng/ml) and FGF2 (16ng/ml).

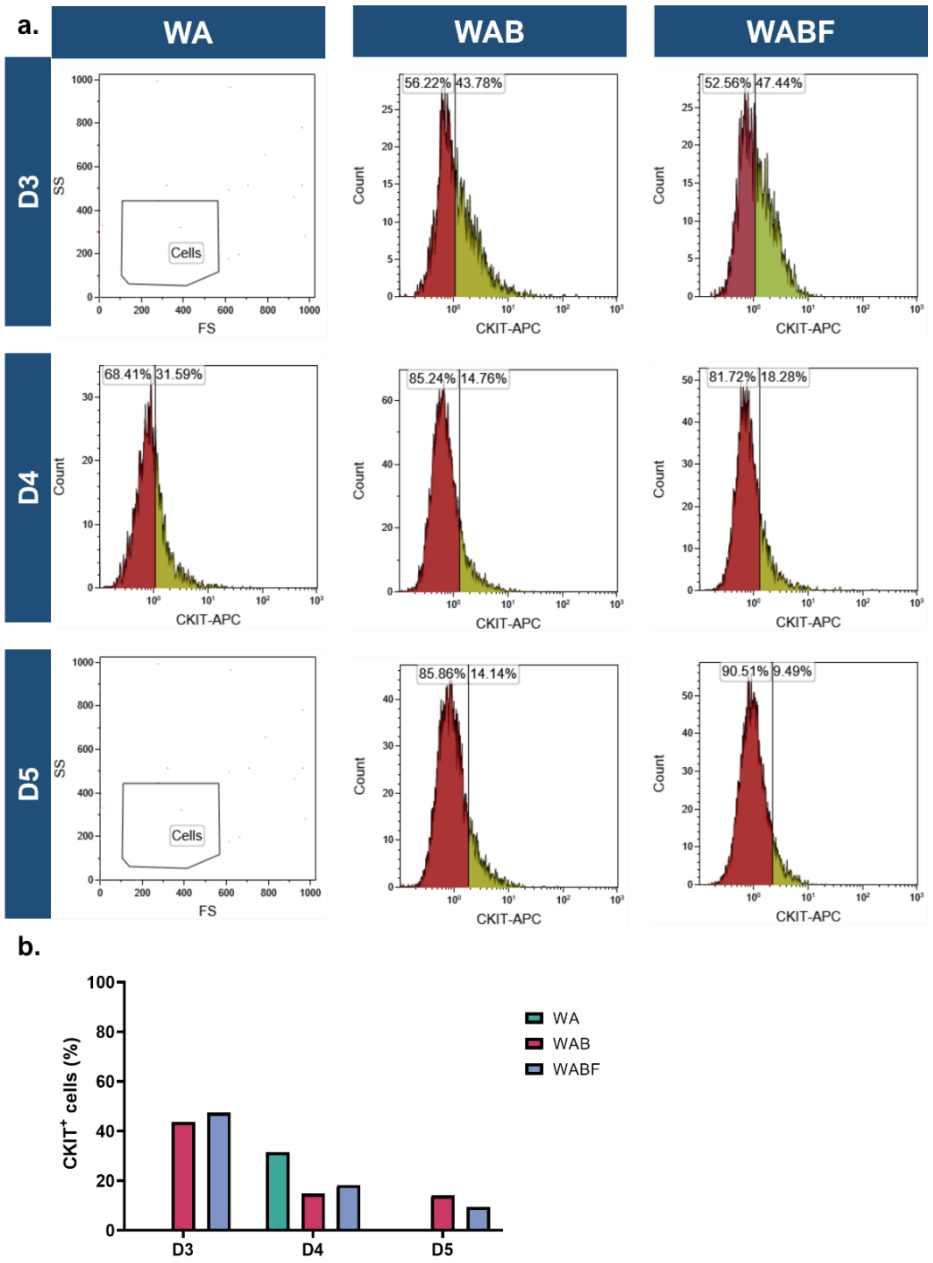


Figure 3-19. Induction of BE.32 (SFTPC$^{WT/WT}$) CKIT-positive endoderm cells is sub-optimal using recombinant proteins.

(a) Flow cytometry analysis of CKIT expression on day 3, 4 and 5 of endoderm differentiation with combinations of growth factors and **(b)** quantification of CKIT-positive cells (N=1). WA = WNT3a + Activin A; WAB = WNT3a + Activin A + BMP4; WABF = WNT3a + Activin A + BMP4 + FGF2. The following concentrations were used; WNT3a (50 ng/ml), Activin A (100 ng/ml), B = BMP4 (10 ng/ml) and FGF2 (16ng/ml).

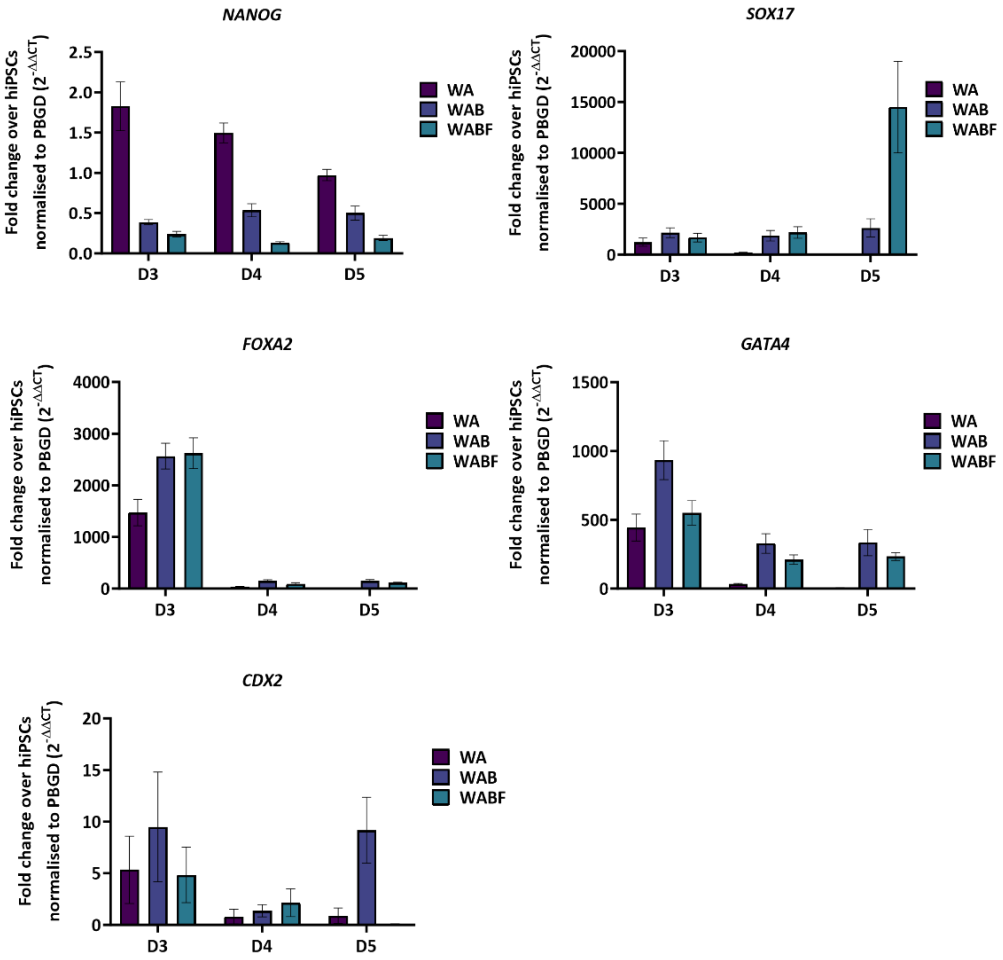


Figure 3-20. Definitive endoderm cells specified with recombinant proteins express endoderm gene markers.

Expression of pluripotency and definitive endoderm genes via RT-qPCR per condition on day 3, 4 and 5 of differentiation. Bars represent normalised expression to the housekeeping gene PBGD relative to hiPSCs ($2^{-\Delta\Delta Ct}$) ± SEM (N = 1, 3 technical repeats for BE.31 ($SFTPC^{WT/WT}$) and BE.32 ($SFTPC^{WT/WT}$) hiPSCs).

3.2.10. Addition of LY294002 significantly improves differentiation efficiency to definitive endoderm

The presence of cell aggregates and *NANOG* expression after endoderm differentiation indicated a resistance of hiPSCs to migration and differentiation outwards from hiPSC colonies. To promote further differentiation, we tested the effect of the addition of the small molecule PI3K inhibitor LY294002 from day 0 to day 2 as it is utilised for cell lines resistant to DE differentiation (**Fig. 3-21a**). Interestingly, in the presence of LY294002, cells exposed to WNT3a and activin A started to undergo apoptosis on day 3 and by day 4 most cells had detached. With the additional supplementation of BMP4 and FGF2 there was increased migration of cells from colonies and a subsequent protective effect, although aggregation of hiPSC colonies was still visible on day 5 (**Fig. 3-21b**). However, LY294002 significantly potentiated the induction of endodermal cells. On differentiation day 3 there was limited induction to DE, but on day 4 there was a robust increase in the expression of CXCR4 (**Fig. 3-22a, b**) and CKIT (**Fig. 3-23a, b**), with ~95% CXCR4+ and ~83% CKIT+ cells quantified on day 5 in the presence of all 4 recombinant proteins with LY294002 (**Fig. 3-22b, 23b**). Next, we aimed to reproduce this efficiency of endoderm differentiation without reliance on as many recombinant proteins.

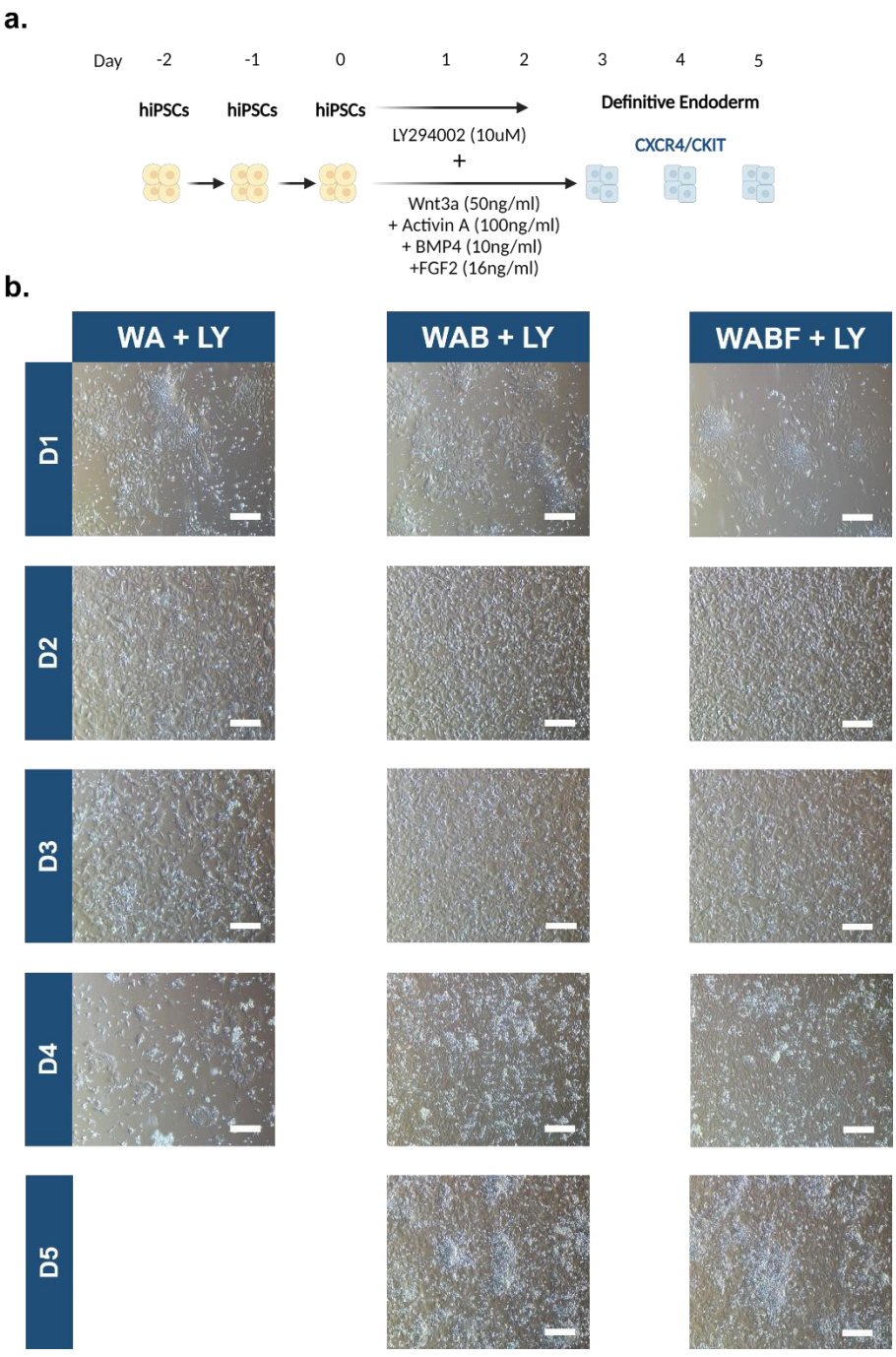


Figure 3-21. Optimisation of definitive endoderm specification with the addition of LY294002 in BE.32 (SFTPC$^{WT/WT}$) hiPSCs.

(a) Schematic of the growth factor combinations with the addition of the PI3K inhibitor LY294002 (LY) for definitive endoderm differentiation. **(b)** Representative light microscope images of cells at each day of endoderm differentiation. Blank spaces reflect cells crashing. Scale bars, 200 µm. WA + LY = WNT3a + Activin A + LY; WAB + LY = WNT3a + Activin A + BMP4 + LY; WABF + LY = WNT3a + Activin A + BMP4 + FGF2 + LY. Cells were treated with LY294002 from day 0 to day 2 only.

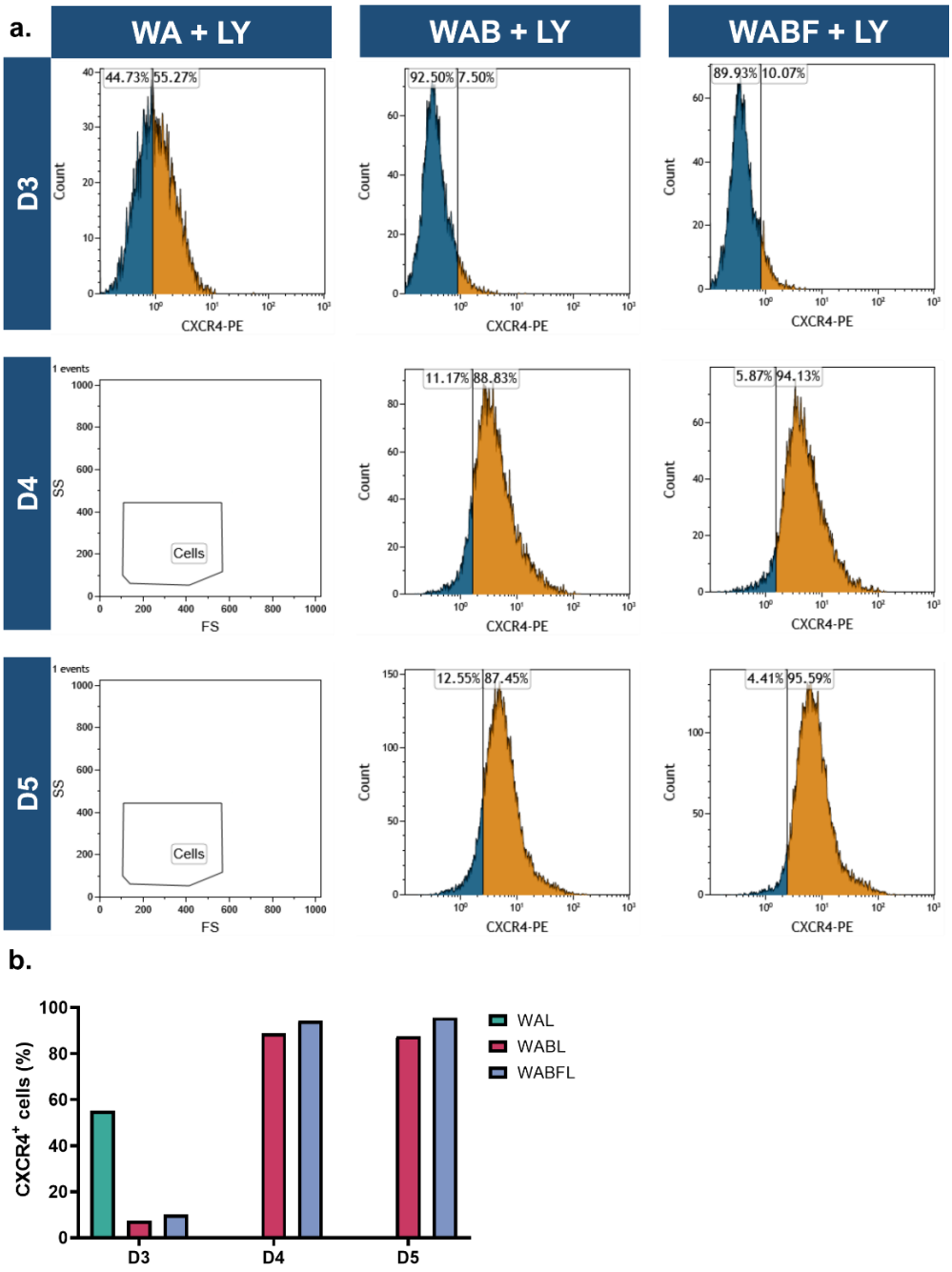


Figure 3-22. Addition of LY294002 improves induction of BE.32 (SFTPC$^{WT/WT}$) CXCR4-expressing endoderm cells.

(a) Flow cytometry analysis of CXCR4 on day 3, 4 and 5 of endoderm differentiation with the addition of LY294002 and **(b)** quantification of CXCR4-positive cells (N=1).Blank plots reflect too few cells for analysis owing to cell apoptosis. WA + LY = WNT3a + Activin A + LY; WAB + LY = WNT3a + Activin A + BMP4 + LY; WABF + LY = WNT3a + Activin A + BMP4 + FGF2 + LY. The following concentrations were used; WNT3a (50 ng/ml), Activin A (100 ng/ml), B = BMP4 (10 ng/ml), FGF2 (16ng/ml) and LY294002 (10µM).

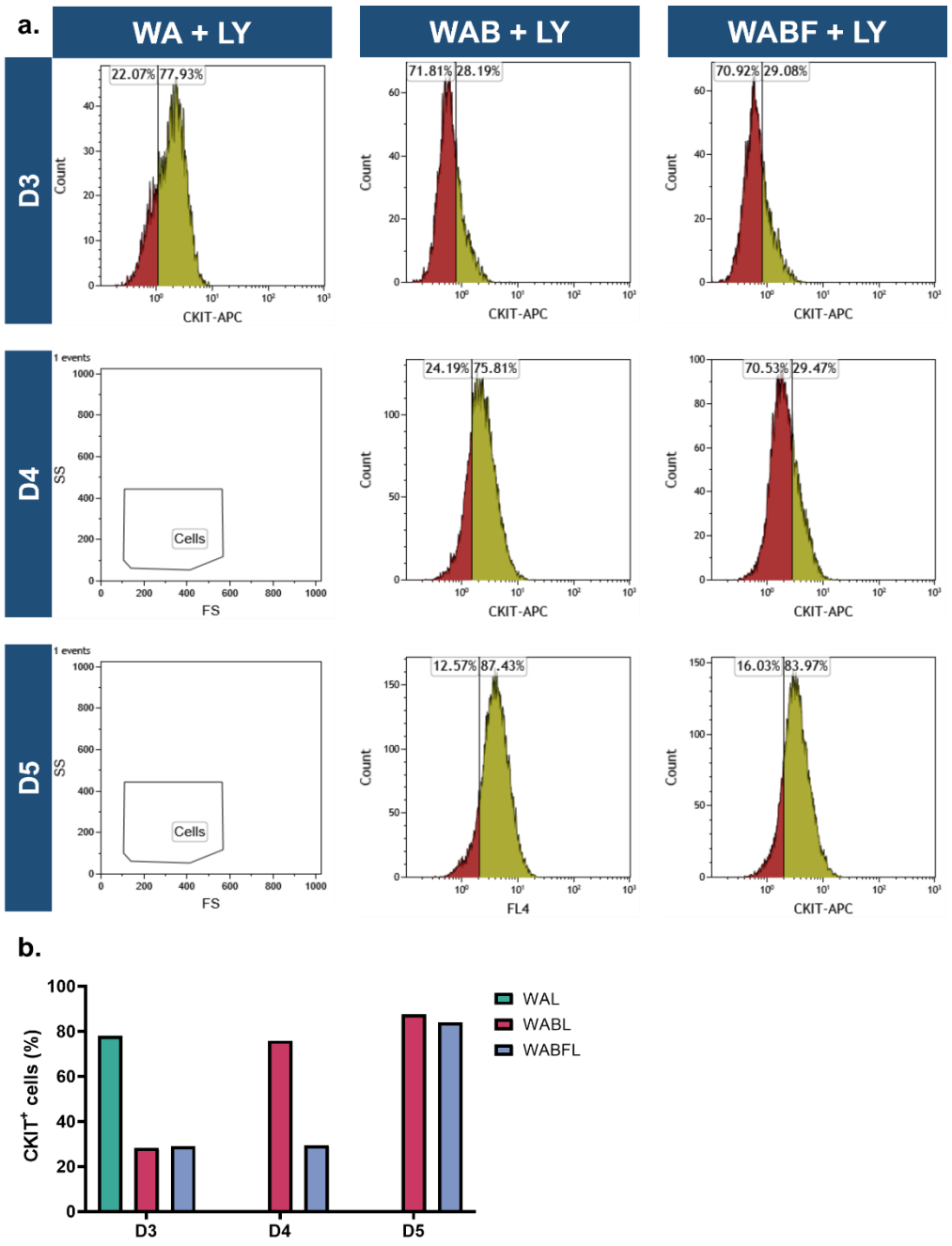


Figure 3-23. Addition of LY294002 improves induction of BE.32 (SFTPC$^{WT/WT}$) CKIT-expressing endoderm cells.

(a) Flow cytometry analysis of CKIT on day 3, 4 and 5 of endoderm differentiation with the addition of LY294002 and **(b)** quantification of CXCR4-positive cells (N=1). Blank plots reflect too few cells for analysis owing to cell apoptosis. WA + LY = WNT3a + Activin A + LY; WAB + LY = WNT3a + Activin A + BMP4 + LY; WABF + LY = WNT3a + Activin A + BMP4 + FGF2 + LY. The following concentrations were used; WNT3a (50 ng/ml), Activin A (100 ng/ml), B = BMP4 (10 ng/ml), FGF2 (16ng/ml) and LY294002 (10µM).

respective conditions (**Fig. 3-24a**). In the presence of Wnt3a, activin A and LY294002, cells underwent apoptosis after a day, but in the presence of CHIR cells expanded and migrated from their colonies to form a monolayer by day 2 (**Fig. 3-24b**). In the presence of CHIR99021, Activin A and LY294002, two CXCR4 populations were observed on day 3 of endoderm specification, but by day 4 and 5 there was a single homogenous CXCR4-expressing population (**Fig. 3-25a**) with a peak on day 4 with >99% CXCR4+ cells (**Fig. 3-25b**). There was no improvement in the efficiency of CXCR4-expressing cells with BMP and FGF signalling activation (**Fig. 3-25b**). Similarly, by day 4 and 5 of differentiation there was a single CKIT-expressing population with >90% of CKIT+ cells on day 4 (**Fig. 3-26a, b**). As a result of these findings, we decided to use the CA + LY condition for 4 days as our optimised DE protocol.

To quantify the number of endoderm cells co-expressing CXCR4 and CKIT, we used a co-staining flow cytometry strategy (**Fig. 3-27a**), and translated this to a new, replacement cytometer (**Fig. 3-27b**) to be used for subsequent experiments owing to the current flow cytometer being decommissioned. Collectively, our optimised endoderm differentiation protocol of CHIR99021 (D0 – D1), LY294002 (D0 – D2) and Activin A (D0 – D4) (**Fig. 3-28a**) resulted in a mean CXCR4 expression of 93.55 ± 5.6% and CKIT expression of 73.89 ± 7.6% (N=10) (**Fig. 3-28b**). For analysis on the new cytometer, the mean number of CXCR4/CKIT-double positive cells on day 4 for the BE.31 ($SFTPC^{WT/WT}$) hiPSCs was 97.68 ± 1.54% (N=4) and for the ABE.27 ($SFTPC^{MUT/MUT}$) hiPSCs was 95.57 ± 1.91% (N=3) (**Fig. 3-28c**). These results demonstrated that our optimised protocol led to robust differentiation to DE using small molecule inhibitors in combination with activin signalling.

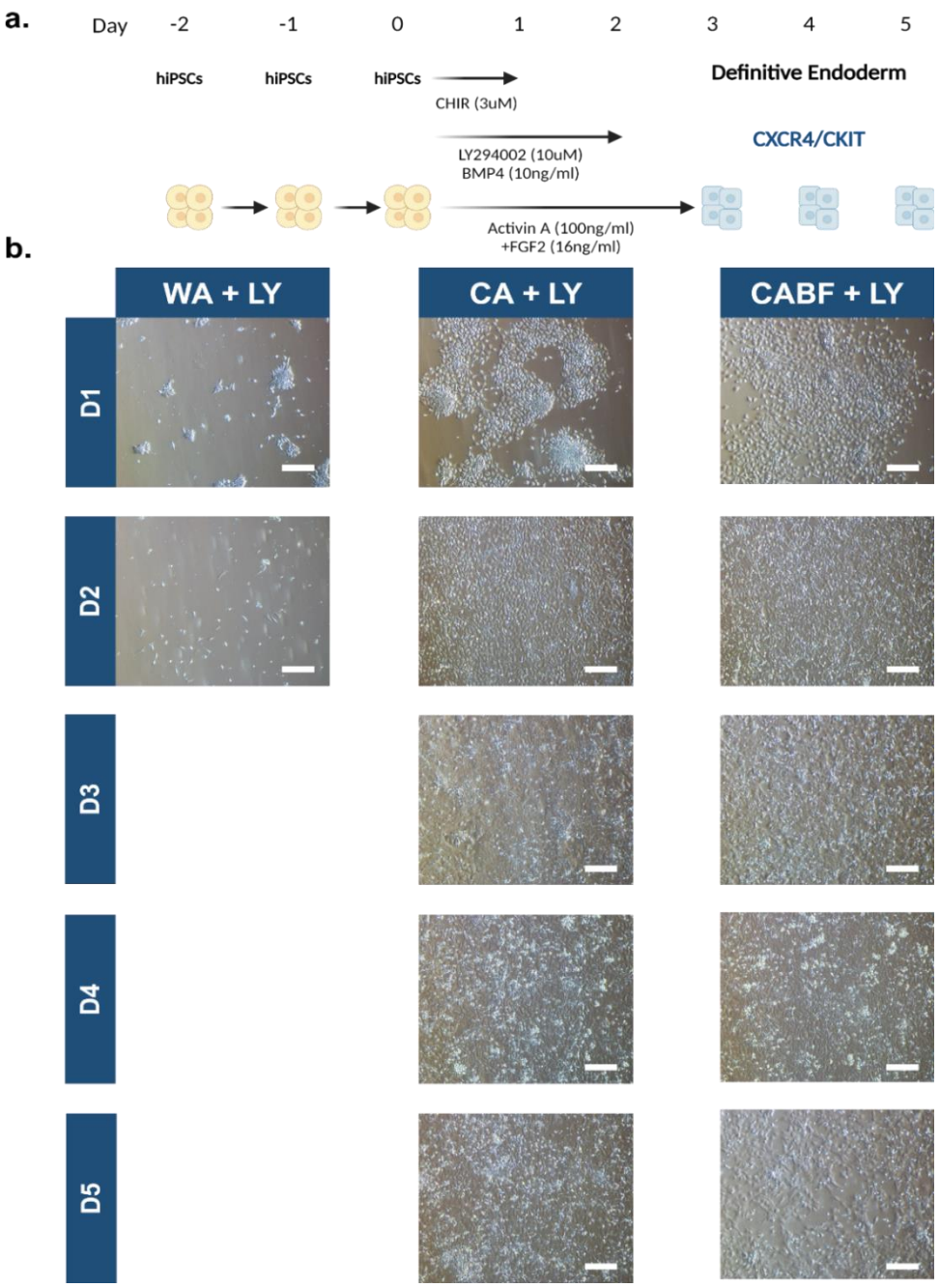


Figure 3-24. Optimisation of definitive endoderm differentiation using the small molecule inhibitor CHIR99021 in BE.32 (SFTPC$^{WT/WT}$) hiPSCs.

(a) Schematic of the differentiation of hiPSCs to endoderm cells using the small molecule inhibitor CHIR99021 (CHIR) in combination with LY294002 and recombinant proteins. **(b)** Representative light microscope images of endoderm cells. Scale bars, 200 µm. Blank spaces reflect cells crashing. WA + LY = WNT3a + Activin A + LY; CA + LY = CHIR + Activin A + LY; CABF + LY = CHIR + Activin A + BMP4 + FGF2 + LY. Cells were treated with CHIR99021 from day 0 to day 1 of differentiation only.

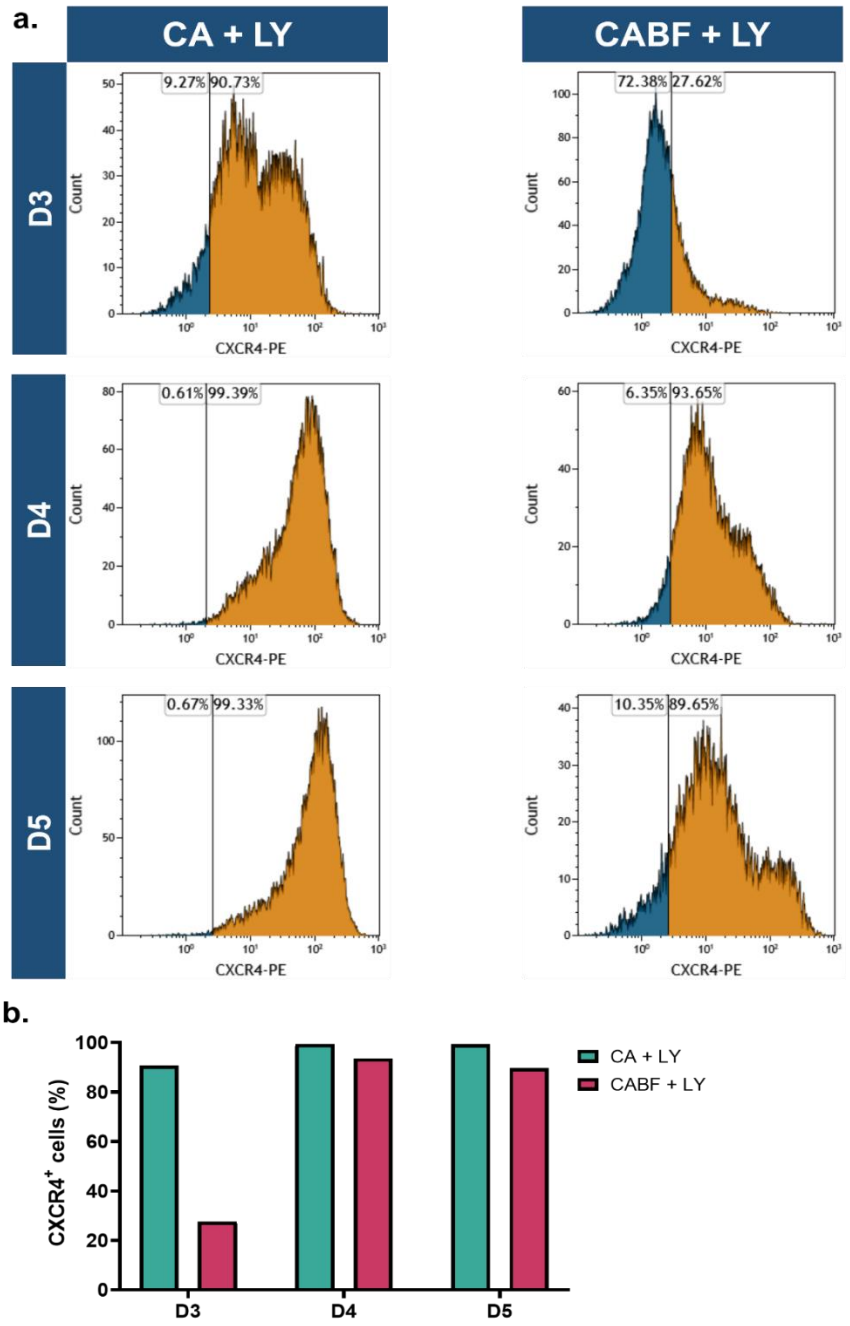


Figure 3-25. Induction of BE.32 (SFTPC$^{WT/WT}$) CXCR4-positive cells are promoted by CHIR99021.

(a) Flow cytometry analysis of CXCR4 on day 3, 4 and 5 of endoderm differentiation in the presence of CHIR99021 with **(b)** quantification of cells positive for CXCR4 (N=1). CA + LY = CHIR + Activin A + LY; CABF + LY = CHIR + Activin A + BMP4 + FGF2 + LY. The following concentrations were used; CHIR99021 (3 µM), Activin A (100 ng/ml), B = BMP4 (10 ng/ml), FGF2 (16ng/ml) and LY294002 (10µM).

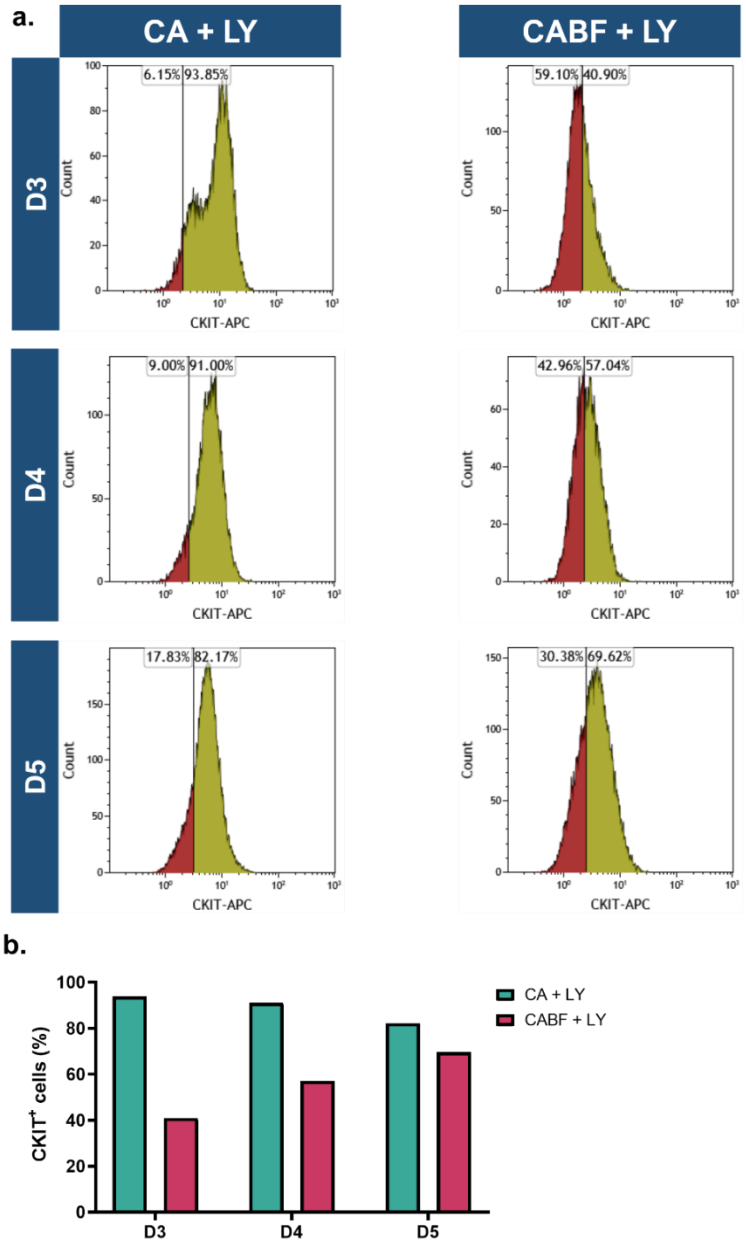


Figure 3-26. Induction of BE.32 (SFTPC$^{WT/WT}$) CKIT-positive cells are promoted by CHIR99021.

(a) Flow cytometry analysis of CKIT on day 3, 4 and 5 of endoderm differentiation in the presence of CHIR99021 with **(b)** quantification of cells positive for CKIT (N=1). CA + LY = CHIR + Activin A + LY; CABF + LY = CHIR + Activin A + BMP4 + FGF2 + LY. The following concentrations were used; CHIR99021 (3 µM), Activin A (100 ng/ml), B = BMP4 (10 ng/ml), FGF2 (16ng/ml) and LY294002 (10µM).

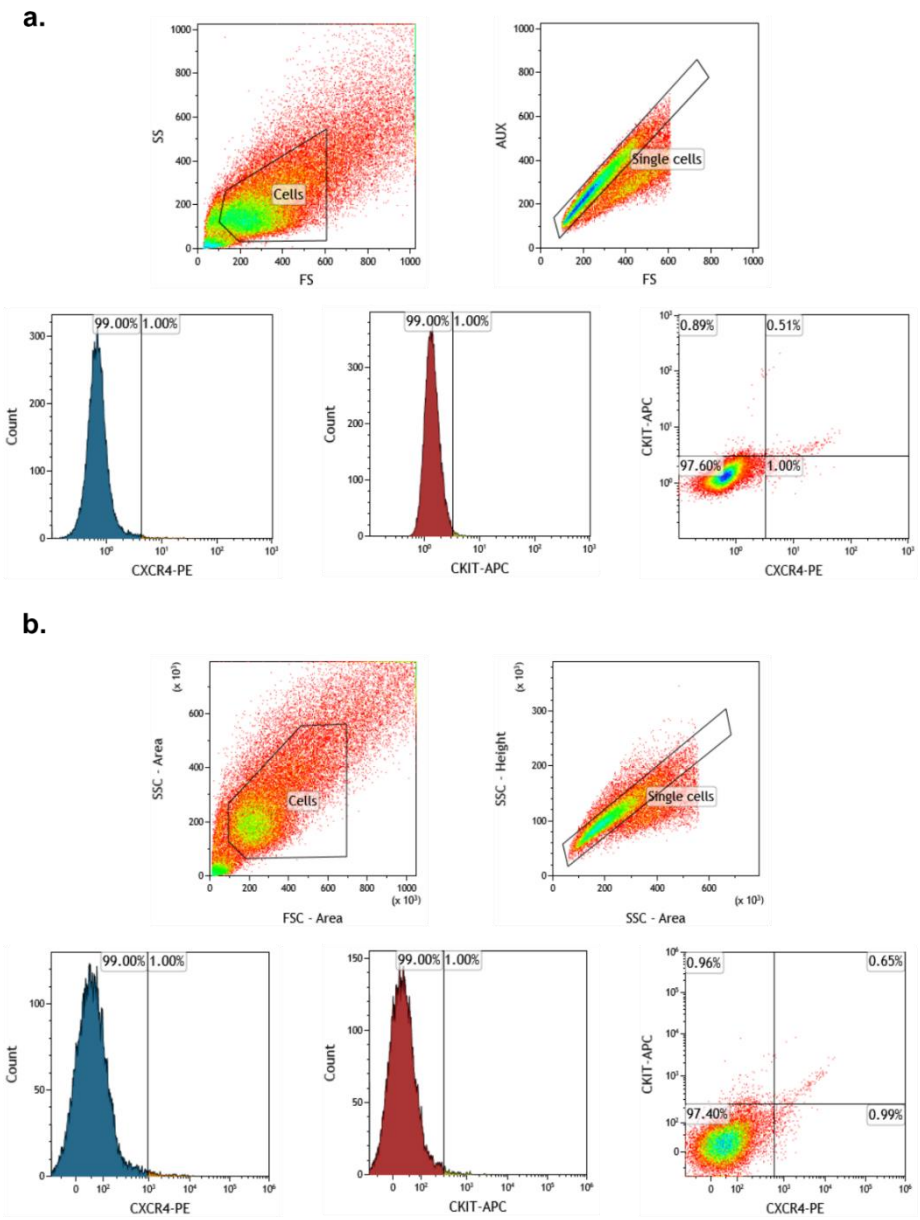


Figure 3-27. Flow cytometry gating strategy for co-staining of CXCR4-PE and CKIT-APC.

(a) Flow cytometry plots of forward scatter (FSC) vs. side scatter (SSC) and gating for cells to remove debris, FS area vs. FS height to identify single cells and exclude doublets, single plots of isotype antibody controls and dot plot for co-staining of isotype controls on the decommissioned FC500 flow cytometer. (b) Flow cytometry plots of the strategy to co-stain CXCR4-PE and CKIT-APC as above with the only difference being SSC area vs. SSC height for isolating single cells instead of FS area vs. FS height on the Sony ID7000 spectral flow cytometry analyser.

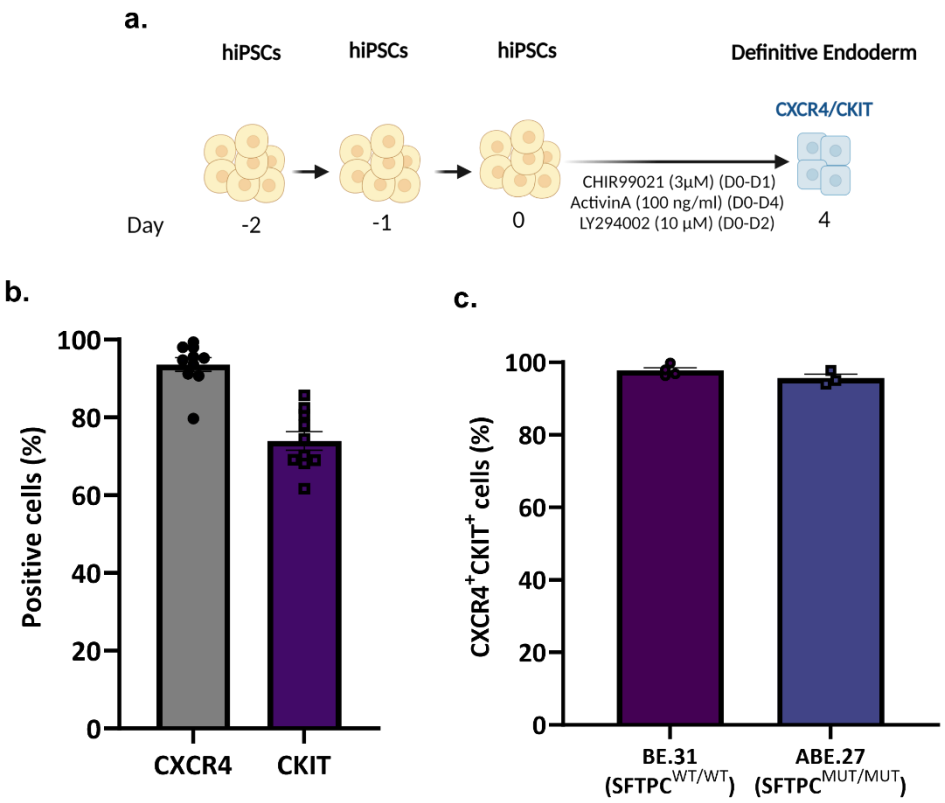


Figure 3-28. Quantification of CXCR4 and CKIT expression after 4 days of optimised endoderm differentiation.

(a) Schematic of the optimised 4-day protocol for differentiation of hiPSCs to DE. Figure generated using Biorender **(b)** Percentage of single BE.31 (SFTPC$^{WT/WT}$) CXCR4 and CKIT positive cells on day 4 of endoderm differentiation with bars representing mean ± SEM analysed by the decommissioned FC500 flow cytometer (N=10). **(c)** Double CXCR4/CKIT-positive cell percentages on day 4 of endoderm differentiation for SFTPC$^{WT/WT}$ (N=4) and SFTPC$^{MUT/MUT}$ (N=3) hiPSCs. Bars represent mean ± SEM analysed by the new ID7000 cytometer.

3.2.12. Characterisation of definitive endoderm cells

With our optimised endoderm differentiation protocol, we observed progressive upregulation of DE genes *SOX17*, *FOXA2*, *CXCR4* and *GATA4* across 4 days of differentiation. The hindgut marker *CDX2* was expressed on day 1 but readily downregulated from day 2 onwards (**Fig. 3-29a**) (N=1). For BE.31 ($SFTPC^{WT/WT}$) and ABE.27 ($SFTPC^{MUT/MUT}$) hiPSCs, 4 days of endoderm differentiation resulted in significant upregulation of *FOXA2*, *CXCR4*, *GATA4* and *SOX17* although there was some of variability in expression between biological replicates (**Fig. 3-29b**). On day 4 of differentiation, the majority of BE.31 ($SFTPC^{WT/WT}$) DE cells exhibited protein expression of SOX17 and FOXA2 (**Fig. 3-30a, b**). These results demonstrated that our optimised differentiation protocol could robustly generate DE cells. Next, we wanted to determine whether the improved efficiency in endoderm specification would positively impact the differentiation to lung epithelial progenitor cells.

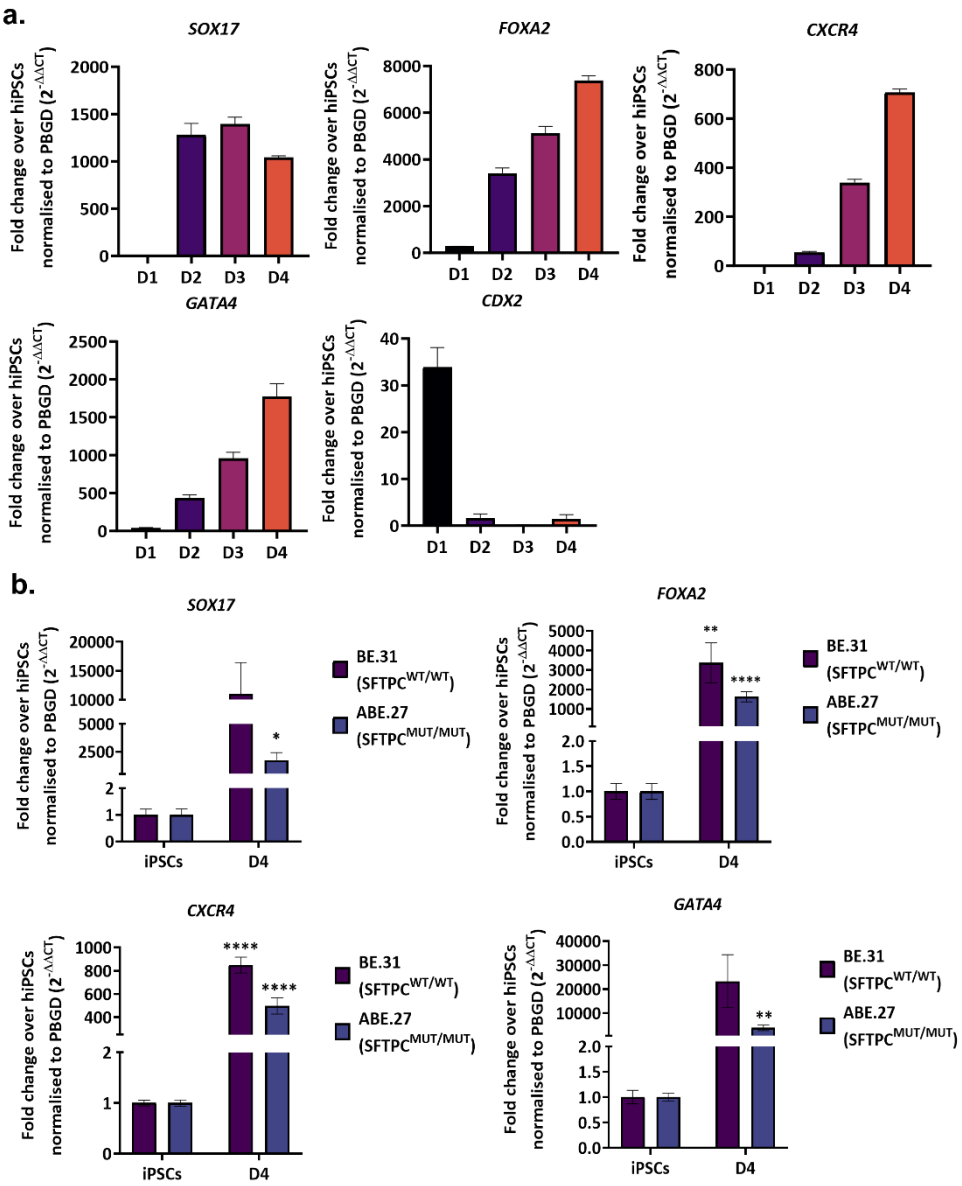


Figure 3-29. Definitive endoderm genes are expressed with an optimised endoderm differentiation protocol.

(a) Expression of endoderm-related genes over four days of endoderm differentiation in BE.32 (SFTPC$^{WT/WT}$) hiPSCs via RT-qPCR with bars representing normalised expression to the housekeeping gene PBGD relative to hiPSCs (2-ΔΔCt) ± SEM (N=1, 6 technical replicates). **(b)** RT-qPCR analysis of DE marker expression on day 4 of optimised differentiation in SFTPC$^{WT/WT}$ and SFTPC$^{MUT/MUT}$ hiPSCs. Bars represent normalised expression to the housekeeping gene PBGD relative to hiPSCs (2-ΔΔCt) ± SEM (N=3). *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001, unpaired, two-tailed Student's t test.

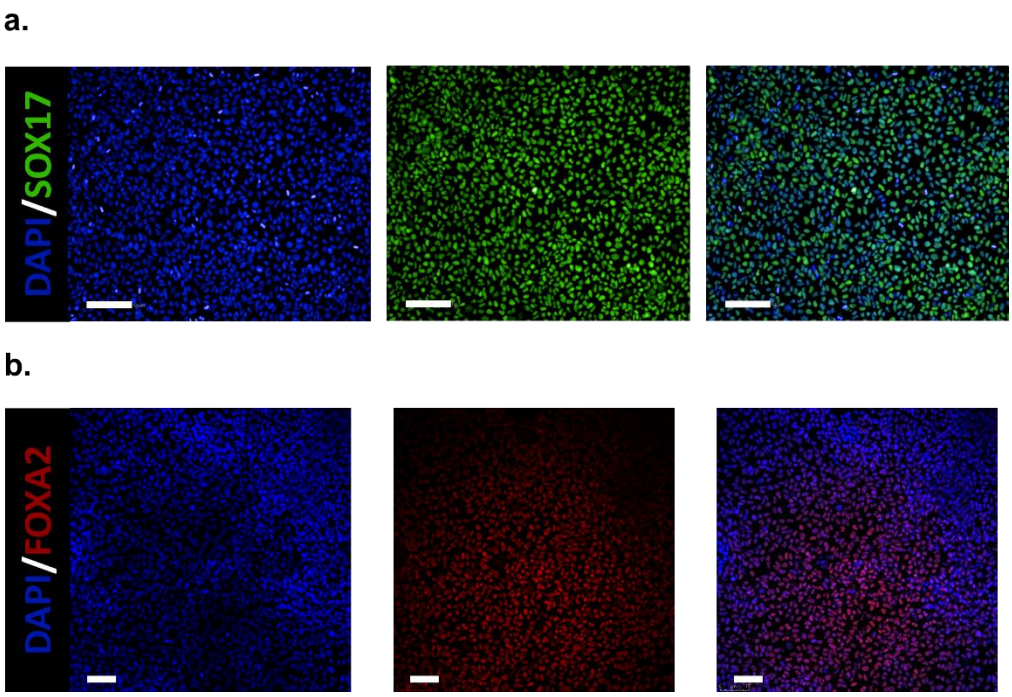


Figure 3-30. Protein expression of definitive endoderm markers.

(a) Representative immunofluorescence staining of SOX17 in BE.31 ($SFTPC^{WT/WT}$) day 4 endodermal cells imaged with the Operetta high content imaging system (nuclei stained with DAPI, blue). Scale bars, 100 µM. **(b)** Representative immunofluorescence staining of FOXA2 in BE.31 ($SFTPC^{WT/WT}$) day 4 endodermal cells with a Leica TCS SPE confocal microscope (nuclei stained with DAPI). Scale bars, 75 µm.

3.2.13. Optimisation of the differentiation of endoderm to lung epithelial progenitor cells using carboxypeptidase M as a surrogate marker for NKX2.1-positive cells

We reasoned that a homogenous DE starting population would limit the differentiation to off-target cell types. To quantify the efficiency of differentiation without the need for fluorescent reporter cell lines, we used CPM again as a surrogate marker for NKX2.1+ lung cells. For all experiments going forward we used a new CPM antibody validated for flow cytometry. We used our established protocol consisting of RPD medium supplemented with CHIR99021, BMP4 and retinoic acid as a control (**condition A**) (**Fig. 3-31a**). Findings from Green et al. (2011) showed that treatment of

DE cells with inhibitors of BMP and TGFβ signalling (dual-SMAD inhibition) was critical to enable cells to respond to signalling cues for lung specification. We therefore included a 3-day treatment with inhibitors of BMP and TGFβ signalling via SB431542 and dorsomorphin, respectively, after DE induction and prior to lung differentiation. We tested the effect of dual-SMAD inhibition with our RPD medium and the same growth factors (**condition B**) and compared this to SFL medium with slightly altered growth factor concentrations (**condition C**) (**Fig. 3-31a, b**). Analysis for CPM by flow cytometry was performed on day 8, 10 and 12 of lung differentiation. For each condition, a highly confluent monolayer was formed but there were areas where cells had aggregated and formed 3D clusters (**Fig. 3-31d**). Across all conditions, we observed the induction of cells expressing CPM on all days analysed which was a significant improvement from prior results when no CPM+ cells were isolated. There were no major differences in the number of CPM+ cells between condition A and condition B, indicating treatment with DS/SB had no effect. However, we did observe a difference in the expression profile of condition C where a high CPM-expressing population had emerged. At the time we deemed this to be insignificant (**Fig. 3-32a**). The percentage of CPM+ cells peaked at 56% on day 10 under condition C, but there was significant biological variability (**Fig. 3-32b**). As CPM+ cell numbers did not differ significantly between conditions, we continued with our original protocol (**condition A**) (**Fig. 3-32b**). We repeated this condition and generated CPM+ cells across 3 cell lines with a mean percentage of 29.99 ± 20.44% (mean ± SEM) CPM+ cells, highlighting continued variability in efficiency between cell lines and biological replicates (**Fig. 3-32c**).

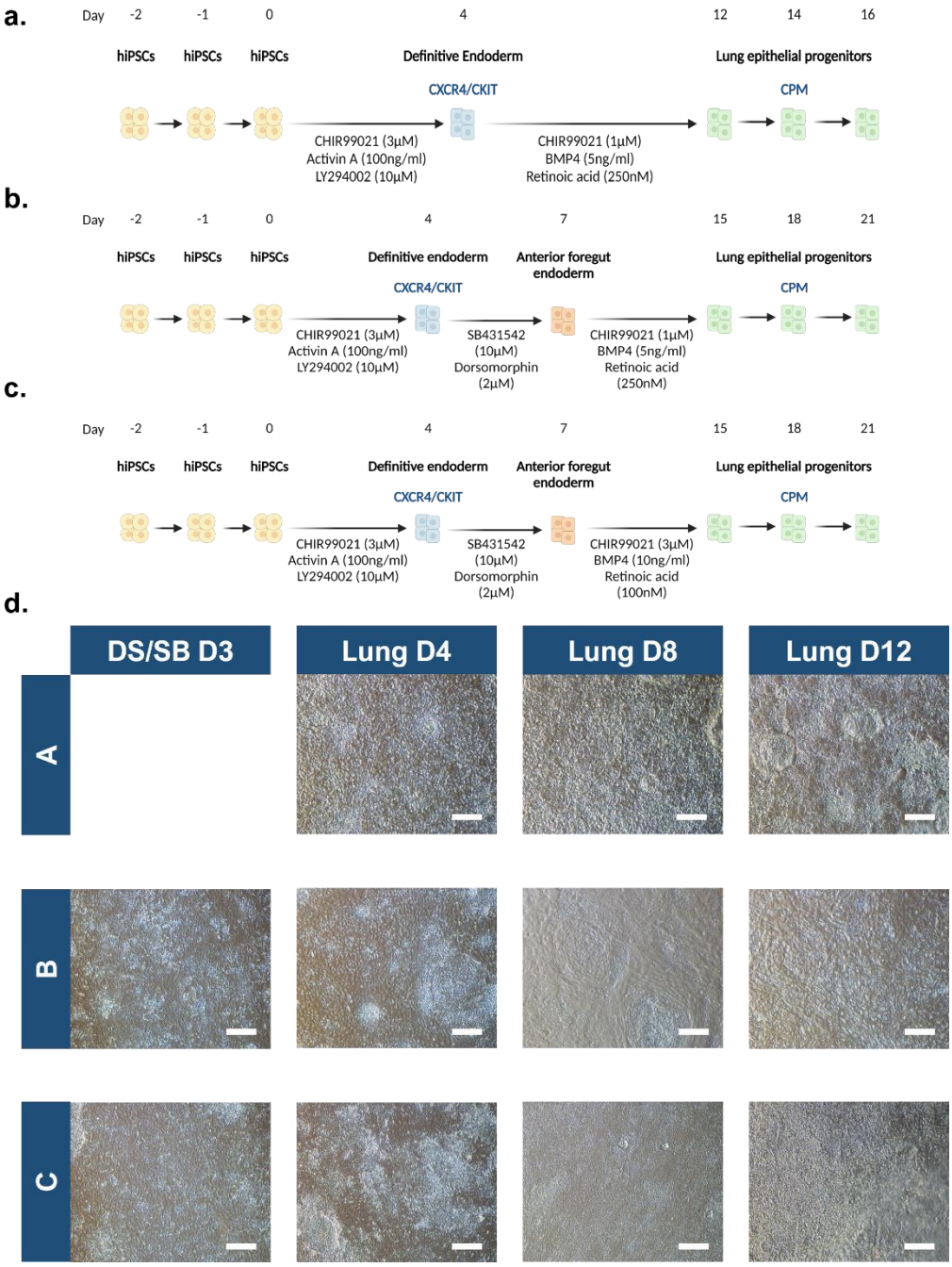


Figure 3-31. Optimisation of the differentiation of definitive endoderm to lung epithelial progenitor cells.

Schematic of the optimisation of the differentiation of hiPSCs to lung epithelial progenitor cells comparing three protocols **(a)** condition A, **(b)** condition B and **(c)** condition C with analysis of CPM expression on day 8, 10 and 12 of lung differentiation. **(d)** Light microscope images of BE.32 ($SFTPC^{WT/WT}$) cells at stages of lung epithelial progenitor cell differentiation for each condition. Scale bars, 200 µm.

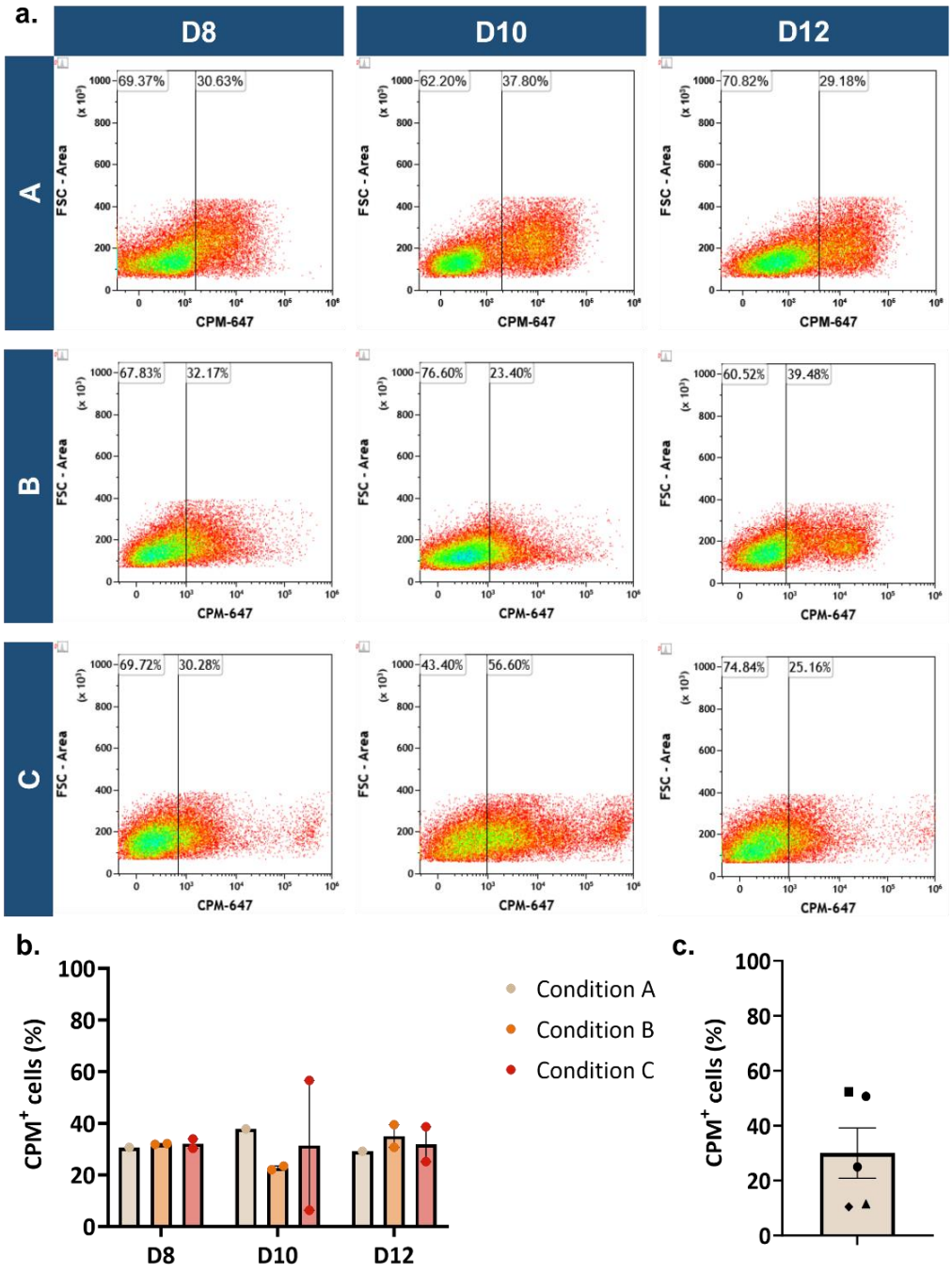

a.
D8
D10
D12
A
B
C
69.37%
30.63%
62.20%
37.80%
70.82%
29.18%
67.83%
32.17%
76.60%
23.40%
60.52%
39.48%
69.72%
30.28%
43.40%
56.60%
74.84%
25.16%
FSC - Area
CPM-647
b.
CPM+ cells (%)
Condition A
Condition B
Condition C
D8
D10
D12
c.
CPM+ cells (%)

Figure 3-32. Each lung progenitor cell differentiation protocol induces CPM-positive cells.

(a) Flow cytometry analysis of CPM on day 8, 10 and 12 of lung differentiation following specification in conditions A – C (**fig. 3-31a-c**). Flow cytometry gating was determined by secondary antibody only staining for each condition and timepoint as a negative control to determine non-specific binding of the secondary antibody that can change in heterogeneous populations of hiPSC-derived populations. **(b)** quantification of the percentage of CPM-positive cells for BE.32 (SFTPCWT/WT) cells. Bars represent mean ± SEM (individual symbols represent biological repeats; condition A (N=1), condition B and C (N=2)). **(c)** Quantification of the percentage of CPM-positive cells following differentiation in condition A with bars representing mean ± SEM (N=5; Triangle = ABE.27 ($SFTPC^{MUT/MUT}$); Diamond = IPF1.12 ($SFTPC^{WT/MUT}$); Circle = BE.32 ($SFTPC^{WT/WT}$); Square = BE.31 ($SFTPC^{WT/WT}$).

3.2.14. CPM-positive cells express markers of non-lung endoderm

In order to confirm expression of lung-related markers in our CPM+ population, we established a strategy to FACS purify lung day 8 CPM+ and CPM- cell fractions (**Fig. 3-33a**). We observed the lack of expression of lung markers *NKX2.1*, *SFTPC*, *ABCA3*, *SLC34A2*, *SFTPB* in CPM+ cells, but there was upregulation of markers related to early (*HHEX*, HNF1) and late (*AFP*, *TTR*) hepatocyte development, although genes involved in pancreatic cell differentiation (*PDX1*, *NGN3*) and early pharyngeal endoderm specification (*PAX9*) were absent (**Fig. 3-33b**). This strongly indicated that under our differentiation conditions CPM+ cells do not enrich for NKX2.1+ lung progenitor cells, but possibly isolate cells of a hepatocyte fate. Interestingly, despite lack of *NKX2.1* gene expression, we did observe punctate NKX2.1 protein staining in both FACS isolated CPM- and CPM+ cells seeded into Matrigel™-coated plates (**Fig. 3-34a**). This might indicate very low NKX2.1 expression in cells, but the immunofluorescent staining would need to be repeated to validate these findings.

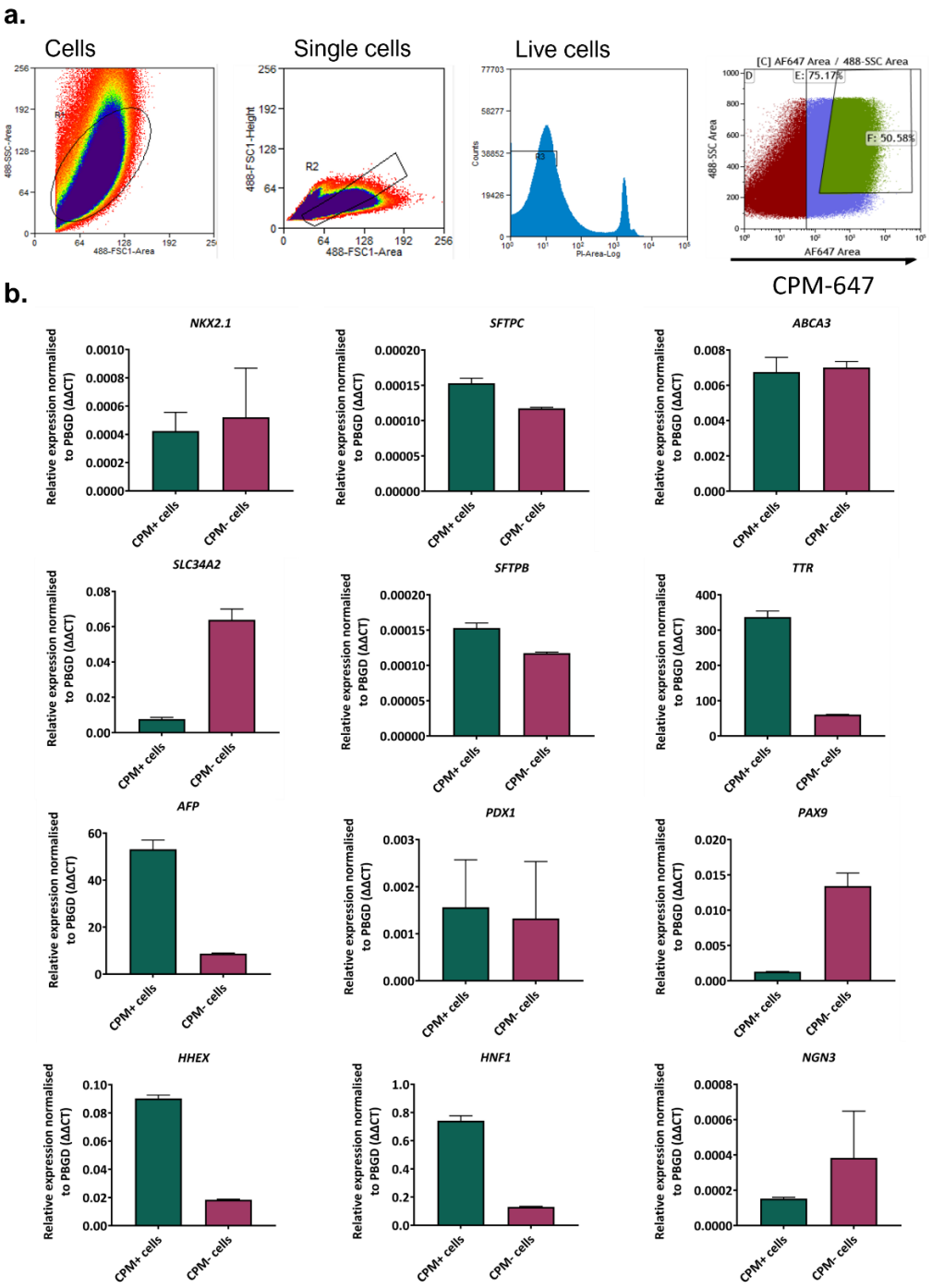


Figure 3-33. BE.32 (SFTPC$^{WT/WT}$) CPM-positive cells express markers of non-lung endoderm.

(a) Representative FACS plots showing the gating strategy used to isolate CPM+ cells on lung D8. **(b)** Expression of genes of lung and non-lung endoderm in CPM-positive and -negative sorted populations via RT-qPCR. Bars represent normalised expression to the housekeeping gene PBGD (ΔΔCt) ± SEM (N=1, 3 technical replicates).

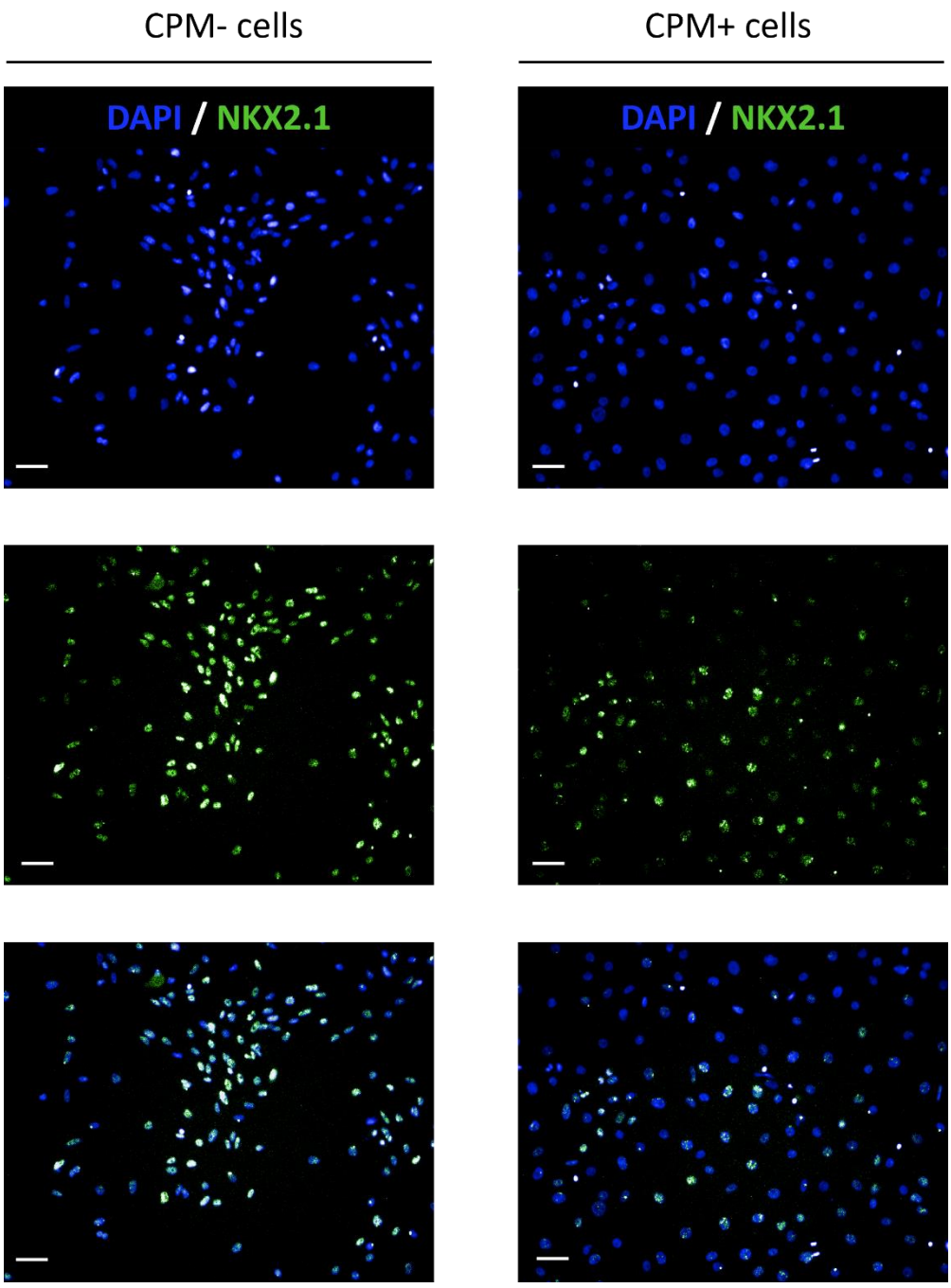


Figure 3-34. Sorted CPM cells express NKX2.1.

Immunofluorescent images of BE.31 ($SFTPC^{WT/WT}$) sorted CPM+ and CPM- cells seeded onto 2D Matrigel™-coated plates expressing NKX2.1. Scale bars, 50 µm.

3.2.15. Patterning of anterior foregut endoderm using dual-SMAD inhibition increases expression of lung progenitor markers

To understand why CPM+ cells with a hepatocyte-like gene signature were being generated in our differentiation, we returned to previous optimisation conditions and assessed the effect of dual-SMAD inhibition on the induction of lung markers. This was because previous findings showed dual-SMAD inhibition could pattern DE cells and enable response to signalling cues to differentiate to the lung lineage (Green et al., 2011). As we observed no increase in the number of CPM+ cells beyond day 10 of lung differentiation, we analysed gene expression on day 8, 9 and 10 across the 3 conditions used above (**Fig. 3-31a**). We observed that inhibition of TGF-β and BMP signalling (dual-SMAD inhibition) for 3 days (**condition B and C**) (**Fig. 3-31a**) prior to lung differentiation induced upregulation of lung genes including *NKX2.1*, *SOX9*, and *ABCA3*, which were expressed highest in condition B, and *SFTPC* with highest expression in condition C. The of peak expression varied across all 3 days for each gene. Without dual-SMAD inhibition, expression of these markers was lost (**Fig. 3-35a**). In contrast, there was a strong hepatocyte gene signature (*AFP*, *ALB*, *TTR*) in cells differentiated in condition A that increased progressively from day 8 to 10, but expression of hepatocyte genes was diminished with dual-SMAD signalling inhibition (**Fig. 3-35a**). This indicated the patterning of definitive endoderm to an anterior foregut fate via inhibition of TGF-β and BMP signalling is a critical step to permit cells to differentiate to lung and limit the capacity to upregulate genes related to hepatocyte development. We opted to use condition C as our optimised protocol going forward given the increased expression of lung progenitor markers but also because *SFTPC* expression was strongest.

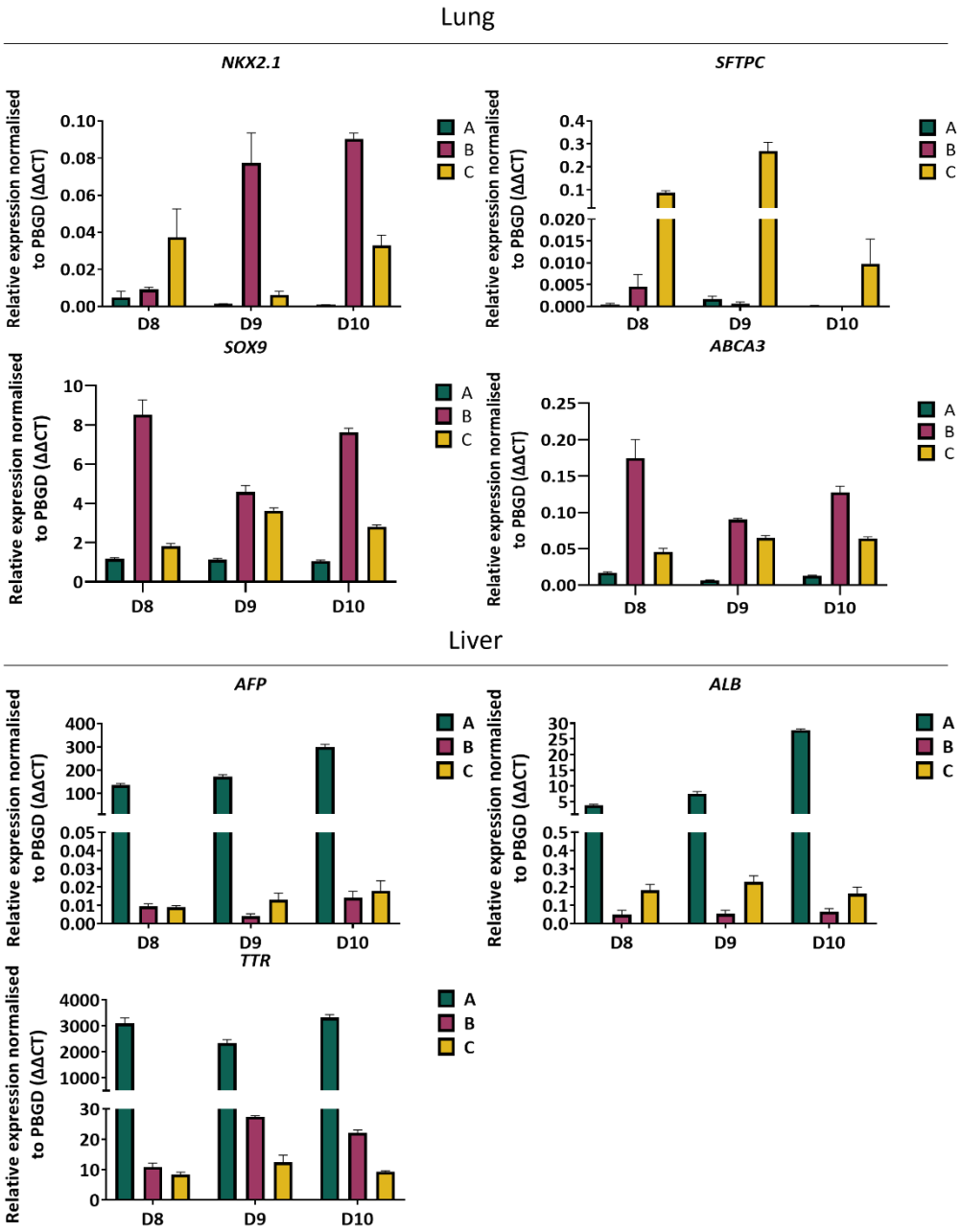


Figure 3-35. Anteriorisation of endoderm via dual-SMAD inhibition preceding lung progenitor differentiation promotes induction of lung markers.

RT-qPCR analysis of lung and liver gene markers on day 8, 9 and 10 of lung differentiation comparing protocols A, B and C (**Fig. 3-31a-c**) for BE.32 ($SFTPC^{WT/WT}$) hiPSCs. Bars represent normalised expression to the housekeeping gene PBGD (ΔΔCt) ± SEM (N=1, 3 technical replicates).

3.2.16. Splitting definitive endoderm cells leads to emergence of a high CPM-expressing cell population

During the differentiation of lung cells, we observed an over-proliferation of cells that led to formation of 3D aggregates across the cell monolayer. Based on previous findings showing that over-growth of cells can inhibit differentiation kinetics, we assessed the impact of splitting definitive endoderm cells prior to dual-SMAD inhibition and subsequent lung specification and CPM expression via flow cytometry. We split endoderm cells at four increasing ratios of the number of wells of day 4 endoderm cells of a 12-well plate to the number of wells of a 12-well plate that endoderm cells were seeded into; 1:1, 1:1.5, 1:2 and 1:3. Low splitting ratios led to over-proliferation with cell accumulation across the monolayer (**Fig. 3-36a; black arrows**), the same as was observed in cultures without splitting DE, whereas higher ratios provided space for cell expansion and by day 8 of differentiation, large cells were observed (**Fig. 3-36a**). A clear effect was observed; increasing the splitting ratio led to a greater number of CPM+ cells but also changed the pattern of CPM expression, causing the emergence of a CPM 'high' cell population via flow cytometry analysis, leading to an alternative gating strategy to discriminate between CPM-, CPM+ and CPM(hi) cells (**Fig. 3-36b**). This effect plateaued at a ratio of 1:2 where numbers of CPM+ and CPM(hi) cells were greatest. For cells that were equivalent to not being split (1:1), the percentage of CPM+ cells were <10% and CPM(hi) cells <1% (**Fig. 3-36b, c**).

We examined the impact of the splitting ratio on the expression of lung-related genes. We showed that there is a positive correlation between the splitting ratio (up to 1:2) and lung differentiation day, and the upregulation of the lung progenitor markers *NKX2.1* and *CPM*. The expression of these genes progressively increased and peaked on lung day 8 for the 1:2 ratio, despite lack of *SFTPC* expression across all conditions and days (**Fig. 3-37a**). Although cells split at higher ratios showed upregulation of the hepatocyte-specific markers *ALB* and *AFP*, expression was low compared to previous analyses of cells differentiated without splitting endoderm and dual-SMAD inhibition. However, *TTR* remained upregulated in these conditions across all days (**Fig. 3-37a**). These findings highlighted that preventing over-proliferation by splitting endoderm cells prior to anterior foregut patterning and lung differentiation significantly improves the induction of cells with a lung gene signature and the appearance of a high CPM-

expressing population. We next sought to understand the significance of these specific CPM populations.

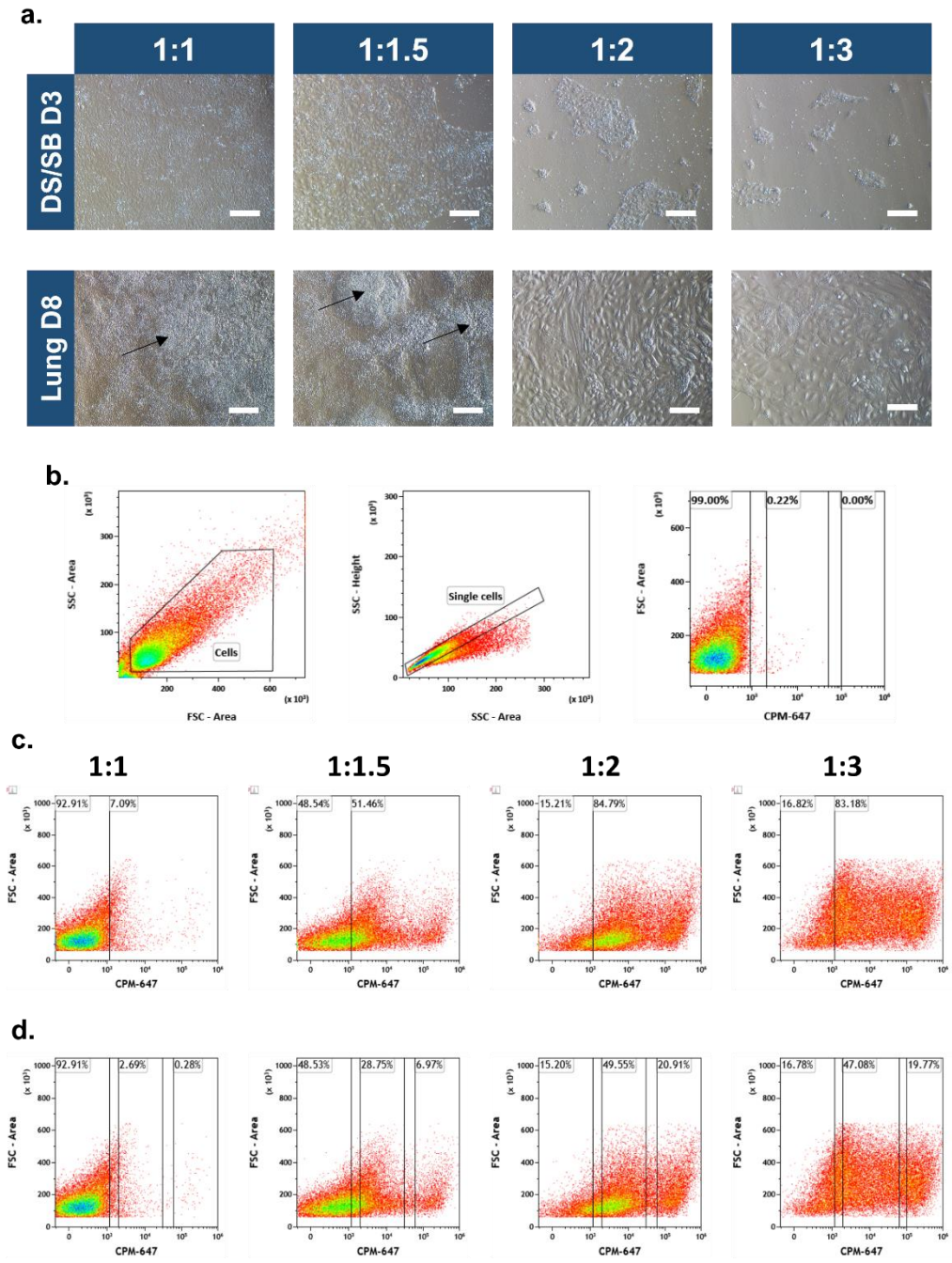


Figure 3-36. Splitting endoderm cells prior to lung differentiation induces three CPM populations.

(a) Flow cytometry plots and gating strategy for cells (left), single cells (middle) and secondary antibody only negative control (right). **(b)** Light microscope images of BE.31 ($SFTPC^{WT/WT}$) cells on day 3 of anteriorisation with DS (dorsomorphin; BMP inhibitor) and SB (SB431542; TGFβ inhibitor) (top panel) and day 8 of lung differentiation (bottom panel) after splitting endoderm cells at increasing ratios for further differentiation. Scale bars, 200 µm. Arrows indicate areas of cell over-proliferation and accumulation. **(c)** Flow cytometry analysis on day 8 of lung specification of CPM expression after splitting endoderm cells at ratios of 1:1, 1:1.5, 1:2 and 1:3 prior to further lung differentiation. **(d)** Alternative flow cytometry strategy to gate for CPM-negative, CPM-positive and CPM-high cell populations after splitting endoderm cells (N=1).

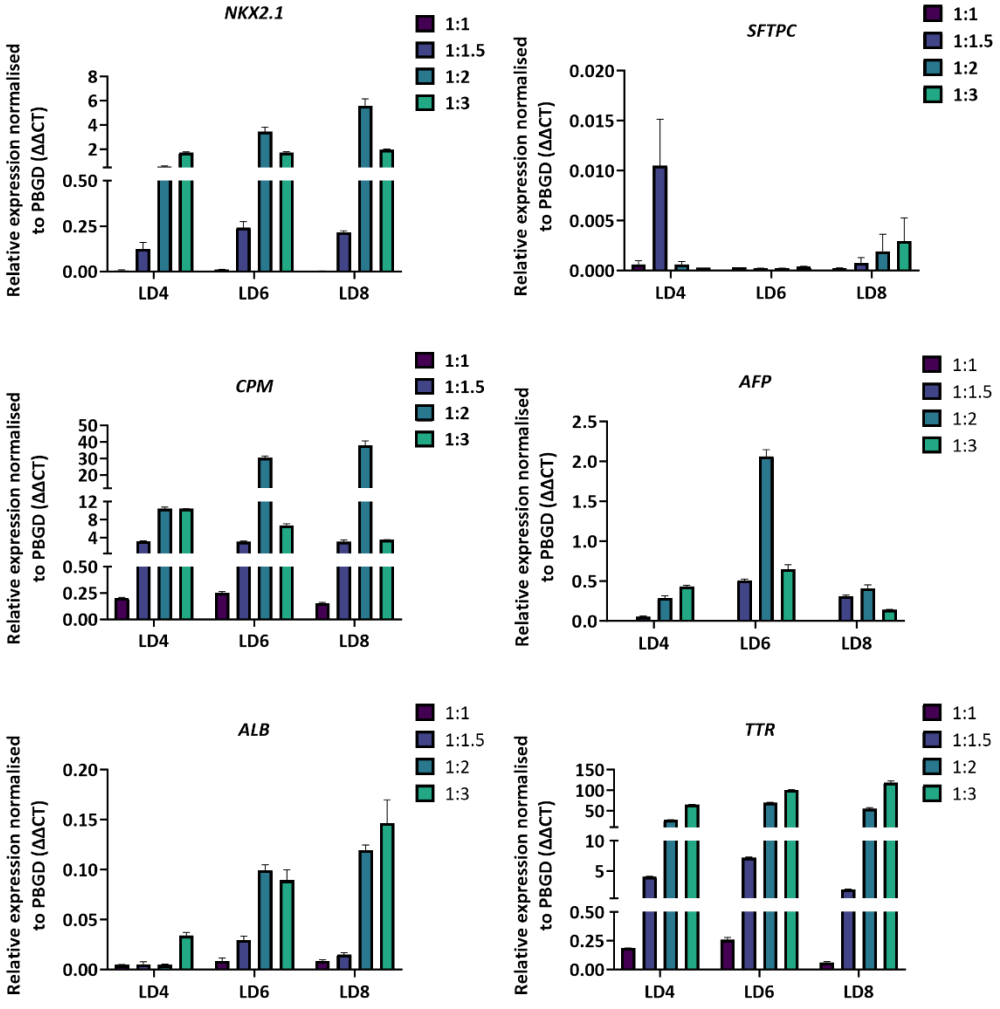


Figure 3-37. Splitting endoderm cells before lung differentiation upregulates lung progenitor specific markers.

(a) Expression of lung (*NKX2.1*, *SFTPC*, *CPM*) and liver (*AFP*, *ALB*, *TTR*) markers on lung differentiation day 4, 6 and 8 after splitting BE.31 ($SFTPC^{WT/WT}$) DE cells at 4 different ratios. Bars represent mean normalised expression to the housekeeping gene PBGD ($\Delta\Delta Ct$) ± SEM (N=1, 3 technical replicates).

3.2.17. CPM 'high' cells are enriched for markers of lung epithelial progenitors

After observing the emergence of three distinct CPM cell populations upon splitting of endoderm cells, we wanted to determine which epithelial cell types were enriched in each CPM fraction. We established a gating strategy for FACS to separate cells into CPM-, CPM+ and CPM(hi) fractions (**Fig. 3-38a**). In FACS isolated CPM(hi) cells, we observed upregulation of the lung progenitor genes *NKX2.1*, *CPM*, *SFTPC* compared to CPM- and CPM+ fractions. In addition, *AFP* and *ALB* expression was diminished in CPM(hi) cells and although upregulated in the CPM+ and CPM- cells, it remained low. Interestingly, *TTR* was expressed in CPM(hi) cells but not in other populations, but the significance of this was not investigated (**Fig. 3-38b**). This clearly demonstrated that the lung competent epithelial progenitor cells we wanted to generate were enriched in a population expressing CPM at high levels as determined by flow cytometry. To isolate these cells and limit contamination from non-lung endodermal cells, we decided to adopt this sorting strategy as part of our optimised differentiation protocol going forward. To improve cell survival after splitting, we used CloneR™2 in our cultures. We observed a large degree of biological variability between differentiations, and the percentage of CPM(hi) cells isolated was 7.06 ± 4.29% (mean ± SEM) (N=3) for BE.31 ($SFTPC^{WT/WT}$) hiPSCs. As a proof of principle in a second cell line we showed ABE.27 ($SFTPC^{MUT/MUT}$) hiPSCs could also generate CPM(hi) cells but with similar low efficiency, although further replicates are required for comparison to BE.31 ($SFTPC^{WT/WT}$) hiPSC differentiation (**Fig. 3-39a**). Nonetheless, we demonstrated robust protein expression of NKX2.1 and CPM in day 8 ABE.27 ($SFTPC^{MUT/MUT}$) lung cells prior to FACS (**Fig. 3-39b**). These findings suggest NKX2.1 expression is strongly induced

during lung progenitor specification, but also indicates CPM might not detect all NKX2.1+ cells in the culture.

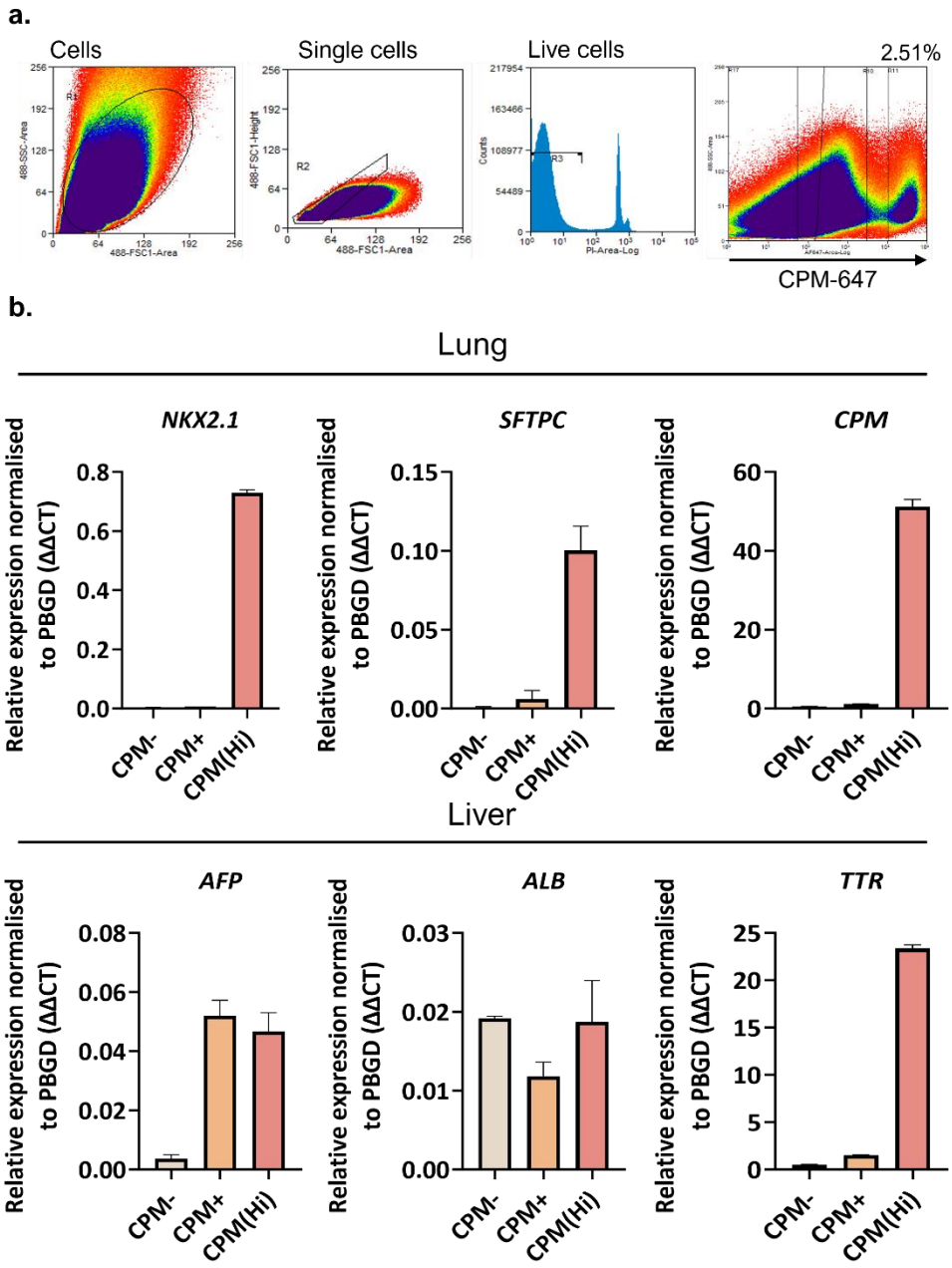


Figure 3-38. BE.31 (SFTPC$^{WT/WT}$) CPM-high expressing cells are enriched for lung epithelial progenitor cell markers.

(a) Representative FACS plots showing the gating strategy used to isolate CPM(hi) cells from D8 of lung differentiation. **(b)** qRT-PCR analysis of lung and liver gene markers in cell populations sorted via CPM expression with bars representing mean normalised expression to the housekeeping gene PBGD (ΔΔCt) ± SEM (N=1, 3 technical replicates).

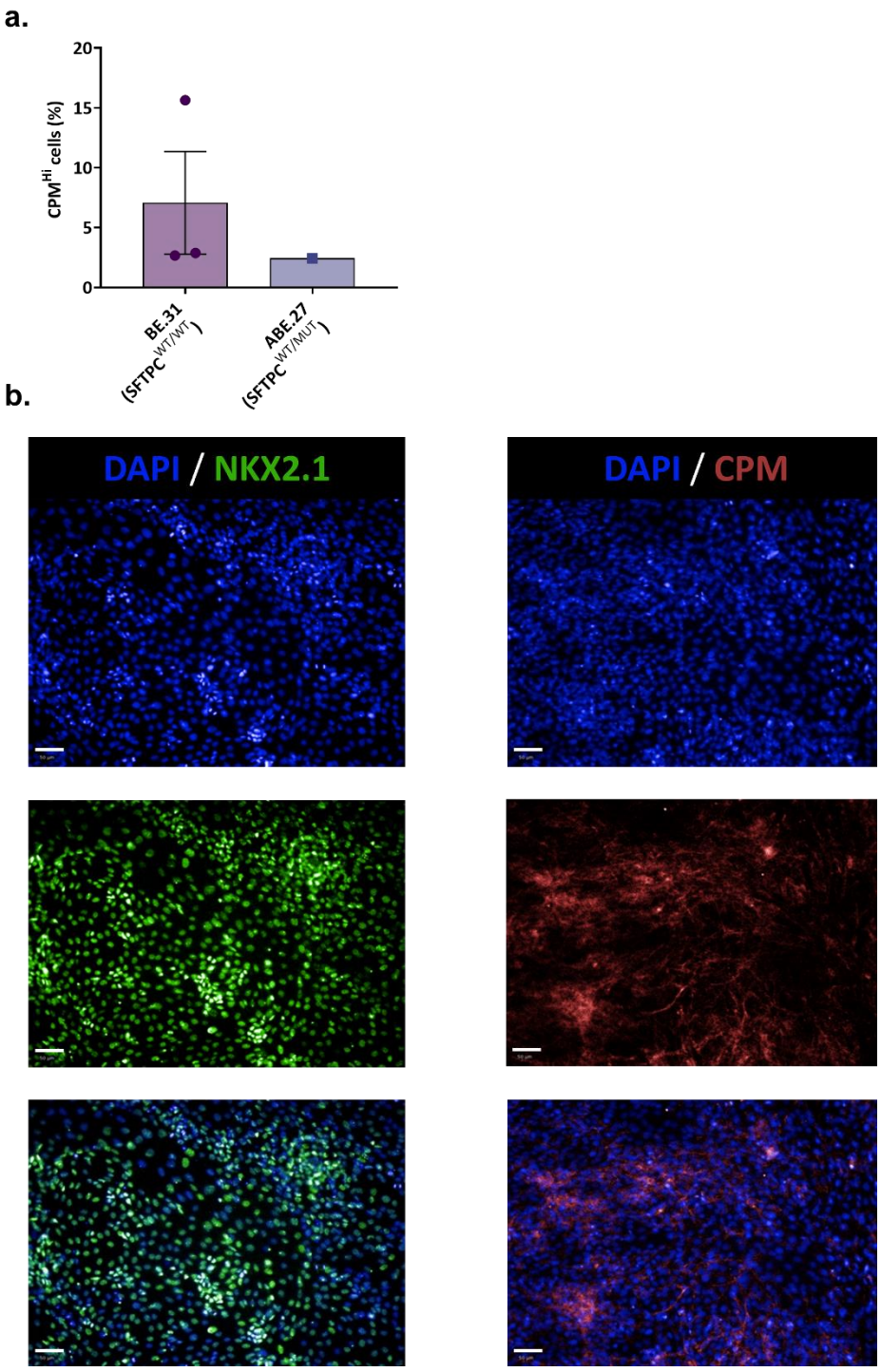


Figure 3-39. Cells on day 8 of lung differentiation express lung progenitor cell markers.

(a) Quantification of CPM(Hi) cells by flow cytometry with an optimised differentiation protocol (BE.31; N=3, mean ± SEM. ABE.27; N=1). **(b)** Immunofluorescent images of ABE.27 ($SFTPC^{MUT/MUT}$) cells on day 8 of lung differentiation expressing NKX2.1 and CPM. Scale bars, 200 µm.

3.2.18. Robust induction of CPM+ cells after splitting following dual-SMAD inhibition

During development DE cells are multipotent and give rise to many epithelial cell types of the respiratory and gastrointestinal epithelium. We hypothesised that restricting the potential of DE to anterior foregut fate, where lung progenitors emerge, and then splitting cells might improve our differentiation efficiency. After 3 days of SB431542 and dorsomorphin treatment, we could dissociate cultures to single cells and seed them into lung specification medium. At day 6 of lung differentiation, a monolayer of large cells was observed (**Fig. 3-40a**). Flow cytometry analysis revealed the induction of a significant percentage of CPM(Hi) cells which increased with time in culture (**Fig. 3-40b**). On day 8 the percentage of CPM(hi) cells was 84.5% (**Fig. 3-40c**). In lung day 7 cultures, lung progenitor genes *NKX2.1* and *CPM* were upregulated and the type 1 AECs markers *AGER* and *HOPX* were also expressed. Expression of *TUBA1A* which encodes for acetylated α-tubulin found in ciliated cells was significantly upregulated and the intestinal marker *VILLIN* was expressed. Genes related to airway cells (*KRT5*, *TP63*, *SCGB3A2*), hepatocyte development (*ALB*, *AFP*) and type 2 AECs (SFTPC, SFTPB) were not expressed (**Fig. 3-40d**). However, attempts to repeat these findings was unsuccessful. The dissociation of anterior foregut to single cells was difficult, perhaps owing to continued deposition of ECM from cells during culture. There is scope to revisit this work to improve the efficiency of lung progenitor cell specification.

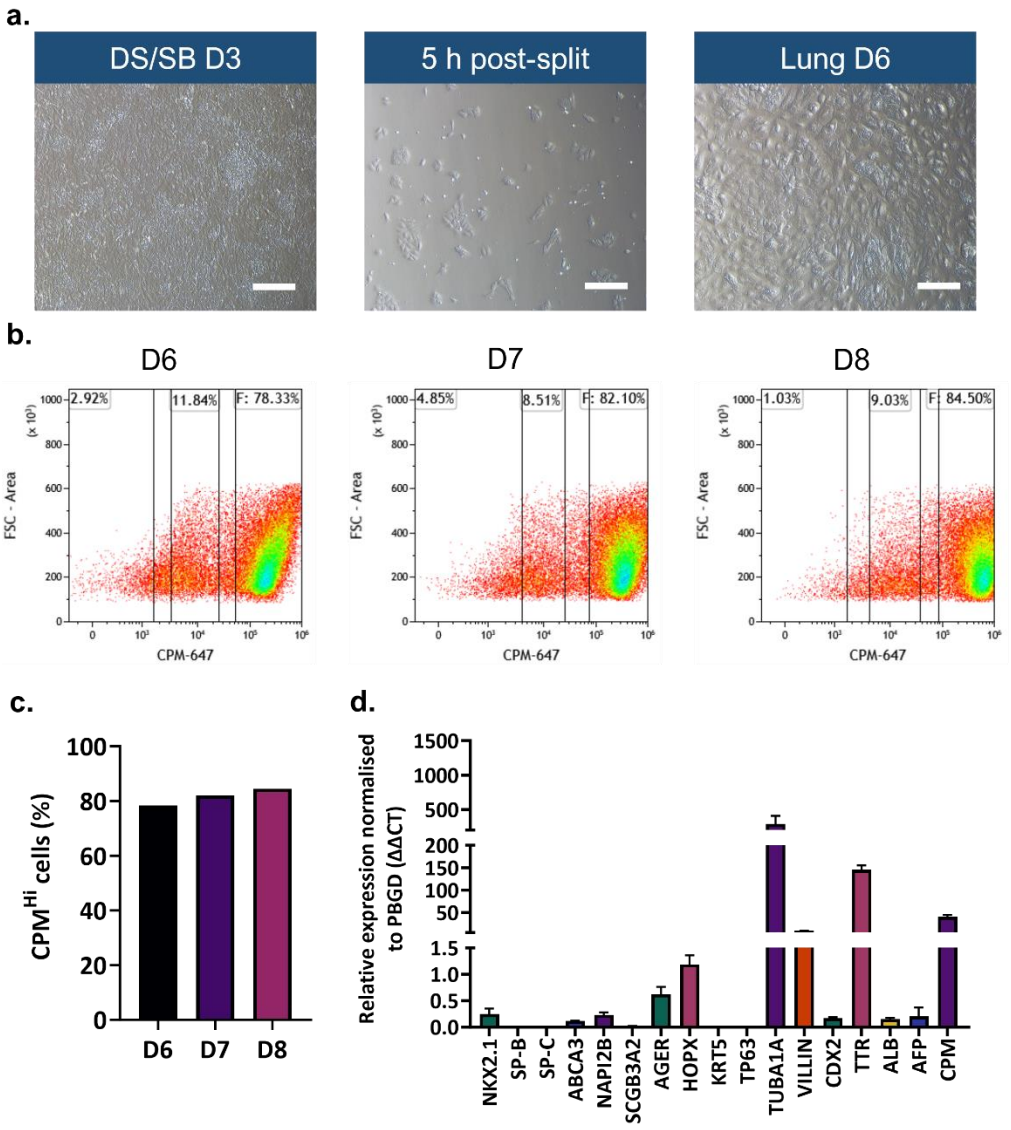


Figure 3-40. Splitting cells after anteriorisation of foregut induces a high percentage of CPM(hi) cells in BE.31 (SFTPC$^{WT/WT}$) hiPSCs.

(a) Light microscope images of BE.31 (SFTPC$^{WT/WT}$) cells split after treatment with DS/SB before differentiation to lung progenitor cells. Scale bars, 200 µm. **(b)** Flow cytometry analysis of CPM expression on sequential days of lung differentiation and **(c)** quantification of CPM(hi) cells (N=1). **(d)** qRT-PCR analysis of lung and non-lung endoderm genes on day 7 of differentiation after splitting cells after anteriorisation with bars representing normalised expression to the housekeeping gene PBGD (ΔΔCt) ± SEM (N=1, 3 technical replicates).

3.2.19. Development of a strategy to generate and isolate hiPSC-type 2 AECs

To determine whether our isolated lung progenitor cells could give rise to type 2 AECs and whether they could be enriched based on NaPi2B expression. We cultured single CPM(hi) cells in 3D Matrigel™ in the presence of lung maturation factors as previously stated. We could maintain organoids for several passages, and they retained the capacity to form 3D structures in culture. We established a FACS strategy to isolate cells co-expressing NaPi2B and the epithelial cell marker EPCAM but efficiency remained low (2.66% of the total population) (**Fig. 3-41a**). Sorted NaPi2B/EPCAM+ cells could be seeded into Matrigel droplets and self-assembled into 3D cystic organoids with a thin epithelium.

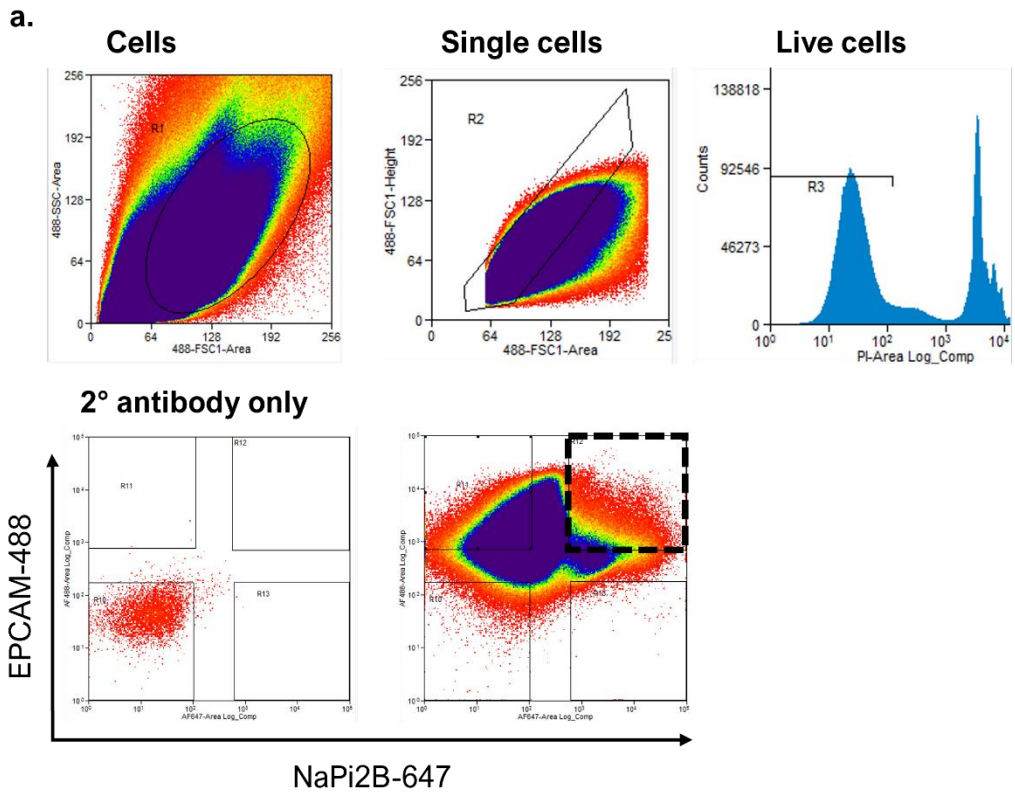


Figure 3-41. Strategy for enriching NaPi2B/EPCAM double-positive cells.

(a) FACS plots showing the gating strategy for isolating NaPi2B+/EPCAM+ cells after 5 passages in 3D culture (dotted square). **(b)** Representative light microscope images of ABE.27 (SFTPC$^{MUT/MUT}$) single NAPI2B+/EPCAM+ cells seeded at a density of 400 cells/µl Matrigel™ self-assembling into 3D organoids. Scale bars, 200 µm.

3.2.20. NaPi2B+ cells and organoids express markers of type 2 AECs

To confirm the gene signature of the NaPi2B/EPCAM populations we performed gene expression analysis in FACS isolated populations. In NaPi2B-/EPCAM+ cells hepatocyte (*ALB*, *AFP*, *TTR*) and hindgut/intestinal (*VILLIN*, *CDX2*) markers were upregulated, but

lung-related genes were not expressed (**Fig. 3-42a**). Before processing organoids, the lung genes *NKX2.1* and *SFTPC* were upregulated, but their expression was enriched in NaPi2B+/EPCAM+ cells after FACS but diminished in NaPi2B-/EPCAM+ populations. In contrast, the hindgut/intestinal markers *VILLIN*, *LGR5* and *CDX2* were upregulated before FACS, and expression was maintained in NaPi2B-/EPCAM+ cells but reduced in double-positive cells after enrichment (**Fig. 3-42b**). These findings suggest the type 2 AEC-like cells can be enriched in a NaPi2B+/EPCAM+ population in our organoid culture system. We showed further that our 3D organoids express SFTPC and NaPi2B-isolated cells display punctate SFTPC staining (**Fig. 3-43a, b**). Further, NaPi2B+/EPCAM+ isolated cells generated from both BE.31 ($SFTPC^{WT/WT}$) and ABE.27 ($SFTPC^{MUT/MUT}$) hiPSCs express Pro-SFTPC, SFTPC and NaPi2B (**Fig. 3-44a, b**). This confirms the type 2 AEC phenotype of isolated cells and organoids.

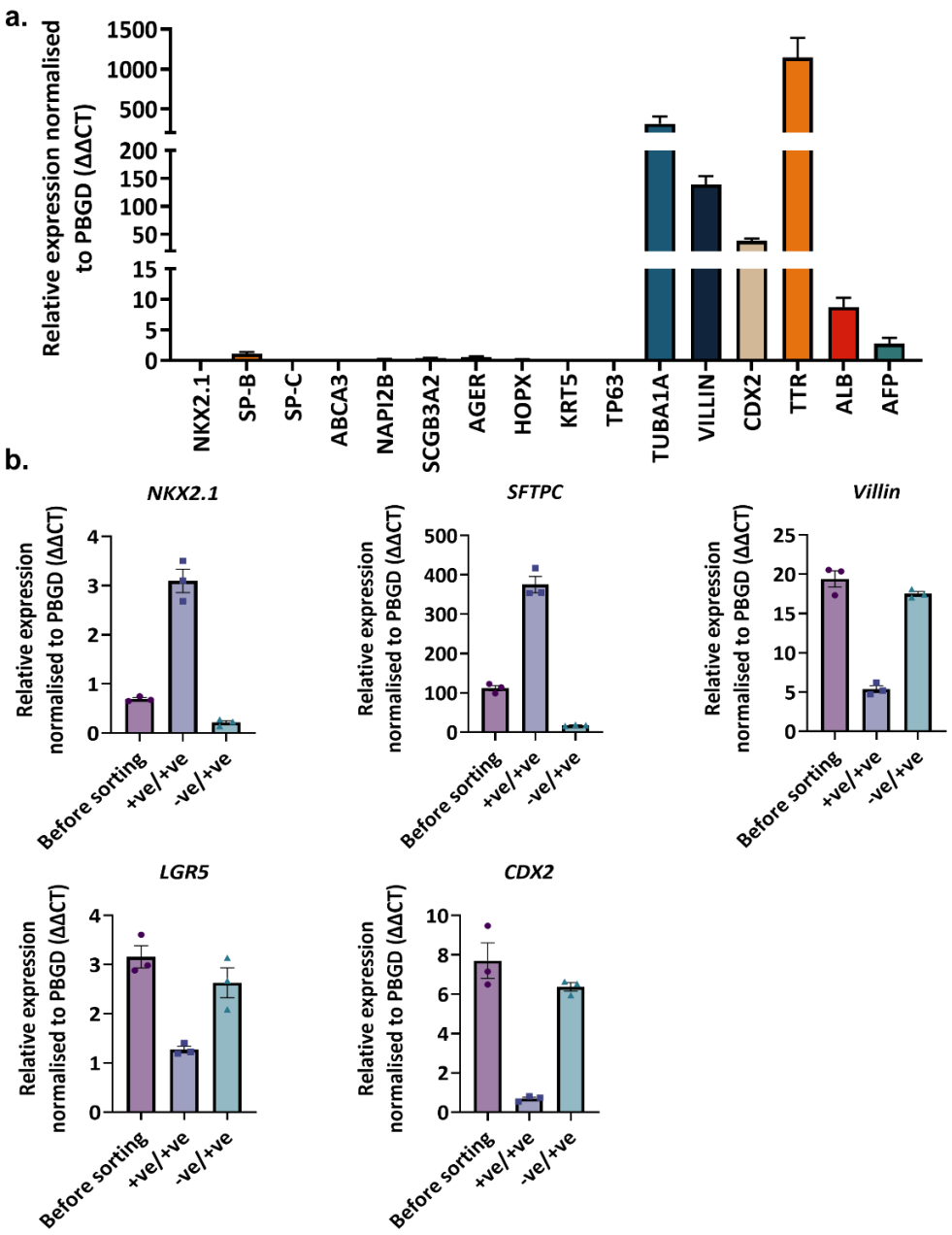


Figure 3-42. BE.31 (SFTPC$^{WT/WT}$) NAPI2B+/EPCAM+ sorted cells are enriched for type 2 AEC markers.

(a) qRT-PCR analysis of lung and non-lung endoderm markers in ABE.27 (SFTPC$^{MUT/MUT}$) NaPi2B-/EPCAM+ cells. Bars represent mean normalised expression to the housekeeping gene PBGD (ΔΔCt) ± SEM (N=1). **(b)** qRT-PCR analysis of type 2 AEC and non-lung endoderm genes in BE.31 (SFTPC$^{WT/WT}$) NAPI2B+/EPCAM+ (+ve/+ve), NAPI2B-/EPCAM+ (-ve/+ve) and pre-sorted cell populations with bars representing mean normalised expression to the housekeeping gene PBGD (ΔΔCt) ± SEM (N=1, 3 technical replicates). (Figure 42b was performed in collaboration with Rizal Rizal as previously described).

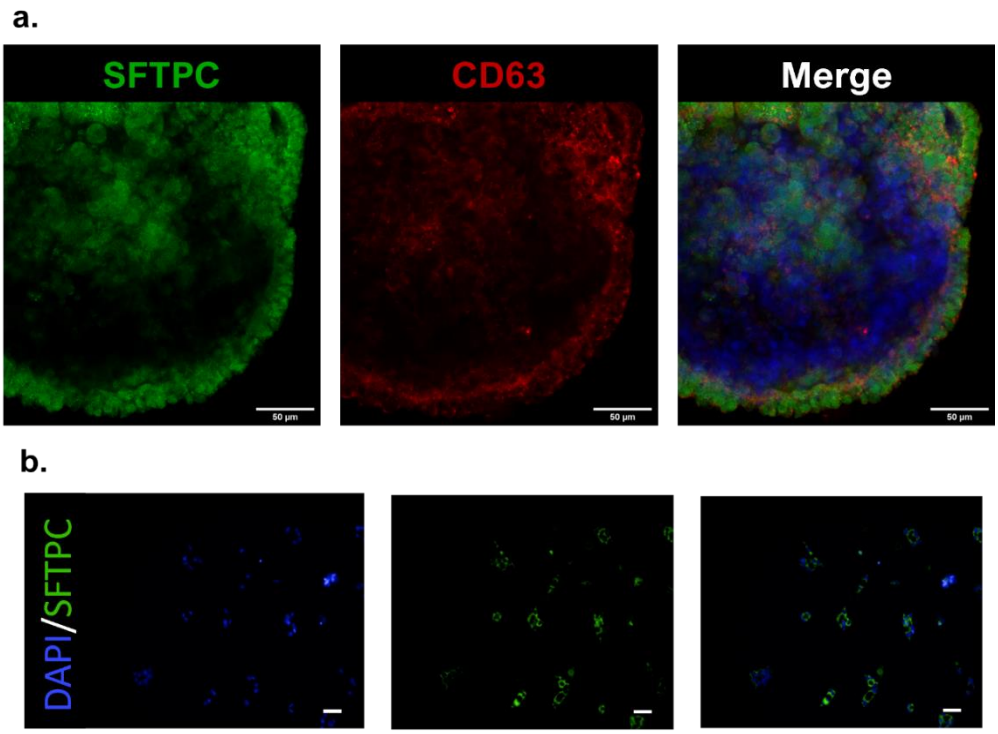


Figure 3-43. NaPi2B-enriched cells and organoids express markers of type 2 AECs.

(a) Immunofluorescence images of IPF1.12 3D organoids expressing SFTPC and CD63 after enrichment via NaPi2B expression. Scale bars, 50 µm. **(b)** Immunofluorescence images of sorted NaPi2B+ cells plated on 2D Matrigel™-coated plates expressing SFTPC. Scale bars, 50 µm. (Figure 43 was performed in collaboration with Dr Analy Serna as previously described).

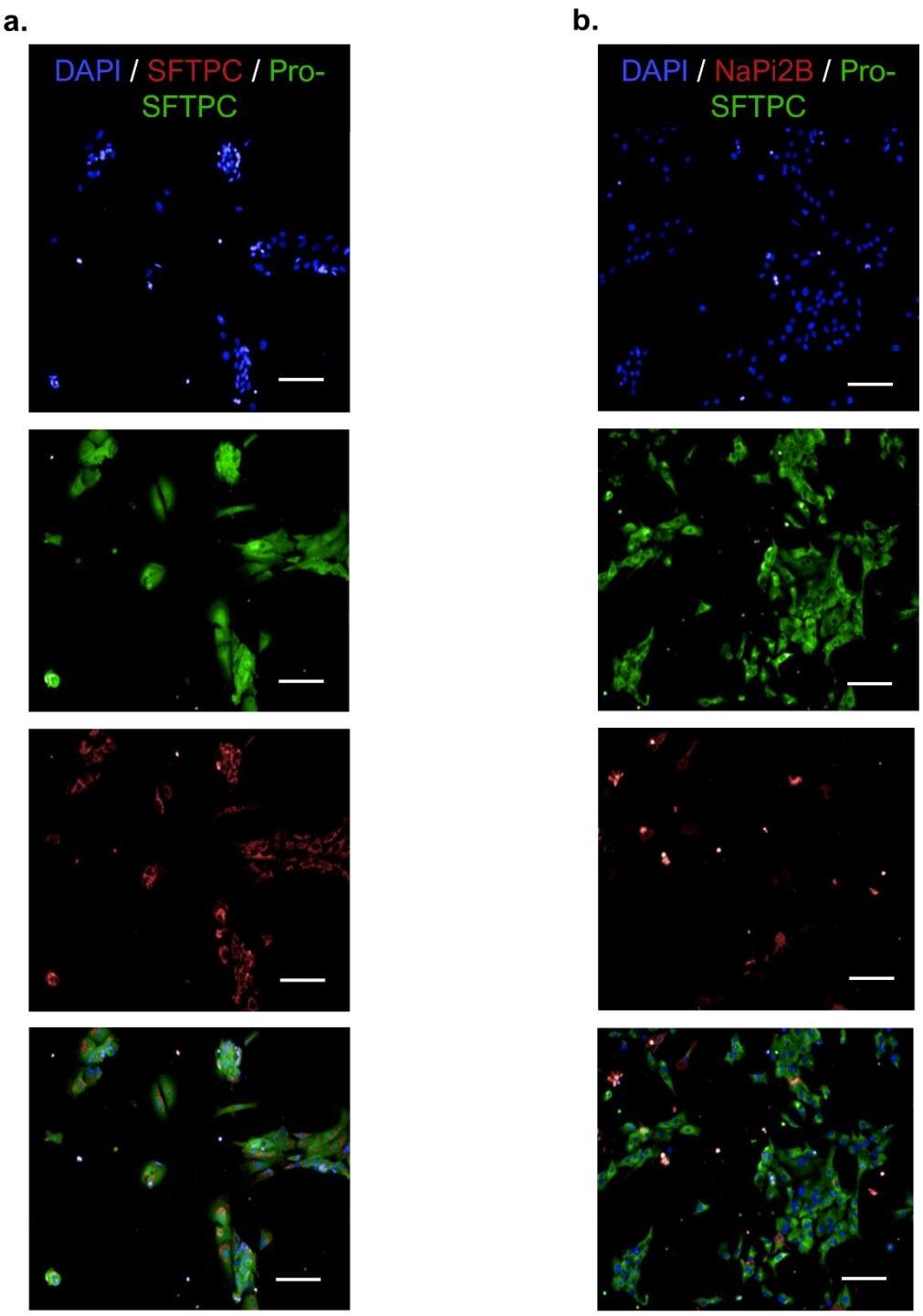


Figure 3-44. Cells enriched for NaPi2B express markers of type 2 AECs.

(a) Immunofluorescence images of sorted ABE.27 ($SFTPC^{MUT/MUT}$) NaPi2B+ cells expressing SFTPC and pro-SFTPC. **(b)** Immunofluorescence images of sorted BE.31 ($SFTPC^{WT/WT}$) NaPi2B+ cells expressing Pro-SFTPC and NAPI2B. Scale bars, 50 µm. (Figure 44 was performed in collaboration with Dr Analy Serna, as previously described).

3.3. Discussion

In the current work we established patient-derived hiPSCs with a heterozygous mutation in the BRICHOS domain of the SFTPC gene and successfully employed base editing to genetically engineer isogenic control cell lines to provide a spectrum of genetic lung disease. We demonstrated these hiPSCs had normal karyotype and expressed the core pluripotency genes NANOG, OCT4 and SOX2. We established a platform to differentiate patient-derived hiPSCs to type 2 AEC organoids via induction of intermediate cell types to replicate normal embryonic lung development. We optimised a strategy using previously identified cell surface markers to isolate and purify NKX2.1+ lung epithelial progenitors from contaminating populations of cells of a similar developmental origin. Enriched NKX2.1+ lung-competent cells could give rise to type 2 AECs in 3D epithelial organoids and were purified via surface expression of NaPi2B and maintained in culture long-term. Antibodies targeted to CPM and NaPi2B are commercially available enabling high reproducibility of this protocol. The production of alveolar cells from both normal and aberrant genetic backgrounds offers a novel opportunity to investigate mechanisms of lung disease.

We first attempted to differentiate hiPSCs to lung progenitor cells using a previously established protocol and showed the activation of WNT and activin A signalling led to upregulation of DE genes, but protein expression of DE markers was low. Following lung specification, we observed a lack of expression of lung progenitor-specific genes and enrichment via CPM and NaPi2B failed. When aggregates of differentiated cells were transferred into 3D culture, they demonstrated a robust ability to self-assemble into 3D organoids, but unsurprisingly, after further differentiation to type 2 AECs with a maturation cocktail of growth factors, there was a lack of expression of type 2 AEC markers. These findings led to us revisit the differentiation approach and optimise each intermediate stage.

Reports on the efficiency of DE specification use a variety of methods and markers, such as immunostaining and/or flow cytometry of SOX17, FOXA2, CXCR4, CKIT, although these markers are not exclusive to embryonic endoderm (Holtzinger et al., 2015). Co-expression of CXCR4 and CKIT is commonly used to quantify and validate *in vitro* differentiation to DE cells, so we used this approach as a reproducible method to assess efficiency of DE specification (Cheng et al., 2012; Gouon-Evans et al., 2006).

Similarly, we used the transcription factors SOX17 and FOXA2 which are critical regulators of mouse endoderm development (Burtscher & Lickert, 2009; Kanai-Azuma et al., 2002) and established markers of *in vitro* derived human DE, to further validate induction of DE cells (Yiangou et al., 2018).

Our finding that the combination of WNT and activin A signalling is refractory to efficient DE differentiation and insufficient for repressing the pluripotency programme is consistent with previous reports (Teo et al., 2014). Although specification of DE has been reported with WNT3a and activin A, it has required the presence of low levels of serum, extended culture periods and high recombinant protein concentrations to achieve efficient differentiation (D'Amour et al., 2006; Teo et al., 2014; Toivonen et al., 2013). We observed that BMP4 could act in a synergistic manner to marginally promote DE induction and repress pluripotency genes during differentiation, but the efficiency of DE specification remained sub-optimal. This corroborates previous findings that BMP4 antagonises pluripotent gene expression, but in contrast to our findings, efficient DE differentiation has been reported in serum-free conditions with activin A and BMP4 alone. This could potentially be explained by the higher concentration of BMP4 used in these studies or variation in differentiation potential between pluripotent cell lines (Phillips et al., 2007; Teo et al., 2014; Teo et al., 2012). Further treatment of hiPSCs with FGF2 in our cultures showed enhanced DE differentiation and greater pluripotency gene suppression. Other reports have shown improved efficiency of DE specification with transient FGF2 activation alongside continued activin A signalling under serum-free conditions (Sui et al., 2012). These findings showed that FGF2-mediated activation of the downstream MAPK/ERK1/2 intracellular pathway acted in coordination with signalling through the SMAD2,3 effector of the TGFβ/Activin/Nodal pathway to robustly induce hPSC differentiation to DE (Brown et al., 2011; Sui et al., 2012; Teo et al., 2011). Nonetheless, under these conditions our differentiation efficiency was sub-optimal and not comparable to other reported DE efficiencies.

We demonstrated that the treatment of hiPSCs with LY294002, an inhibitor of PI3K signalling, in combination with WNT, BMP and activin A significantly improved induction of DE cells which was enhanced by treatment with FGF2. Insulin and insulin-like growth factor 1 (IGF1) signalling through the PI3K/AKT pathway is important for

self-renewal and maintenance of pluripotent stem cells but is refractory to DE development (McLean et al., 2007; Xu et al., 2011). Therefore, several methods for differentiating hiPSCs to DE include LY294002 treatment in combination with aforementioned combinations of growth factor proteins (McLean et al., 2007; Teo et al., 2011; Touboul et al., 2010; Yu et al., 2015). In medium lacking insulin it has been shown that inhibition of PI3K signalling through LY294002 is not necessary for efficient differentiation to DE, but in this particular study DE cells were not quantified and conclusions were established using gene expression analysis only (Xu et al., 2011). In contrast, it has been reported that LY294002 is needed in combination with WNT3a, Activin A, BMP4 and FGF2 signalling for efficient DE differentiation in serum-free conditions, as LY294002 with Activin A alone was insufficient to supress expression of pluripotency genes. This supports our findings that LY294002 is critical for robust onset of the DE programme and further suggests BMP and FGF signals are important to enable cells to initiate differentiation in response to activin A (Touboul et al., 2010).

The mechanisms governing DE differentiation have been further elucidated with studies into the downstream intracellular signalling pathways. When PI3K/AKT activity is diminished, ERK and effectors of WNT signalling, such as β-catenin activation and repression of GSK3β, are stimulated and synergise with activin A induced SMAD2,3 signalling to target lineage specific genes and promote differentiation (Singh et al., 2012). Further to this, signalling through PI3K and the downstream protein complex mTORC2 phosphorylates a linker of SMAD2/3 subsequently leading to its degradation. Inhibition of PI3K signalling prevents active SMAD2,3 degradation and permits activin-dependent differentiation to DE (Yu et al., 2015). Recently, Lau et al. (2023) described the involvement of signalling through ERK1/2 in DE induction. They demonstrated that in the presence of WNT and activin A ligands and PI3K inhibition, FGF2-mediated signalling through FGFR2 could activate ERK1/2 and could drive the generation of SOX17+ DE cells with high efficiency. When FGFR and ERK1/2 activity was inhibited, activin A alone was insufficient for DE specification. This has led to a proposed model whereby activin A-SMAD2 and ERK1/2 activation drive initial mesendodermal fate and inhibition of PI3K enhances DE differentiation by heightening activin A-induced SMAD2 signalling. Subsequent activation of ERK1/2 via FGF2-signalling through FGFR then enables SOX17+ DE commitment. Although we did not investigate intracellular signalling pathways in our study, these findings align with our results showing robust

differentiation of hiPSCs to DE with treatment of WNT3a, Activin A, BMP4, FGF2 and LY294002. However, PI3K inhibitors can impair cell viability during DE formation (Lau et al., 2023; Sekine et al., 2012), and have apoptotic effects (Will et al., 2014), so timing of LY294002 treatment is critical. We observed cell apoptosis with LY294002 in the presence of Wnt3a and activin A, but in all other conditions, brief treatment with LY294002 did not have an adverse effect on cell viability.

We observed the greatest efficiency of hiPSC differentiation to DE when cells were treated with the GSK3β inhibitor CHIR99021 for 1 day with PI3K inhibition for 2 days and continued activin A signalling. Stimulation of the Wnt pathway via GSK3β has been effective in specifying DE cells in serum-free conditions (Bone et al., 2011; Hannan, Segeritz, et al., 2013; Naujok et al., 2014) and other findings have indicated transient treatment with CHIR99021 in combination with BMP4 and activin A signalling is superior to recombinant WNT3a in elevating β-catenin levels, maximally driving DE differentiation and supressing pluripotency genes (Teo et al., 2014). This supports our results, but we demonstrated that BMP4 and FGF2 were dispensable for efficient DE differentiation which is in contrast to previous reports showing BMP4 and FGF2-ERK1/2 signalling is important for DE commitment (Lau et al., 2023). However, CHIR99021-mediated WNT signalling might activate downstream ERK effectors as cross-talk between these pathways is known (Bikkavilli & Malbon, 2009). Although it is beyond the scope of this work, further investigation would be necessary to confirm this. Nonetheless, we showed our optimised protocol was highly reproducible and our hiPSC-DE cells expressed SOX17 and FOXA2.

Following optimisation of DE differentiation we induced a foregut population with continued activin A signalling, and following treatment with factors for lung progenitor cell specification, we demonstrated the onset of cells expressing CPM but observed strong expression of genes of non-lung endoderm, in particular hepatic lineages, within this population. Self-renewing multipotent foregut cells have been generated with continued activin A signalling and can differentiate into cells expressing markers of multiple foregut lineages such as the liver, pancreas and lung/thyroid (Hannan, Fordham, et al., 2013). The emergence of contaminating cell types in our culture could be a result of similar signalling pathways required for differentiation of several lineages from a similar developmental origin such as foregut. Indeed, differentiation to

hepatocyte cells involves FGF2, BMP4 and hepatocyte growth factor (HGF) signalling, but FGF10 and RA signalling and TGF-β inhibition have also been used (Graffmann et al., 2022). This highlights important considerations when selecting cell surface markers to enrich for target populations. Although CPM has been used to purify NKX2.1+ lung progenitors, it is also expressed in human fetal liver and has been used to enrich hiPSC-derived bipotential liver progenitors with the capacity for hepatocyte differentiation (Gotoh et al., 2014; Kido et al., 2015). Other groups have reported contamination with cells of liver and intestinal lineages during lung differentiation even after enrichment steps (Hawkins et al., 2017; Hein et al., 2022). Trajectory analysis of scRNAseq datasets showed the co-emergence of liver and lung progenitor cells along distinct pathways during differentiation of hiPSCs to lung. They showed the liver progenitor program is determined as early as the DE stage and lung progenitor cells become established at later timepoints. These trajectories were completely independent and there was no evidence of a bipotent lung-liver progenitor (Ori et al., 2021). This aligns with our findings and suggests early liver progenitors could be determined during our differentiation to DE and continued WNT, BMP and RA signalling might promote differentiation and expansion of this population.

We showed inhibition of TGF-β and BMP signalling after DE formation and prior to lung differentiation was required to diminish the expression of hepatic genes and induce lung-related transcripts which aligns with other reports that showed this step was necessary to induce anterior foregut fate and permit lung induction (Green et al., 2011; Huang et al., 2014). This is similar to the mouse system where activin A induced DE is biased towards hindgut identity and is incapable of generating lung cells, but treatment with inhibitors of TGF-β and BMP signalling restricts foregut potential and permits lung specification, mimicking a short phase of mouse embryonic foregut patterning (Huang et al., 2014; Longmire et al., 2012; Mou et al., 2012). Other approaches to anteriorise DE have used FGF2 and sonic hedgehog (SHH) signalling to enable emergence of lung epithelial progenitor cells, but transcripts of thyroid, pharyngeal, liver and pancreatic lineages were detected (Wong et al., 2012). The timing of inhibition is an important parameter as it can significantly impact the capacity for cells to become lung progenitors, with 72 hours appearing optimal as we demonstrated (Hawkins et al., 2017).

One important finding from our work was the need to dissociate and re-plate endodermal cells prior to treatment with SB431542/dorsomorphin. This promoted the expression of lung progenitor genes and NKX2.1 and led to changes in the expression profile of CPM following lung differentiation. This aligns with previous findings that showed without optimisation of endoderm plating density, lung differentiation efficiency can be significantly impaired across several cell lines (Hawkins et al., 2017; Jacob et al., 2017). Similarly, other protocols plate a specific number of single endodermal cells prior to TGF-β/BMP inhibition (Huang et al., 2015; Rodrigues Toste de Carvalho et al., 2021). We observed that dissociation to single cells induced massive cell death even with supplements that promote cell survival, so opted to split cells as colonies. In contrast, Gotoh et al. (2014) did not find it necessary to dissociate endoderm cells before foregut patterning, but differences in pluripotent cell lines and culture conditions could explain these conflicting findings. There is currently no proposed mechanism explaining why over-confluent cultures inhibit induction of NKX2.1+ cells. However, in the context of cardiac differentiation, cell density can impact cell cycle regulation and the response of cells to differentiation with the GSK3β inhibitor CHIR99021 (Laco et al., 2018). When we performed a 3-way comparison of protocols for lung induction, we observed the highest expression of *SFTPC* in the condition with SFD basal medium and 3 µM CHIR, 10 ng/ml BMP4 and 100 nM. We therefore used these conditions in future experiments.

We demonstrated that splitting endoderm cells and treatment with the minimal factors CHIR, BMP4 and RA was sufficient to induce a population of cells expressing high levels of CPM which was enriched for lung progenitor cell transcripts (Huang et al., 2014; Serra et al., 2017). When CPM was first described as a surface marker of NKX2.1+ lung epithelial progenitor cells, the entire CPM+ fraction was purified for further differentiation and there was no report of the presence of non-lung cells types (Gotoh et al., 2014; Konishi et al., 2016). In further reports, a low CPM-expressing population was shown to contain few NKX2.1+ cells, but in CPM(hi) cells NKX2.1 was highly enriched in accordance with our findings (Korogi et al., 2019; Yamamoto et al., 2017). The fact that not all cells exhibited a high-CPM expression pattern after differentiation indicates the persistence of non-lung lineages in our culture despite a loss of hepatocyte markers. We would need to assess the transcriptomic profile of the CPM-low and negative populations to determine what other lineages were present,

but this highlights a continued limitation of current differentiation protocols where non-lung cell types emerge (Hawkins et al., 2017; Hein et al., 2022; Huang et al., 2014; Mithal et al., 2020; Ori et al., 2021; Yamamoto et al., 2017). This suggests the signalling parameters can be further refined to improve the specification of NKX2.1+ cells, but currently purification remains an essential step.

An interesting observation from our findings was the expression of transthyretin (*TTR*) across all lung conditions but also in CPM(hi) cells. The liver and endocrine glands of the pancreas are major sites of transthyretin synthesis which facilitates the transport of thyroid hormones (Refai et al., 2005). We reasoned *TTR* was upregulated because of contaminating hepatic cells. However, our CPM(hi) population was enriched for NKX2.1 and other lung transcripts suggesting *TTR* should not have been expressed. TTR has been recognised as a biomarker for lung cancer and can stimulate the proliferation of lung tumour cells (Lee et al., 2019). This might suggest there is an unidentified role for TTR during lung development, but this requires further investigation.

When we differentiated DE cells to AFE for 3 days and then split as colonies into lung specification medium we observed the induction of a significantly higher percentage of CPM(hi) cells. This finding raises several questions about early endoderm patterning that requires further investigation. One potential explanation is that splitting cells at the DE stage when they have greater differentiation potential might induce changes in cell function and response to differentiation cues resulting in emergence of non-lung endoderm derivatives. Further restriction of potential in AFE cells might limit the capacity to generate contaminating lineages after cell splitting in response to lung-specific differentiation signals. Performing scRNA-seq to compare AFE populations after 3-day dual-SMAD inhibition that were split and not split at the DE stage would help identify if splitting DE results in changes in the cell populations that emerge. Recent scRNA-seq analysis after 3-day SB431542/dorsomorphin treatment and lung differentiation revealed the presence of a significant CDX2+ hindgut population that could give rise to intestinal cells (Mithal et al., 2020). This suggests there are signalling requirements yet to be identified to further restrict the potential to an anterior foregut endoderm fate and/or highlights the plasticity of hPSC derivatives *in vitro*. More studies are required to interrogate this population *in vitro* and determine whether

more restricted AFE cells would lead to improved differentiation efficiency along the lung lineage.

We showed the CPM(hi) lung progenitor cells generated in our cultures could readily self-assemble into 3D organoids and differentiate into distal lung cells expressing gene and protein markers of type 2 AECs in the presence of WNT and FGF signalling, and factors implicated in maturation of type 2 AECs including corticosteroids and activators of cAMP signalling (Gonzales et al., 2002). Currently this is the only combination of growth factors reported for generating type 2 AECs from hiPSC derived-lung progenitor cells (Jacob et al., 2017; Jacob et al., 2019). Key roles for WNT signalling have been identified for driving lung alveologenesis and activating NKX2.1 expression to promote specification of type 2 AECs (Frank et al., 2016; Lim et al., 2023). FGF7 and FGF10 support the expansion of distal lung epithelium although human lung explants cultured with FGF10 do not undergo branching suggesting a role in humans distinct to that in the mouse in regulating branching morphogenesis (Danopoulos et al., 2019; Dye et al., 2015; Miller et al., 2018).

Our findings confirmed we could utilise the cell surface transporter NaPi2B to enrich for epithelial cells which robustly express transcripts of type 2 AECs like *SFTPC* and *NKX2.1* and show protein expression of SFTPC and pro-SFTPC, consistent with previous reports (Korogi et al., 2019). The expression of genes related to intestinal and liver lineages in negative NaPi2B populations confirms the persistence of non-lung derivatives, even after purifying lung progenitor cells and treatment with factors related to type 2 AEC development. This is a recurring problem with several lung differentiation protocols and occurs at earlier differentiation timepoints as well, as discussed above, and highlights the need to further uncover the signalling events in human alveolar specification and maintenance (Hein et al., 2022; Jacob et al., 2017; Yamamoto et al., 2017). Our protocol can mitigate some of these concerns as we can readily differentiate and purify type 2 AECs which is a major advantage of our differentiation platform.

4. Optimisation of a protocol to derive hiPSC-type 2 AECs in chemically defined medium

4.1. Introduction

Directed differentiation protocols to derive type 2 AEC organoids from hPSCs have successfully been developed which has enabled improved modelling of human alveolar development and disease. Despite these advances, few methods have been described to generate type 2 AECs *in vitro* and several limitations remain (Gotoh et al., 2014; Jacob et al., 2017; Jacob et al., 2019; Yamamoto et al., 2017). Off-target non-lung derivatives emerge during differentiation and cell fate can readily switch in culture, demonstrating the plasticity of hPSC-derived cells. This is further complicated as a few conserved signalling pathways including WNT, BMP, FGF and RA are used to specify multiple cell types from a similar origin. Reported differentiation efficiencies also vary considerably and the use of fluorescence reporter cell lines to track differentiation limits reproducibility (Hein et al., 2022; Hurley et al., 2020; Jacob et al., 2017; Korogi et al., 2019). Therefore, a current challenge is to improve our understanding of signalling events during lung specification to develop more robust differentiation protocols.

Small molecule inhibitors of signalling have been utilised during lung differentiation to drive cell patterning and specification. Inducing an anterior identity in endoderm cells can be mediated with pharmacological inhibitors of TGFβ and BMP signalling, but following this treatment hindgut-like populations can emerge and give rise to non-lung endoderm cells (Green et al., 2011; Mithal et al., 2020). Prospective AFE cells are also briefly exposed to the WNT signalling inhibitor Dickkopf 1 (DKK1) during development and has therefore led to use pharmacological inhibitors of WNT/β-catenin signalling like IWP2 *in vitro* (Huang et al., 2014). More potent small molecule inhibitors of TGFβ and BMP signalling such as A83-01 and LND193189, respectively, have been developed and used to generate AFE, but comparison of their effects to other small molecules on subsequent lung differentiation has not been evaluated (Kearns et al., 2013). The small molecule inhibitor of GSK3α/β, CHIR99021, is also commonly used to stimulate WNT/β-catenin signalling to drive lung progenitor cell specification. Given CHIR99021 does not directly activate the WNT pathway, it can have off-target effects and can

significantly impact functions such as cell cycle regulation during differentiation (Laco et al., 2018). Direct activation of WNT/β might lead to greater reproducibility (Hein et al., 2022).

The signalling requirements for development and maintenance of alveolar epithelial cells have not been fully elucidated. Type 2 AECs serve as the facultative stem cells of the mouse lung undergoing self-renewal and differentiation to type 1 AECs under homeostasis and in response to injury (Barkauskas et al., 2013). They have been posited to play a similar role in humans although this is yet to be confirmed (Lee & Rawlins, 2018). Adult type 2 AECs self-renew slowly during normal alveolar maintenance (Barkauskas et al., 2013). In contrast, hPSC-derived type 2 AECs are highly proliferative in the presence of WNT and FGF signalling (Jacob et al., 2017). This reflects the immature, more fetal-like phenotype of hPSC-derived type 2 AECs isolated via *SFTPC* fluorescent reporters (Sun et al., 2021). Human adult type 2 AECs are maintained in the presence of WNT, FGF and EGF ligands and inhibitors of TGFβ and BMP signalling, reflecting additional complexity in signalling requirements of type 2 AECs at varying stages of maturity (Katsura et al., 2020; Youk et al., 2020). Therefore, further investigation is required to determine if maturity of type 2 AECs derived from hPSCs can be promoted by further manipulation of signalling.

There have been increasing efforts to establish defined conditions for differentiation of hPSCs *in vitro*. Components used in culture medium to support cell survival and growth such as animal-derived serum comprise an undefined mix of proteins, cytokines, growth factors and hormones that introduce significant physiological variability. The composition of serum is strongly influenced by several factors including the method of sample collection, the state of health, diet and country of origin of the animal and biological differences between donors resulting in significant batch-to-batch variability. The presence of these factors can markedly impact cell growth and behaviour and hinder the reproducibility of hPSC differentiation protocols while also confounding our understanding of the signalling mechanisms of hPSC-derived cells during development and in disease (Burridge et al., 2014; Konagaya & Iwata, 2019). Although most lung differentiation protocols have established conditions free of animal serum, they have depended on replacement with components such as bovine serum albumin and B-27™ supplement to improve cell survival (Jacob et al., 2017;

Miller et al., 2019; Rodrigues Toste de Carvalho et al., 2021). B-27™ supplement was originally developed to support the growth of hippocampal neurons under serum free conditions but it is now routinely used in culture for multiple cell types (Brewer et al., 1993). It is a complex mix of vitamins, lipids, antioxidants and steroids but also contains BSA that is shown to have differential effects on cell growth (Francis, 2010). B-27™ supplement also suffers from batch variation and needs to be tested before use. For these reasons, developing medium with chemically defined composition is highly favourable. There are also several other drivers for replacing animal derived products in cell culture including translational potential of hPSC-derived cell products, but these will be discussed in chapter 5.

Here, we utilise our previously established directed differentiation protocol to generate type 2 AECs from hiPSCs in xeno-free, chemically defined medium. Under defined conditions, we assess the effect of different small molecule inhibitors of TGFβ, BMP and WNT signalling during foregut patterning and the impact on lung induction. We also investigate the impact of manipulating WNT, BMP, FGF, TGFβ and RA signalling on the efficiency of specification of lung epithelial progenitor cells. We determine the benefit of using recombinant WNT3a and R-spondin for direct WNT/β-catenin instead of CHIR99021 to efficiently induce lung fate and assess the impact of conditions for maintenance of adult type 2 AECs on the phenotype of hiPSC-derived type 2 AECs. Finally, we use bulk-RNA sequencing to benchmark our derived cell types at major stages of the differentiation protocol.

4.2. Results

4.2.1. hiPSC differentiation to DE is robust and reproducible in XFL medium

After successfully establishing a platform to differentiate hiPSCs to type 2 AECs, we wanted to develop a chemically defined, xeno-free protocol to reduce biological variability during differentiation and remove animal-derived components to streamline the clinical application of derived type 2 AECs. To do this, we used a chemically defined, xeno-free basal medium for all steps of differentiation, referred to as xeno-free lung (XFL) medium and performed optimisation experiments on Matrigel™ using our previously optimised protocol as a benchmark (**Fig. 4-1a**). For the majority of experiments in this the BE.31 ($SFTPC^{WT/WT}$) hiPSC cell line was used, unless stated otherwise in the figure legend or text. To assess the capacity for hiPSC differentiation to definitive endoderm in XFL medium, we compared combinations of endoderm induction factors to our previously optimised protocol for DE specification. We used CXCR4 and CKIT co-expression to quantify endoderm induction efficiency as before. In the presence of activin A alone, most cells expressed CKIT but lacked CXCR4 induction, and stimulation of WNT signalling led to emergence of two cell populations, one co-expressing CXCR4 and CKIT, which was significantly potentiated with PI3K inhibition via LY294002, and one expressing CKIT alone (**Fig. 4-2a**). Using CHIR99021 in place of recombinant Wnt3a for Wnt signalling activation improved the induction of CXCR4+/CKIT+ cells, an effect that was further enhanced with supplementation of LY294002, leading to a mean percentage of 98.5% CXCR4+/CKIT+ cells (N=4) (**Fig. 4-2a, b**). To determine the optimal density for differentiation, we seeded hiPSCs at 10,000, 20,000 and 30,000 cells/cm^2 and showed all 3 densities led to efficient induction of CXCR4+/CKIT+ endoderm in XFL medium on Matrigel™, as cells expanded and migrated from colonies to form a complete monolayer by day 3 of differentiation (**Fig. 4-3a, b**). Differentiation to DE in defined conditions was highly reproducible across multiple cell lines. In XFL medium on Matrigel™, the mean percentage of CXCR4+/CKIT+ cells for BE.31 ($SFTPC^{WT/WT}$) hiPSCs were 98.14% (N=23), for ABE.27 ($SFTPC^{MUT/MUT}$) hiPSCs, 99.92% (N=6), for C3g3.14 corrected hiPSCs, 99.56% (N=3) and for LOPCK hiPSCs, 99.23% (N=7) (**Fig. 4-4a**). After 4 days of differentiation, immunofluorescence microscopy further confirmed the expression of the endoderm markers SOX17 and FOXA2 in DE cells (**Fig. 4-4b**). These findings demonstrated the

robustness of hiPSC differentiation to definitive endoderm in fully defined, xeno-free medium.

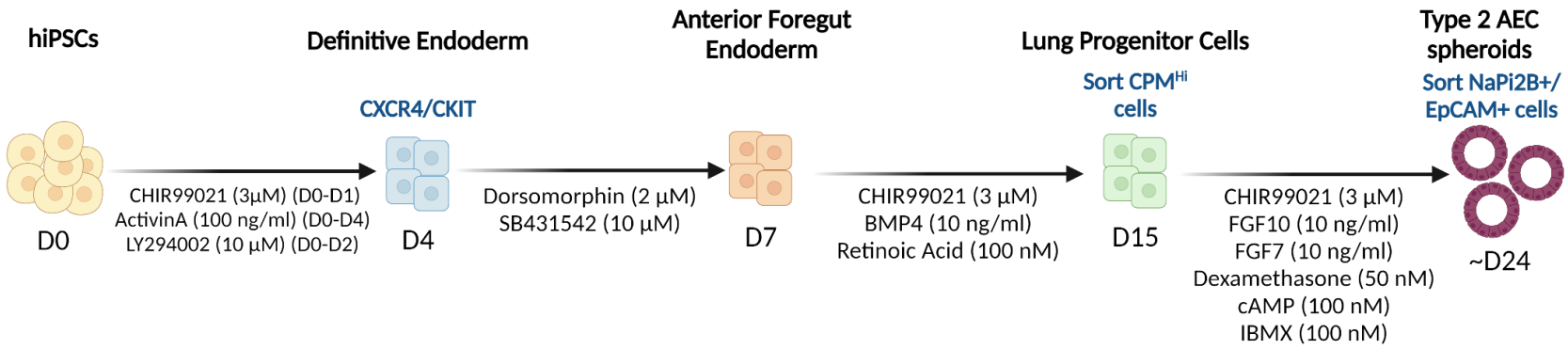


Figure 4-1. Schematic of the optimised directed differentiation of hiPSCs to type 2 AECs.

hiPSCs are differentiated to DE in the presence of CHIR, activin A and LY across a 4-day protocol. Following endoderm specification cells are treated with inhibitors of TGFβ and BMP signalling to promote anterior patterning, followed by differentiation towards lung epithelial progenitor cells via treatment with CHIR, BMP4 and RA. CPM(hi) cells are isolated and maturation to type 2 AECs is induced through treatment with CHIR, FGF7, FGF10, dexamethasone, cAMP and IBMX. Figure generate using BioRender.

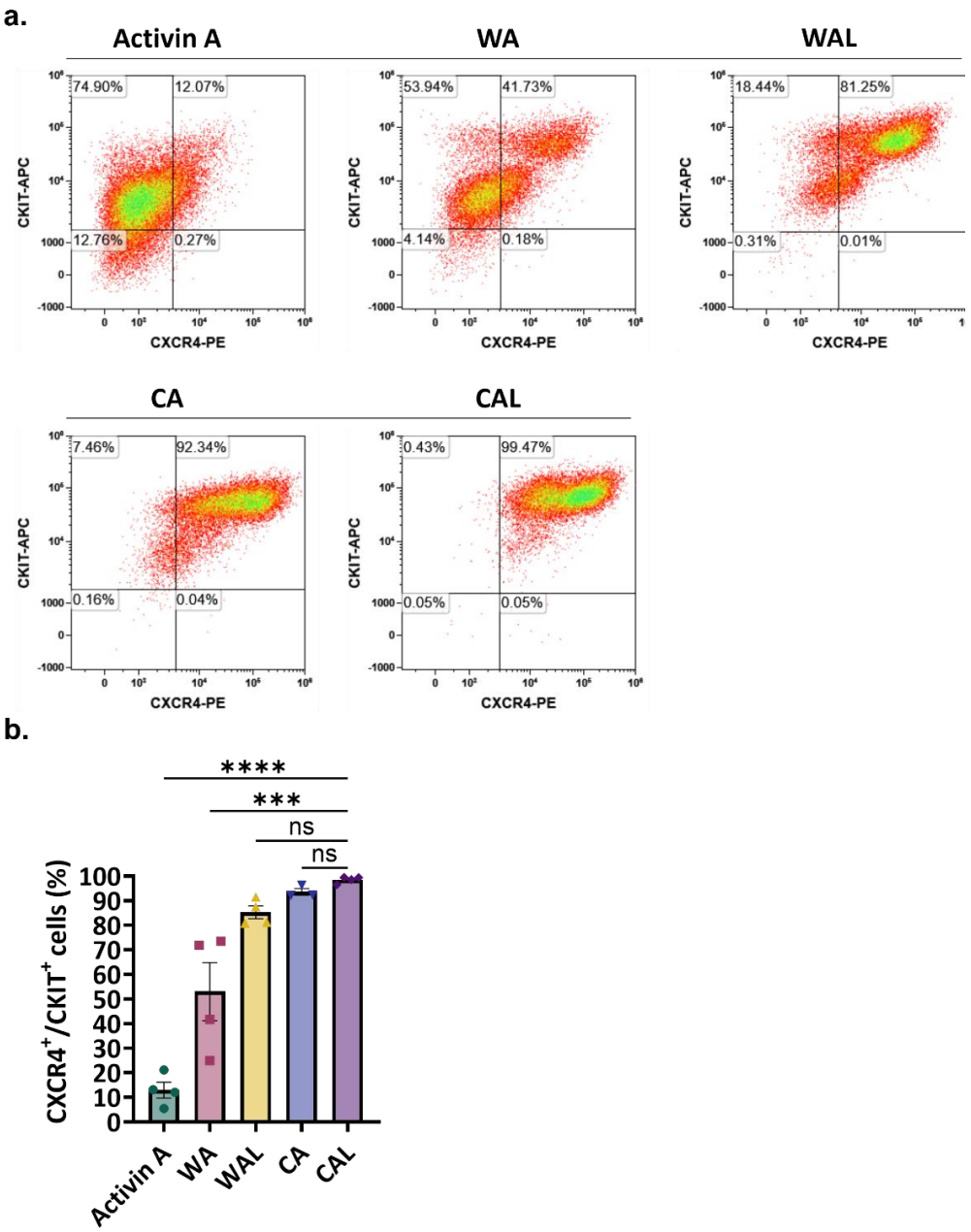


Figure 4-2. Optimised endoderm differentiation generates the highest number of CXCR4+/CKIT+ cells in XFL medium.

(a) Representative flow cytometry analysis of day 4 BE.31 ($SFTPC^{WT/WT}$) endoderm cells for CXCR4 and CKIT differentiated with combinations of growth factors and **(b)** quantification of CXCR4+/CKIT+ cells with bars representing mean ± SEM (N=3, CA; N=4 all other conditions). W = Wnt3a (50 ng/ml), A = Activin A (100 ng/ml), L = LY294002 (10 µM), C = CHIR99021 (3 µM). *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001, one-way ANOVA with Tukey's multiple comparisons test. NS = non-significant.

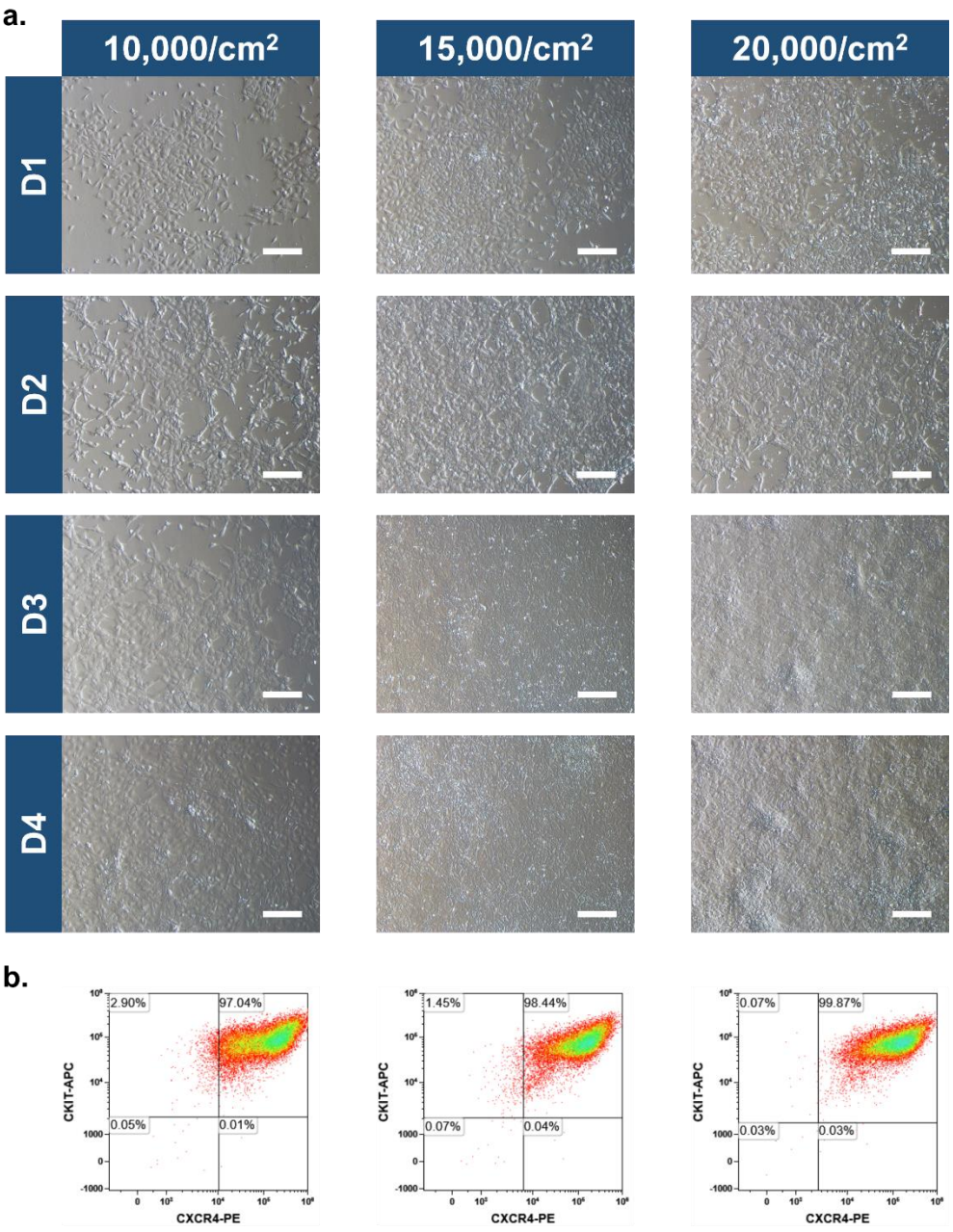


Figure 4-3. Optimisation of cell seeding density for differentiation of hiPSCs to endoderm in XFL medium on Matrigel™.

(a) Representative light microscope images of cells during endoderm differentiation after seeding hiPSCs at increasing densities. Scale bars, 200 µM. **(b)** Flow cytometry analysis of day 4 endoderm cells for CXCR4 and CKIT co-expression for each cell seeding density (N=1).

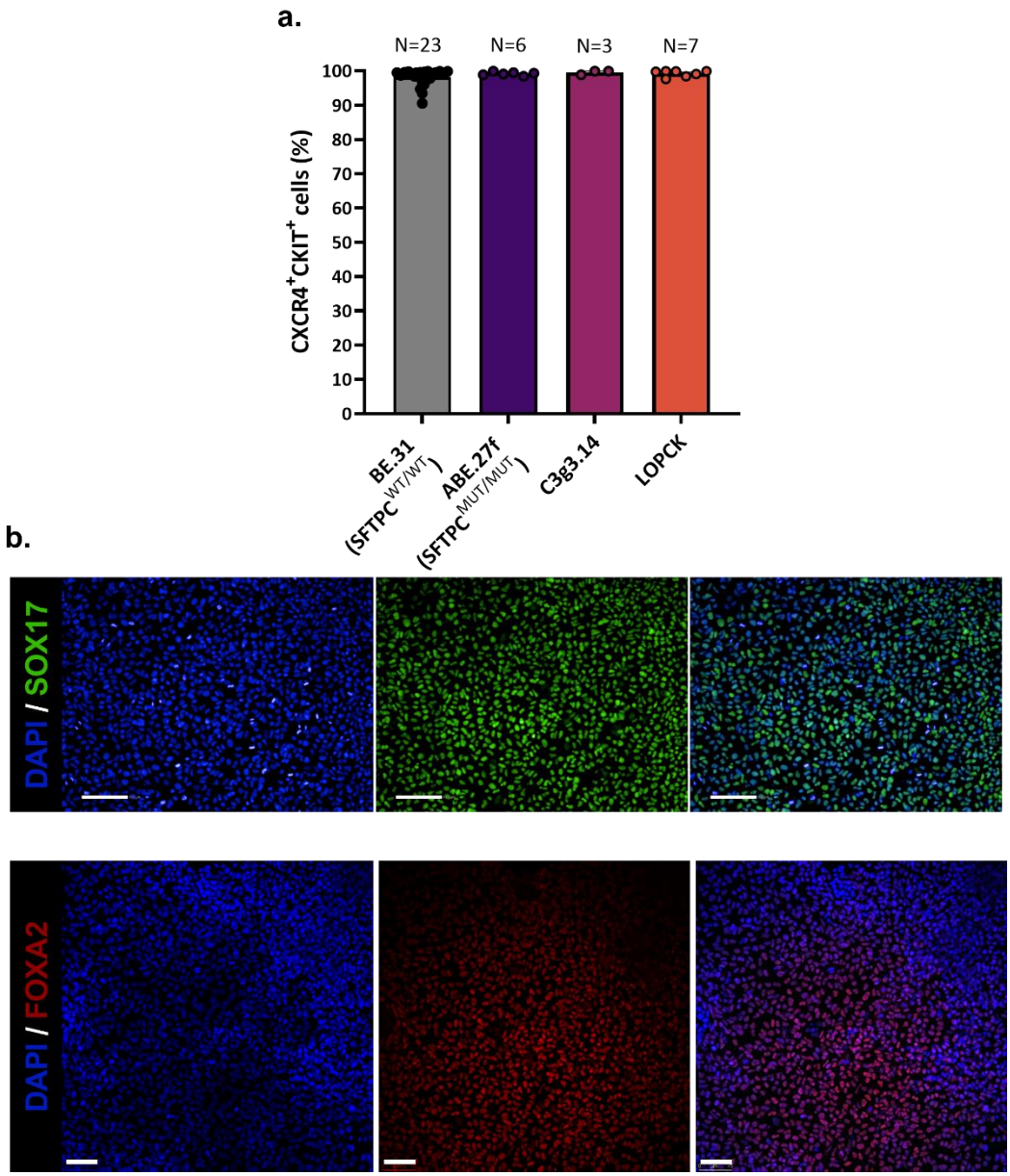


Figure 4-4. hiPSCs can be robustly differentiated to definitive endoderm in chemically defined medium.

(a) Quantification of CXCR4+/CKIT+ cells on day 4 of endoderm differentiation for 3 cell lines in XFL medium on Matrigel™ with bars representing mean ± SEM. **(b)** Immunofluorescent staining for SOX17 and FOXA2 in DE D4 BE.31 ($SFTPC^{WT/WT}$) cells differentiated in XFL medium. Scale bars, top row, 100 µm, bottom row, 75 µm.

4.2.2. Transcriptomic profiling of hiPSC-derived DE cells

To define the transcriptomic profile of our hPSC-derived cells in chemically defined medium, we performed RNA sequencing (RNA-seq) analysis for BE.31 ($SFTPC^{WT/WT}$) hiPSCs at major developmental stages across our differentiation protocol. The first timepoint we analysed was DE D4. There was a significant enrichment of transcripts associated with DE development including *SOX17*, *GATA6* and *TBX3* in DE cells compared to undifferentiated hiPSCs (**Fig. 4-5a**). In the top 20 DEGs, we observed the enrichment of the DE marker *EOMES* (**Fig. 4-6a**), and when we performed profiling of a selected set of DE genes, we observed significant expression of *SOX17*, *FOXA2*, *GATA4*, *GATA6*, *HNF4α* and *CXCR4* in hiPSC-DE cells (**Fig. 4-6b**). These genes were not expressed in pluripotent cells. Expression of markers such as *CDX2* were not upregulated, suggesting there was no early lineage bias toward hindgut fate (**Fig. 4-6b**).

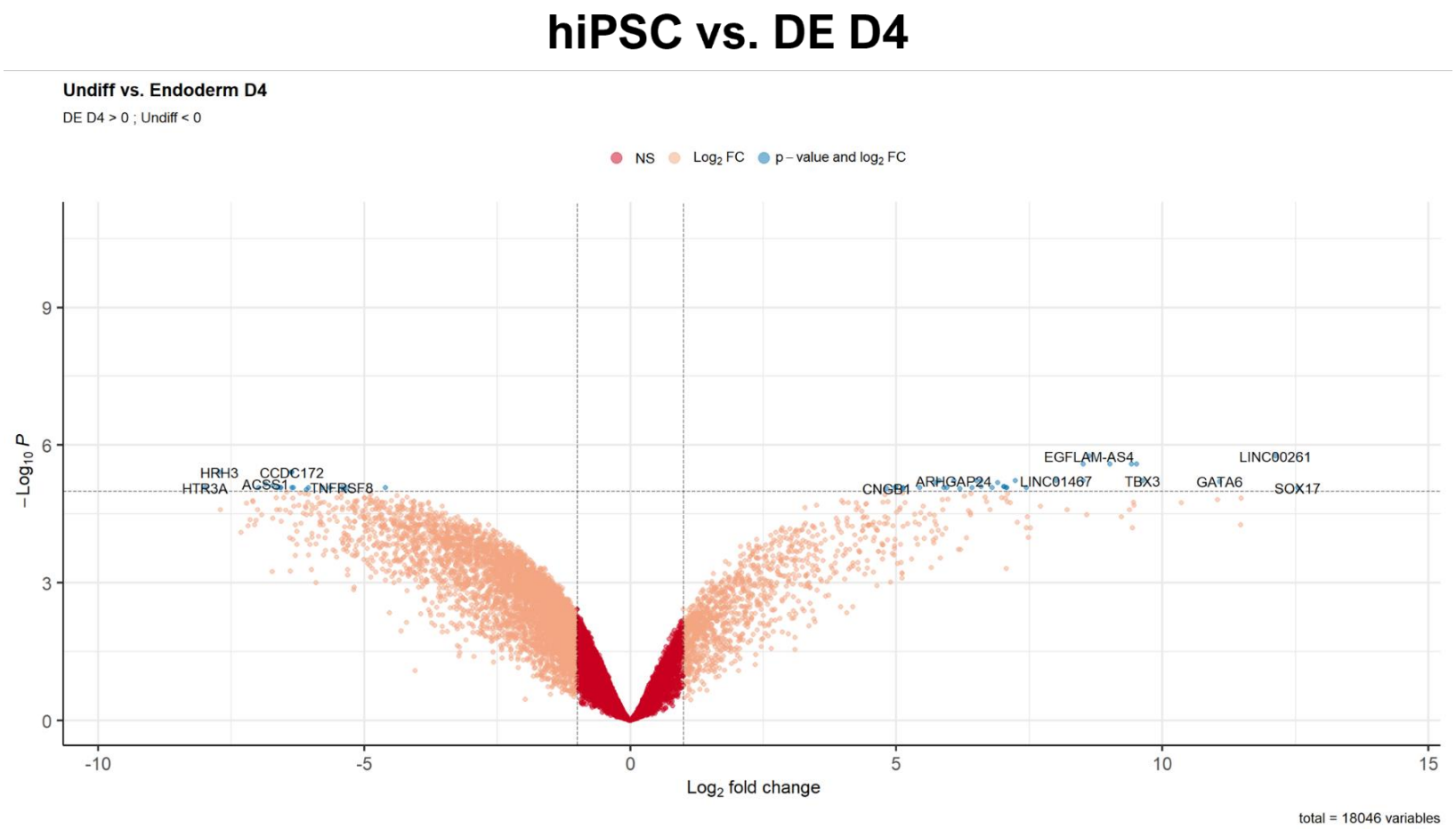


Figure 4-5. Differentiation of hiPSCs to DE elicits transcriptomic changes.

Volcano plot of differentially expressed genes in undifferentiated hiPSCs versus day 4 DE cells. Statistically significant ($P \leq 0.05$) changes are highlighted.

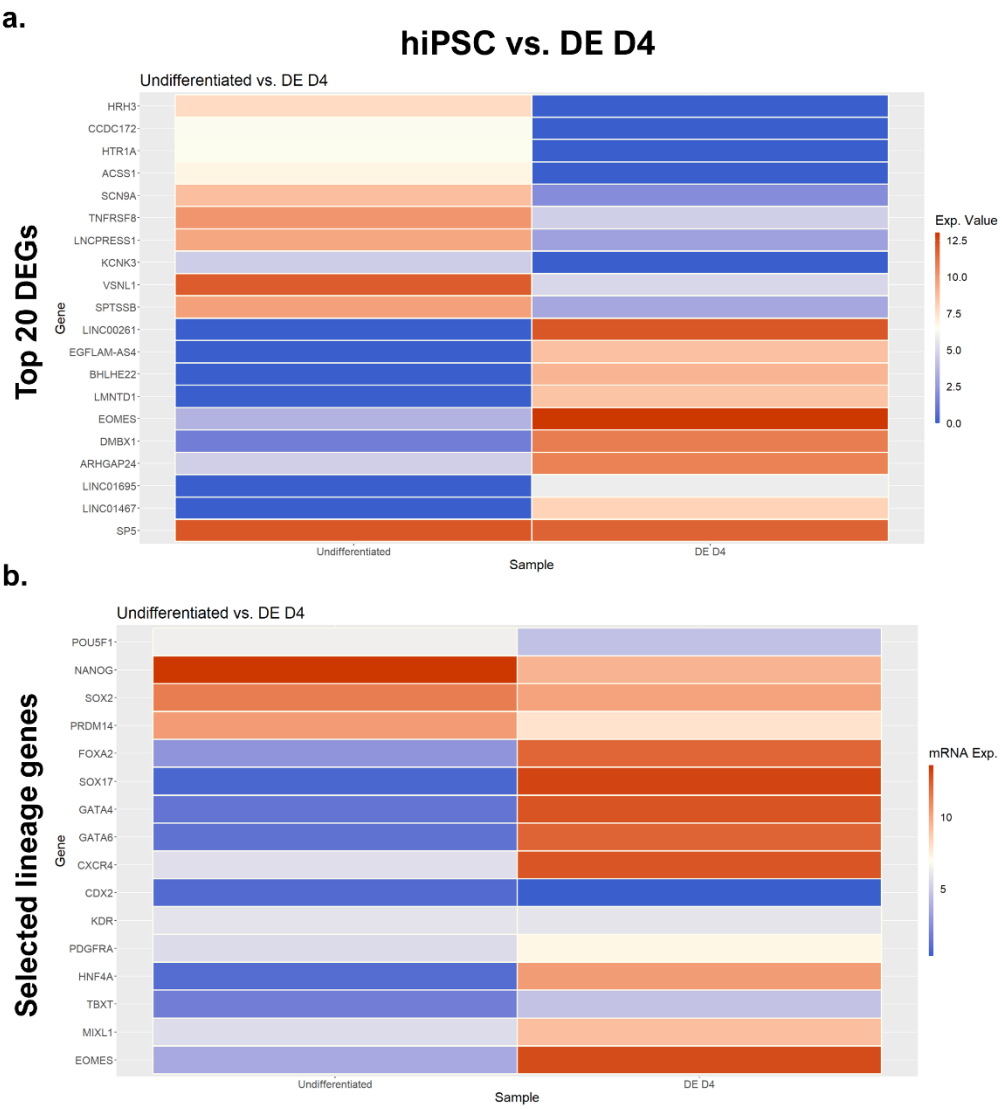


Figure 4-6. Transcriptomic profiling of hiPSC-derived DE cells.

(a) Heatmap of the top 20 DEGs in undifferentiated hiPSCs versus DE cells. **(b)** Heatmap of the expression of selected lineage markers in undifferentiated hiPSCs versus DE cells. All DEGs displayed are statistically significant to $P < 0.05$.

4.2.3. DE cells can be differentiated to CPM-expressing lung progenitor cells in XFL medium

We next sought to confirm the capacity to generate lung epithelial progenitor cells in xeno-free medium with our optimised protocol. We demonstrated previously that confluency of cells significantly impacted the efficiency of differentiation and therefore introduced a stage where DE cells were split and re-seeded at an optimised ratio. To confirm this was still necessary, we performed differentiation without splitting DE cells and showed significantly poor induction of CPM(hi) cells (**Fig. 4-7a**). We assessed several methods to split cells for further differentiation. Splitting DE as single cells before foregut and lung specification induced >38% CPM(hi) cells, but cell survival after dissociation was poor even in the presence of CloneR™2 and this method was not reproducible (**Fig. 4-7b**). When DE cells were split as colonies, low splitting ratios led to poor induction of CPM(hi) cells, but this was improved by increasing the ratio, and 1:8 was optimal (**Fig. 4-7c, d**). We also demonstrated cells could be dissociated after patterning DE to AFE for 3 days, but this similarly led to poor survivability and was difficult to reproduce (**Fig. 4-7e**). This confirmed we could specify a CPM(hi) population could in XFL medium, but cell density remained a critical parameter that needed to be optimised for efficient differentiation. We analysed CPM populations via qRT-PCR and observed the upregulation of the lung progenitor genes *NKX2.1* and *CPM* and distal bud tip marker *SOX9* in CPM(hi) cells. Markers related to hepatocyte (*ALB*, *AFP*) and hindgut specification (*CDX2*) were not expressed in our CPM(hi) population but were upregulated in CPM+ and CPM- cells. Interestingly, *Villin*, a protein of the microvilli in the brush border epithelial cells of the intestine, was expressed in all CPM populations (**Fig. 4-8a**) (Wang et al., 2008). However, brush cells expressing Villin have also been identified in the respiratory epithelium (Reid et al., 2005). To determine whether *Villin* expression in our system is observed because of off target intestinal cells in our enriched population or is a feature of lung epithelial progenitor cells requires further gene expression analysis and protein characterisation via immunocytochemistry. Similarly, *TTR*, a transport protein synthesised predominantly in the liver, was also expressed in CPM(hi) cells, despite lack of *ALB* and *AFP* (**Fig. 4-8**). There are currently no well-known functions of *TTR* in the lung although

it is overexpressed in some lung cancers and can stimulate cell proliferation (Lee et al., 2019). Performing single cell RNA sequencing would reveal any potential heterogeneity in the CPM(hi) population and confirm the presence of non-lung cell types after enrichment to explain the expression of genes such as *Villin* and *TTR*. Nonetheless, taken together these findings indicate our CPM(hi) population was enriched for cells expressing markers of lung progenitor cells and the remaining population consist of contaminating non-lung endoderm cell types. This highlights the necessity of a purification step during differentiation.

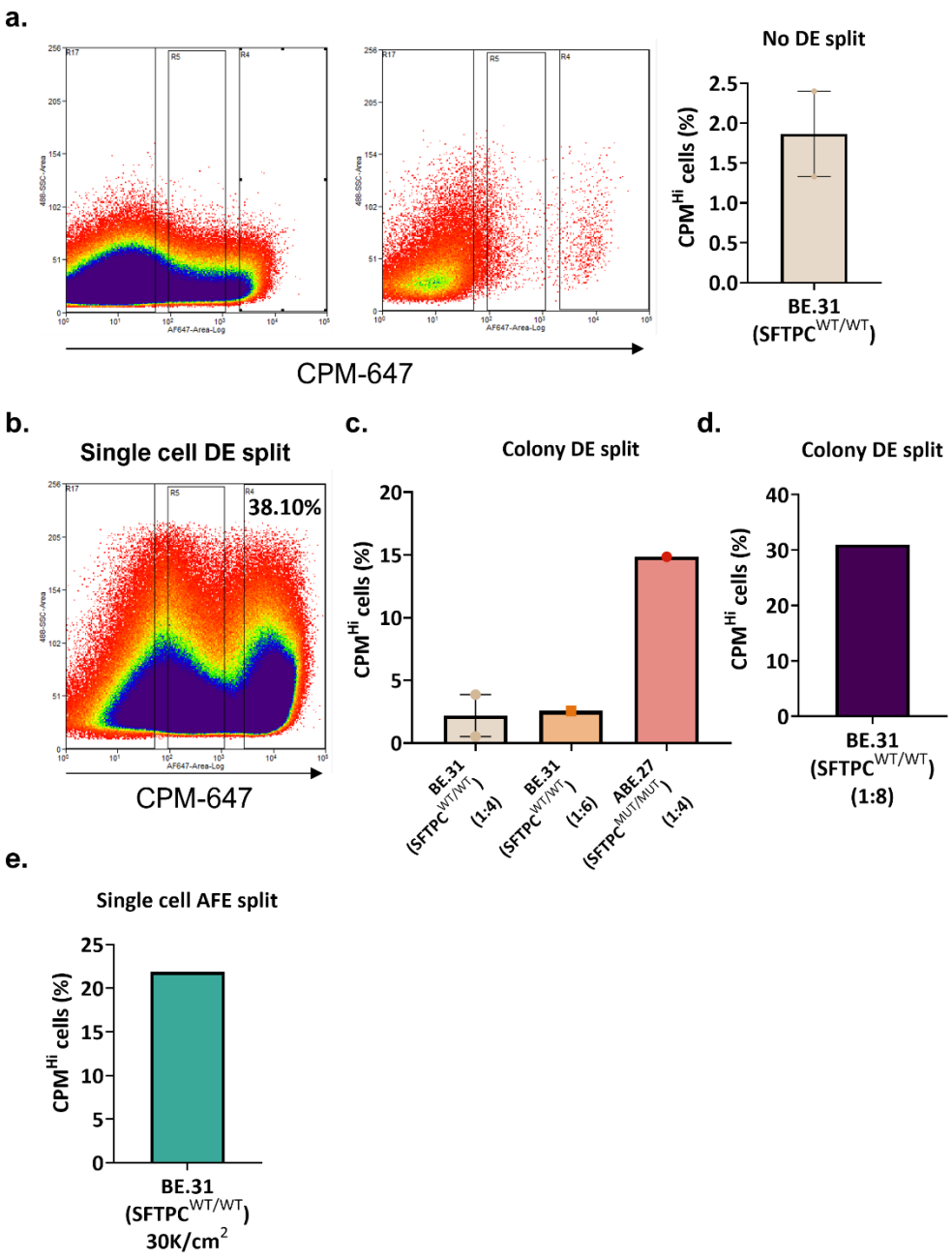


Figure 4-7. Confluency of cells significantly impacts specification of CPM(hi) cells.

(a) Flow cytometry analysis of BE.31 (SFTPC$^{WT/WT}$) lung cells for CPM following no cell splitting of DE cells and quantification of CPM(hi) cells. Bars represent mean ± SEM **(b)** Flow cytometry analysis of BE.31 (SFTPC$^{WT/WT}$) lung cells after a single cell split of DE cells. Quantification of CPM(hi) cells after splitting DE cells as colonies at ratios of **(c)** 1:4 and 1:6 with bars representing mean ± SEM and **(d)** 1:8. **(e)** Quantification of CPM(hi) cells following a single cell split after 3 days of DS/SB treatment.

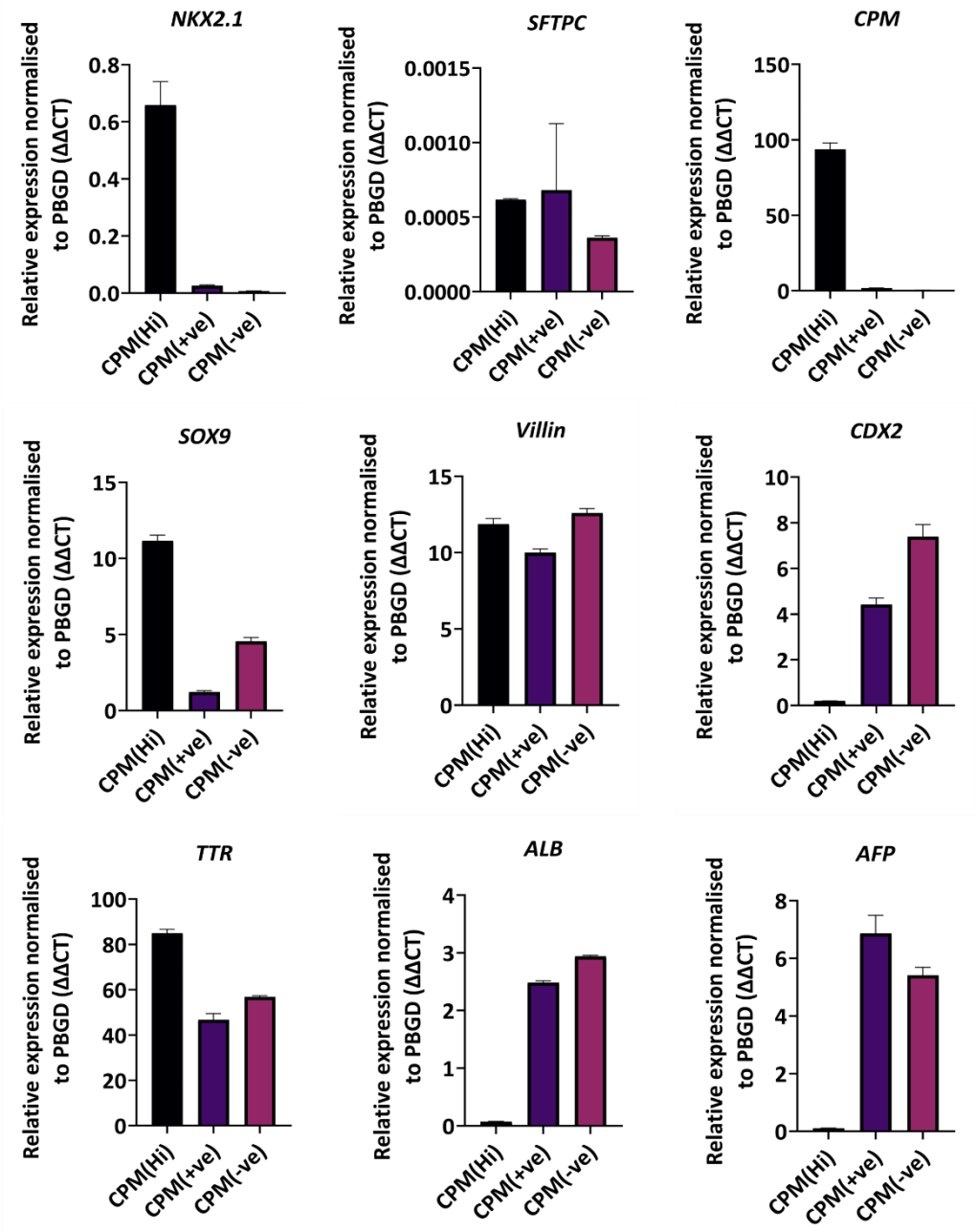


Figure 4-8. CPM(hi) cells differentiated in XFL medium express lung epithelial progenitor genes.

RT-qPCR of lung and non-lung endoderm genes in lung day 6 CPM sorted populations differentiated in animal-free medium with bars representing mean normalised expression to the housekeeping gene PBGD (ΔΔCt) ± SEM (N=1, 3 technical replicates).

4.2.4. Specification of lung progenitor cells is modulated by changes in activation of signalling pathways

Our next aim was to improve the efficiency and reproducibility of our differentiation protocol. To do this, we assessed the effect of modulating major signalling pathways involved in lung development including FGF, BMP, WNT and Retinoic acid signalling on lung progenitor cell specification by quantification of CPM expression via flow cytometry. We included our previously optimised protocol consisting of CHIR99021, BMP4 and retinoic acid as a control. Given CHIR99021 is a small molecule inhibitor of GSK3β signalling and impacts other cellular pathways, we determined the impact of more specific WNT activation by recombinant WNT3a/R-spondin on lung differentiation efficiency. Under our control conditions, <3% CPM(hi) cells were generated. This was improved by replacing CHIR99021 with WNT3a/R-spondin and led to ~10% CPM(hi) cells, an effect that was not enhanced by doubling the recombinant WNT3a concentration (**Fig. 4-9a, b**). Removal of retinoic acid from the culture completely diminished the CPM(hi) population, validating previous findings that RA is indispensable for lung differentiation (Rankin et al., 2016). A small percentage of CPM(hi) cells could still be generated with removal of BMP4 and inhibition of BMP signalling via dorsomorphin. In the presence of CHIR99021, supplementing FGF2, FGF7 and FGF10 independently led to induction of <0.1% CPM(hi) cells. Only a marginal increase in CPM(hi) cell percentage with WNT3a/R-spondin in place of CHIR99021 was observed (**Fig. 4-9a, b**). This suggests FGF signalling is redundant for lung differentiation in the human system as previous findings have suggested despite its involvement in branching morphogenesis during development. Significant differences in the morphology of cells during differentiation with Wnt/R-spondin or CHIR99021 were also apparent. Treatment with CHIR99021 induced significant over-proliferation as shown by the aggregation of cell clusters across the monolayer. These clumps were not visible in cultures differentiated with Wnt/R-spondin and instead large cells formed a 2D monolayer (**Fig. 4-10a**). This indicates CHIR99021 might impact signalling pathways unrelated to WNT to drive over-proliferation or specification to non-lung cell types.

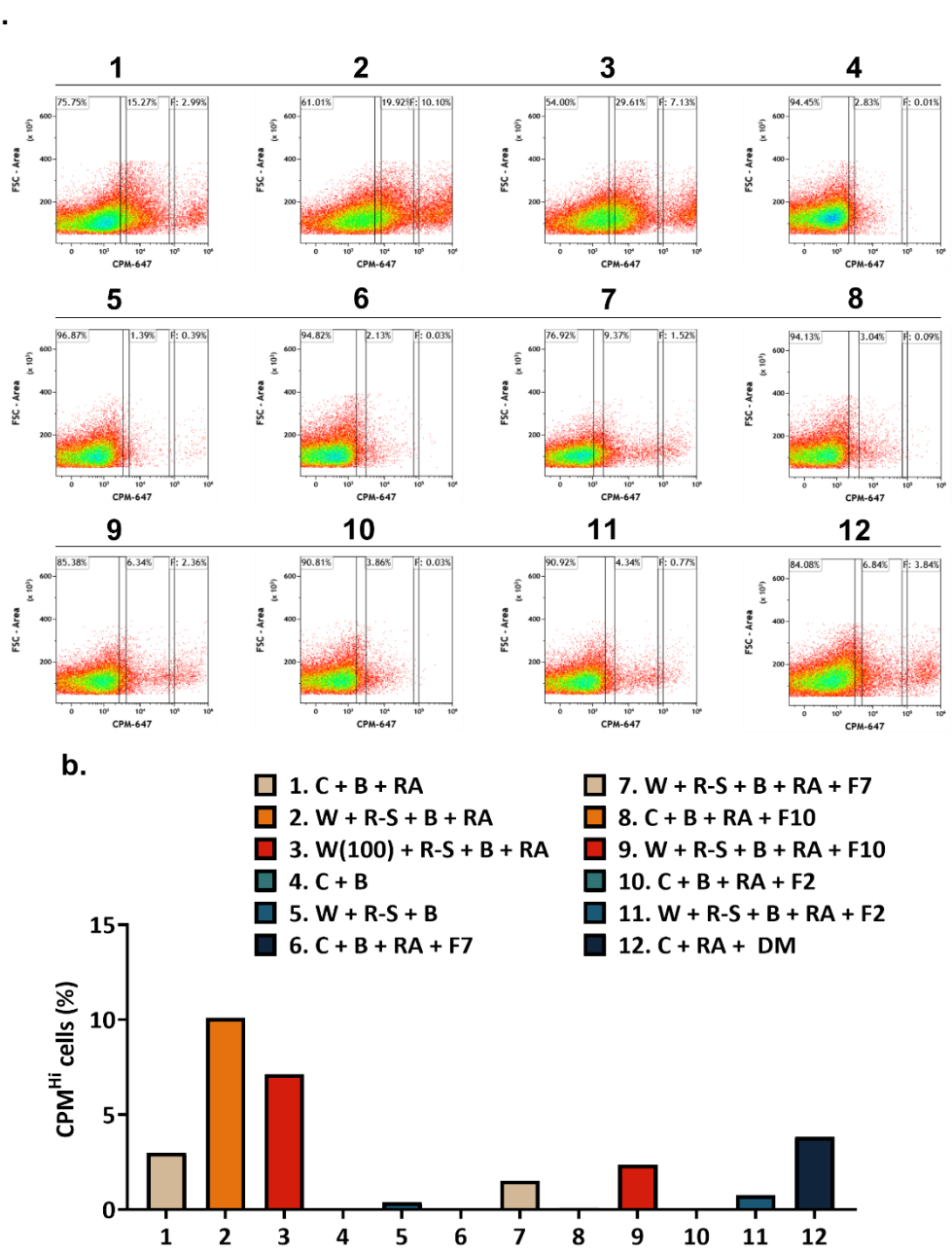


Figure 4-9. Modulation of growth factors during lung progenitor cell differentiation impacts specification of CPM(hi) cells.

(a) Flow cytometry analysis of lung day 6 cells for CPM for 12 different growth factor combinations, with **(b)** quantification of CPM(Hi) cells. C = CHIR99021 (3µM), B = BMP4 (10ng/ml), RA = Retinoic acid (100nM), R-S = R-spondin (100ng/ml), W = Wnt3a (50ng/ml or 100ng/ml where specified), F7 = FGF7 (10ng/ml), F10 = FGF10 (10ng/ml), F2 = FGF2 (10ng/ml), DM = Dorsomorphin (2µM).

a.

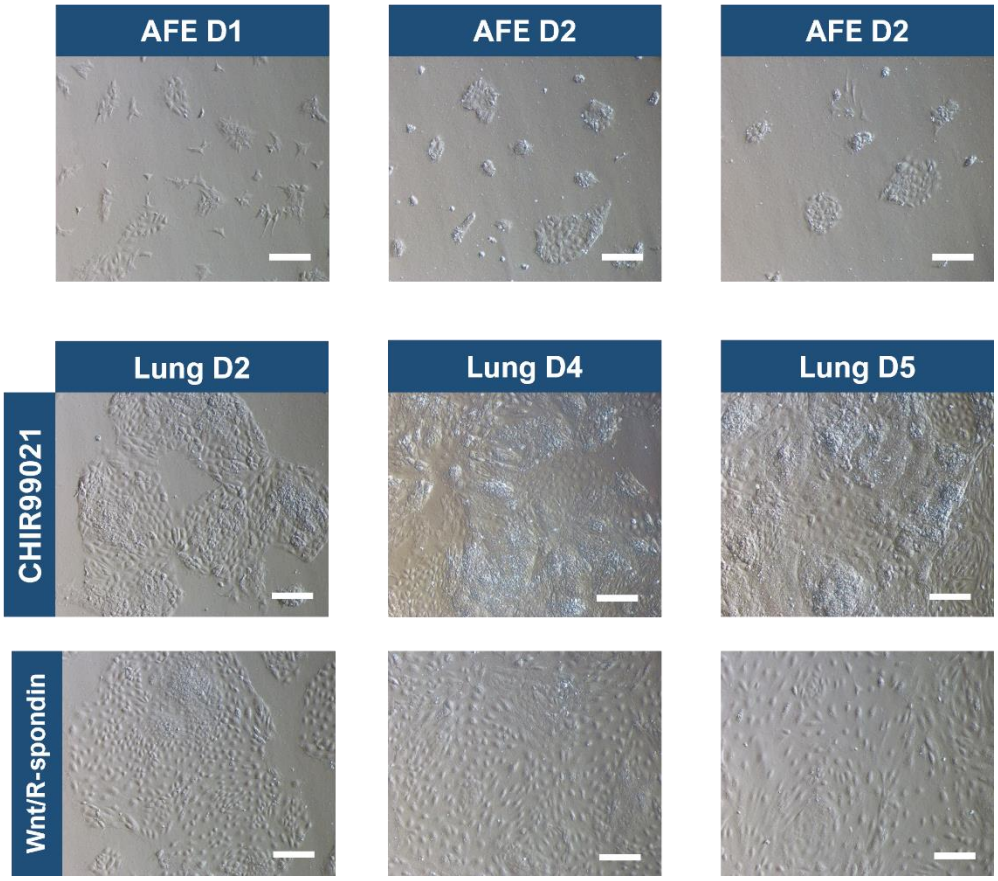


Figure 4-10. CHIR99021 induces over-proliferation during lung progenitor specification.

(a) Representative light microscope images of cells during anterior foregut specification and lung differentiation in the presence of CHIR99021 or Wnt/R-spondin. Scale bars, 200 µm.

4.2.5. Changes in small molecule inhibition during foregut endoderm patterning impact differentiation to lung epithelial progenitor cells

Previous findings have demonstrated the importance of inhibiting TGFβ and BMP signalling (dual-SMAD inhibition) to pattern DE cells to an anterior fate to permit the specification of the lung lineage (Green et al., 2011). We corroborated these results previously and showed dual SMAD-inhibition by SB431542 and dorsomorphin was necessary for emergence of CPM(hi)lung progenitor cells. During development,

anterior foregut endoderm populations are also exposed briefly to the WNT signalling inhibition. This signalling event has been mimicked *in vitro* and transient Wnt inhibition following dual-SMAD inhibition has been shown to improve the efficiency of subsequent lung differentiation (Huang et al., 2014). To investigate the impact of foregut patterning on efficiency of lung differentiation in XFL medium, we tested different combinations of small molecule inhibitors of Wnt, BMP and TGF-β signalling. Differentiation to DE was performed using our optimised protocol and cells were split at a ratio of 1:12 prior to foregut patterning. We included treatment with SB431542 and dorsomorphin for 3 days as a control, and tested A83-01, a potent inhibitor of TGF-β signalling, LDN193189, a selective BMP signalling inhibitor, and IWP2, an inactivator of porcupine to inhibit WNT signalling (**Fig. 4-11a**). The percentage of CPM(hi) cells determined via flow cytometry was used to quantify efficiency. Dual-SMAD inhibition by SB43152 or A83-01 in combination with dorsomorphin for 3 days led to a similar percentage of CPM(hi) cells following lung specification via CHIR99021 or Wnt3a/R-spondin (**Fig. 4-11b, c**). Differentiation with recombinant proteins yielded slightly fewer CPM(high) cells (**Fig. 4-11c**). Endoderm cells treated with SB431542 and dorsomorphin for 1 day followed by SB431542 and IWP2 for 2 days induced ~24% CPM(hi) cells with CHIR99021, but efficiency was significantly lower with WNT3a/R-spondin. Compared to SB431542, use of A83-01 still generated CPM(hi) cells but with lower efficiency in the presence of CHIR99021 and WNT3a/R-spondin during the lung stage of differentiation (**Fig. 4-11c**).

Inhibiting BMP signalling with LDN193189 in combination with SB431542 or A83-01 for 3 days significantly diminished the induction of a CPM(hi) population after differentiation with WNT3a and R-spondin (**Fig. 4-12a, b**). Combining dual-SMAD and WNT inhibition led to emergence of CPM(hi) cells, but the efficiency was still <10% (**Fig. 4-12b**). These findings suggested the mechanism of BMP inhibition exerted by LDN193189 is refractory to lung differentiation and dual-SMAD inhibition for 3 days was the optimal treatment.

a.

Small molecule inhibitor	Signalling pathway	Receptor kinase inhibited
SB431542	TGFβ/Activin/NODAL	ALK4, ALK5, ALK7
A83-01	TGFβ/Activin/NODAL	ALK4, ALK5, ALK7
Dorsomorphin	BMP	ALK2, ALK3, ALK6, AMPK
LDN193189	BMP	ALK1, ALK2, ALK3, ALK6
IWP2	WNT	Porcupine

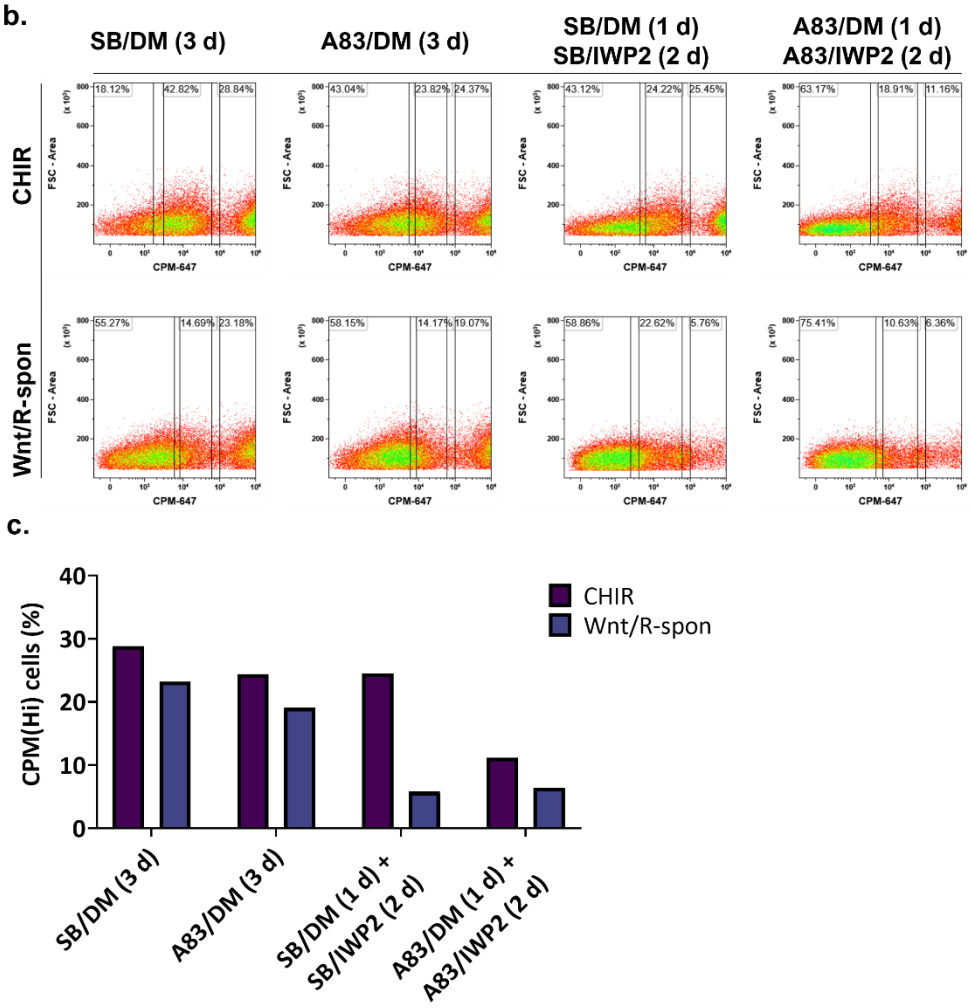


Figure 4-11. Effect of dual-SMAD and Wnt signalling inhibition during foregut patterning on induction of CPM(hi)-expressing cells.

(a) Table showing small molecule inhibitors and their targets. (b) Flow cytometry analysis of cells on day 6 of lung differentiation for CPM for 4 conditions of foregut specification prior to lung differentiation with Wnt signalling via CHIR99021 or Wnt and R-spondin, and (c) quantification of CPM(hi) cells on lung D6 (N=1) (d = days). SB = SB431542 (10 µM), DM = Dorsomorphin (2 µM), A83 = A83-01 (1 µM), IWP2 (1 µM).

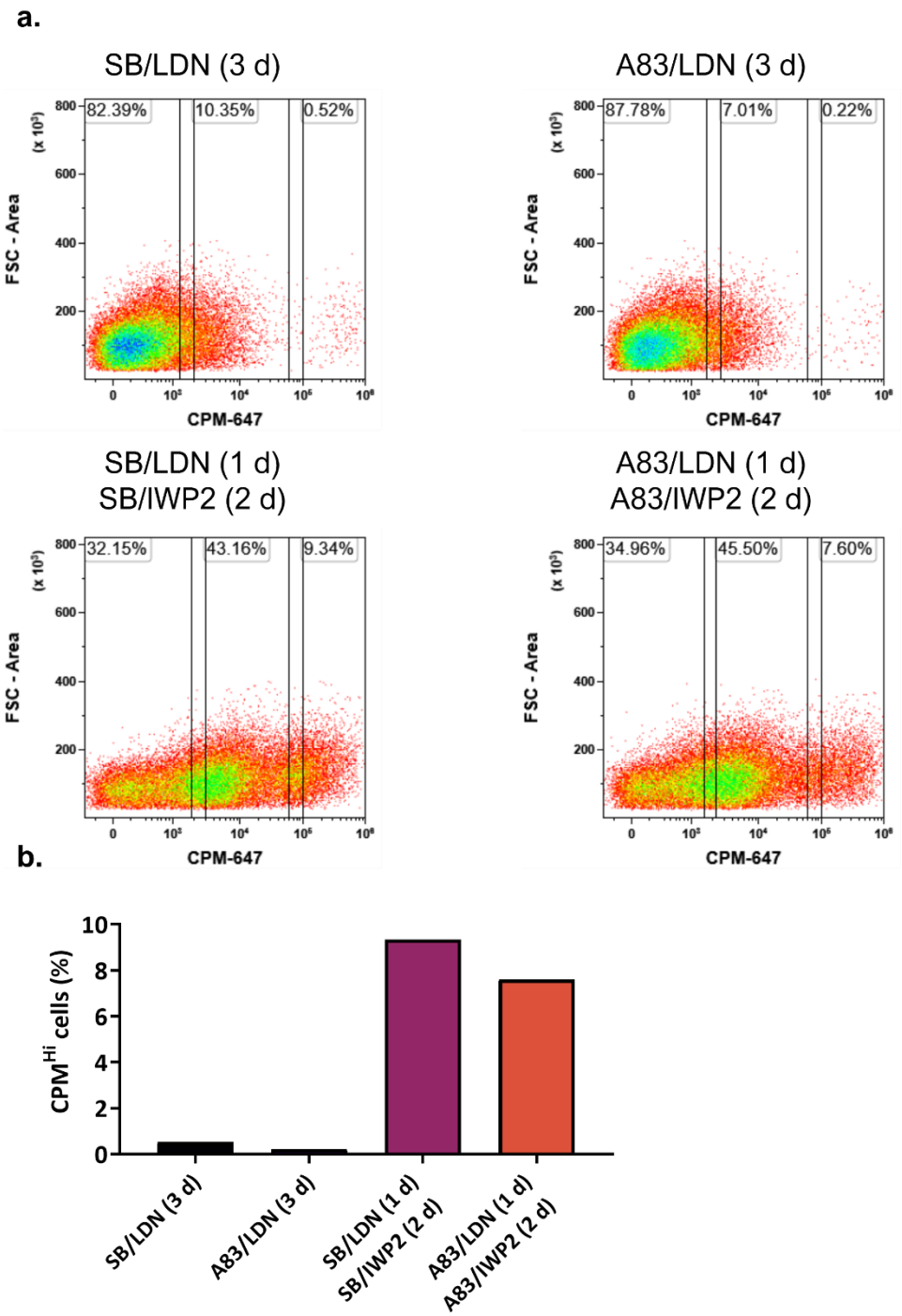


Figure 4-12. BMP signalling inhibition via LDN193189 during foregut patterning reduces specification of CPM(hi) cells.

(a) Flow cytometry analysis of lung progenitor cells on day 6 of differentiation for CPM following foregut patterning with the small molecule inhibitor of BMP signalling LDN193189 and **(b)** quantification of the percentage of CPM(hi) cells. SB = SB431542 (10 µM), LDN = LDN193189 (250 nM), A83 = A83-01 (1 µM), IWP2 (1 µM).

4.2.6. A dose-dependent increase in recombinant WNT3a and R-spondin concentration improves specification of CPM(hi) cells

Although CHIR99021 led to a marginal increase in the percentage of CPM(hi) cells compared to WNT3a/R-spondin during the lung induction stage of differentiation, we sought to determine the effect of increasing the concentration of WNT/R-spondin on lung progenitor differentiation efficiency. We showed an increase in recombinant protein concentration led to an increase in the percentage of CPM(hi) cells in a dose-dependent manner, with a ~5-10% increase in CPM(hi) cells from the lowest to highest concentrations across all foregut patterning conditions tested (**Fig. 4-13a, b**). In this particular experiment, dual-SMAD inhibition via A83-01 and dorsomorphin for 1 day followed by A83-01 and IWP2 treatment for 2 days induced the highest percentage of CPM(hi) cells (>30%) (**Fig. 4-13a, b**). This confirmed our hypothesis that direct WNT signalling activation by recombinant proteins could efficiently induce lung progenitor cells in XFL medium and we concluded that 100 ng/ml WNT3a and 250 ng/ml R-spondin were the optimal concentrations of which we tested for lung differentiation.

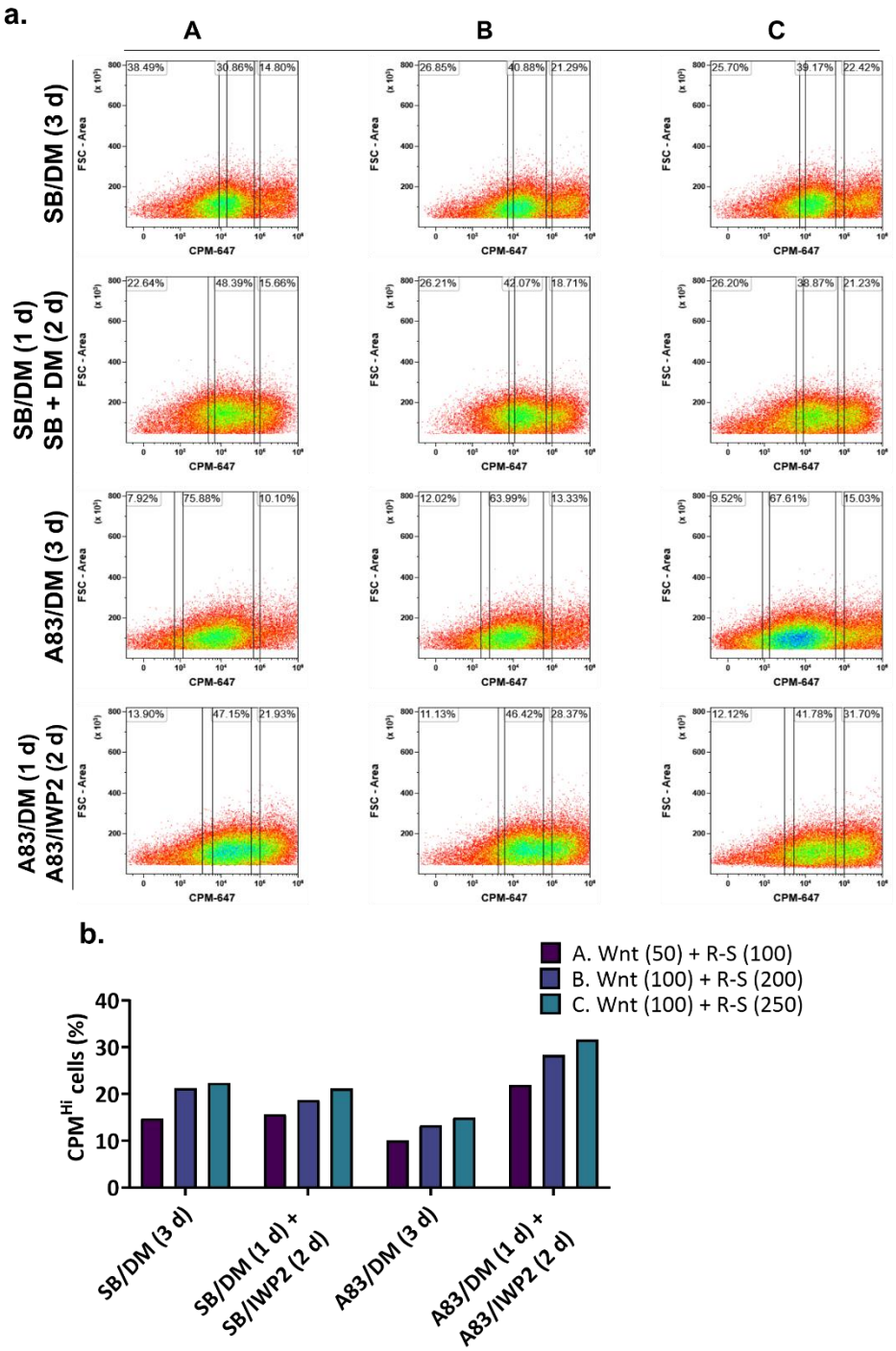


Figure 4-13. Increasing concentrations of recombinant Wnt3a and R-spondin increase induction of CPM(hi) cell numbers.

(a) Flow cytometry analysis of CPM on lung differentiation day 6 for increasing concentrations of Wnt3a and R-spondin via flow cytometry following specification of foregut cells with 4 combinations of small molecule inhibitors. **(b)** Quantification of CPM(hi) cells on lung differentiation day 6 for each condition (N=1). SB = SB431542 (10 µM), DM = Dorsomorphin (2 µM), A83 = A83-01 (1 µM), IWP2 (1 µM).

4.2.7. Removal of retinoic acid and activation of FGF10 signalling impairs lung epithelial progenitor cell differentiation

To evaluate whether RA signalling is required continuously or just transiently for AFE to acquire competence for lung specification, we assessed the effect of a 2-day RA treatment on the induction of CPM(hi) cells in XFL medium. Lung differentiation was performed with our optimised concentrations of WNT3a and R-spondin and continuous RA signalling was included as a control. We repeated all foregut patterning conditions prior to lung induction and also included a condition where RA was replaced after a 2-day treatment with FGF10 treatment. CPM(hi) cells were quantified by flow cytometry on day 6 of lung differentiation. For all foregut conditions, continued retinoic acid signalling led to the highest percentage of CPM(hi) cells, an effect that was slightly reduced by removing retinoic acid after 2 days and further diminished by activating FGF10 signalling (**Fig. 4-14a**). As previously shown, dual-SMAD inhibition with LDN193189 treatment (for BMP inhibition) for 3 days led to complete loss of the CPM(hi) cell population across all 3 lung progenitor conditions. Induction of CPM(hi) cells was recovered by combining dual-SMAD inhibition via LDN193189 with WNT inhibition during foregut patterning, but the percentage of CPM(hi) cells was still <20% (**Fig. 4-14a**). Dual-SMAD inhibition via SB431542 and dorsomorphin for 3 days with continued RA signalling during lung specification remained the optimal condition to induce the highest percentage of CPM(hi) cells (>53%) (**Fig. 4-14a**). These findings indicate that a pulse of RA stimulation is sufficient to prime AFE and direct cells to a lung fate with continued WNT and BMP signalling in XFL medium. However, continued RA signalling induces the highest percentage of CPM(hi) cells, an effect impaired by supplementing FGF10. Overall, dual-SMAD inhibition for 3 days followed by lung differentiation via WNT3a, R-spondin, BMP4 and RA was the optimal differentiation condition (**Fig. 4-14b**). Therefore, we decided to continue with this protocol in XFL medium for future experiments.

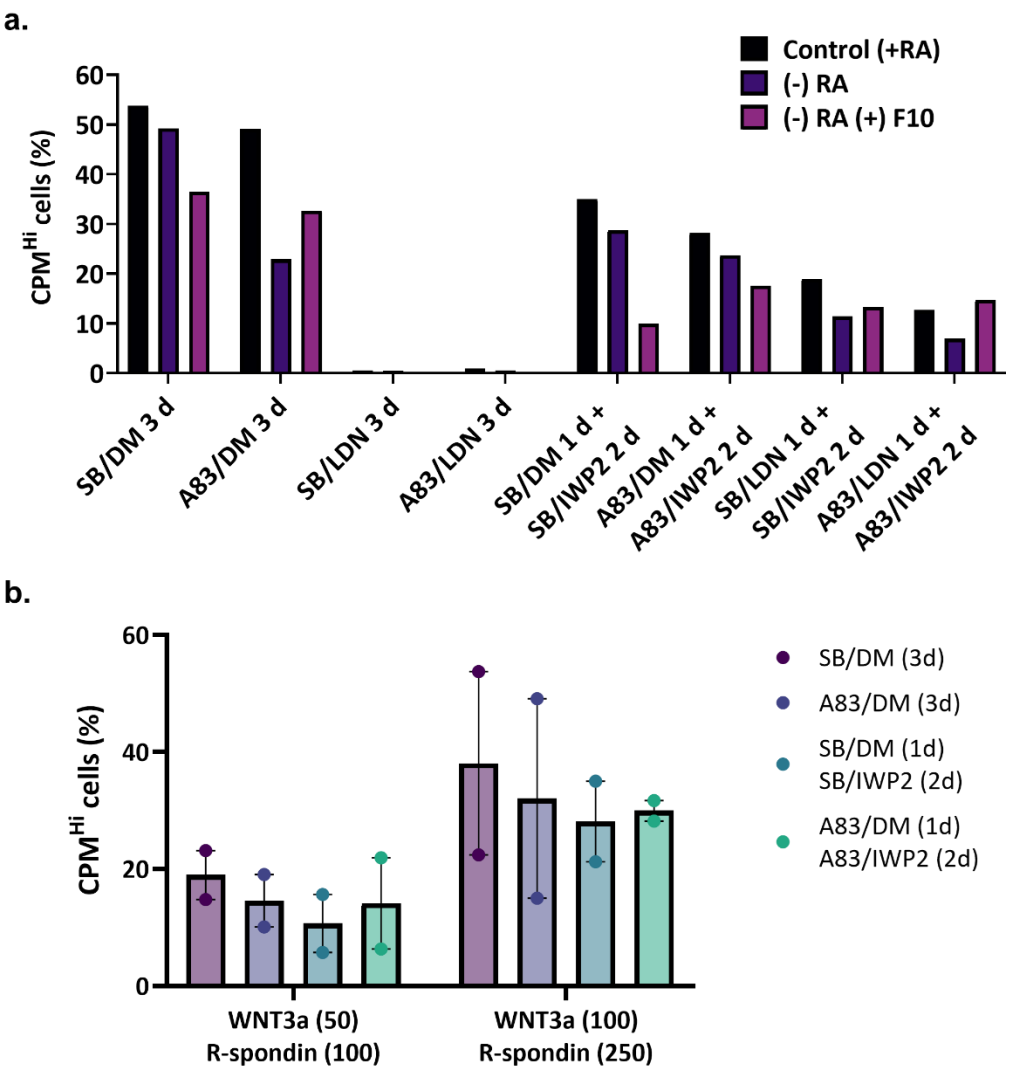


Figure 4-14. Efficiency of differentiation to CPM(high) lung progenitor cells is reduced with removal of retinoic acid and addition of FGF10.

(a) Quantification of CPM(high) cells on day 6 of lung progenitor cell specification via flow cytometry comparing 8 conditions of foregut patterning prior to lung cell differentiation comparing control conditions to removal of retinoic acid and addition of FGF10 (N=1). **(b)** Comparison of the percentage of CPM(hi) cells after treatment with different combinations of small molecule inhibitors during foregut patterning and lung differentiation with WNT3a and R-spondin. Bars reflect mean ± SEM (N=2).

4.2.8. CPM(hi) cells generated in defined medium with an optimised differentiation protocol express a lung progenitor gene signature

To confirm the lung identity of CPM(hi) cells specified in XFL medium by recombinant WNT3a and R-spondin, we sorted CPM populations and analysed expression of lung progenitor markers via qRT-PCR. We compared expression profiles of CPM(hi) and CPM+ cells after AFE patterning with dual-SMAD inhibition via SB432542 and dorsomorphin treatment for 3 days, or dual-SMAD inhibition for 1 day and TGFβ and Wnt inhibition (IWP2) for 2 days, followed by our optimised lung progenitor differentiation. These new differentiation conditions result in expression of CPM in the majority of cells meaning too few CPM- cells were isolated for analysis (**Fig. 4-15a**). The lung epithelial progenitor genes *NKX2.1* and *CPM*, type I AEC genes *AGER* and *HOPX* and type 2 AEC gene *SLC34A2* were all upregulated in the CPM(hi) cell fraction and expressed highest with dual-SMAD inhibition for 3 days (**Fig. 4-15b**). In contrast, *SFTPC* was not expressed in either cell population. Genes related to hepatocyte function, *ALB*, *AFP* and *TTR*, were upregulated CPM(hi) and CPM+ populations treated with SB431542, dorsomorphin and IWP2 during foregut specification but were expressed at low levels in CPM(hi) cells treated with dual-SMAD inhibition for 3 days. Expression of *TTR* was upregulated in both CPM populations for both foregut conditions. Expression of *CKIT*, previously used to enrich for lung progenitor cells, was higher in CPM+ cells, and *Villin* was upregulated in CPM(hi) cells across both foregut conditions (**Fig. 4-15b**). These data indicated the CPM(hi) population generated with our optimised protocol was enriched for cells with a lung epithelial progenitor gene signature with little expression of non-lung endoderm genes. Immunofluorescence imaging further confirmed FACS isolated CPM(hi) cells differentiated in XFL medium expressed NKX2.1 protein when seeded in 2D culture (**Fig. 4-16a**).

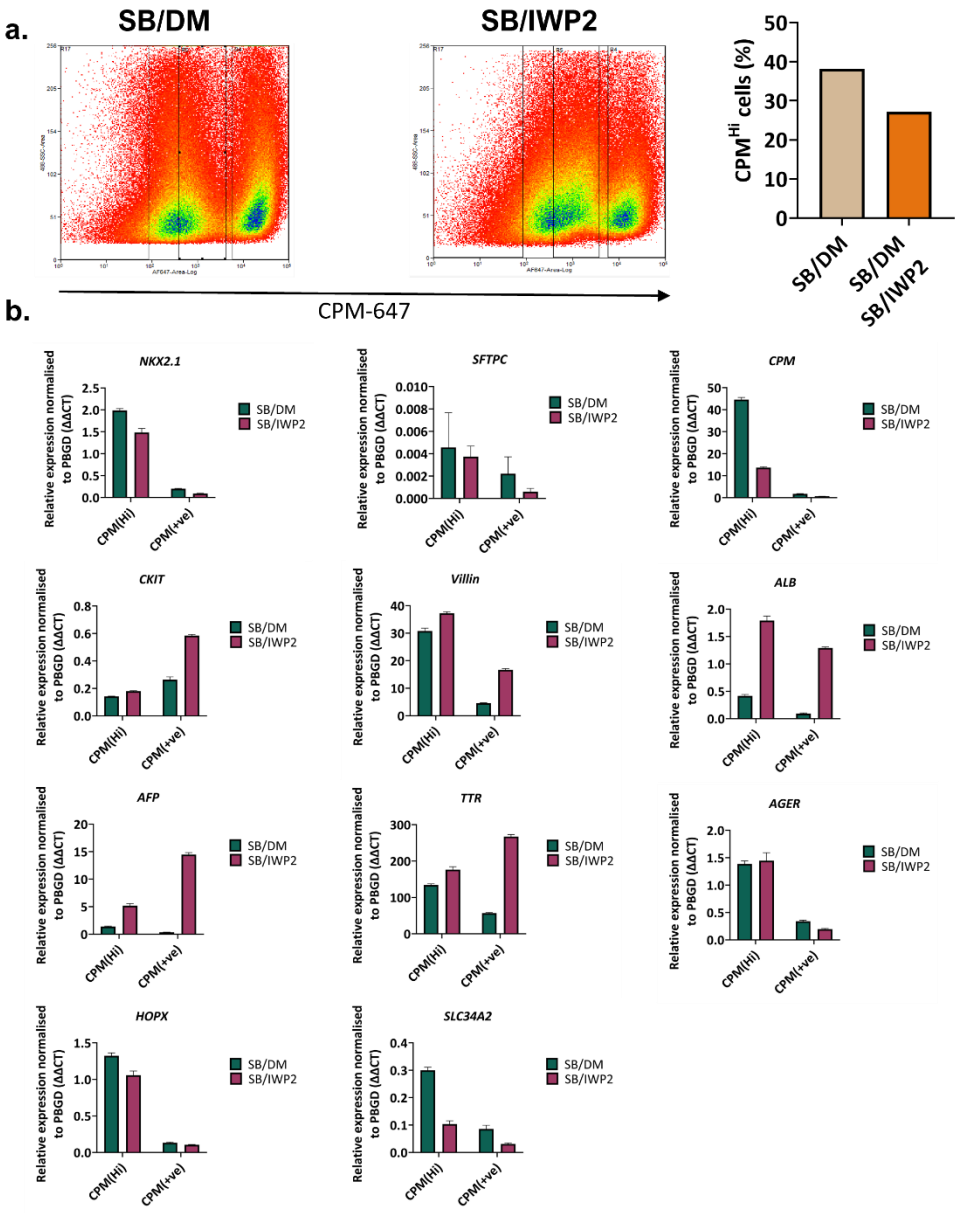


Figure 4-15. CPM(high) cells specified by dual-SMAD inhibition alone or with Wnt inhibition before lung differentiation express markers of lung progenitor cells.

(a) FACS analysis of BE.31 ($SFTPC^{WT/WT}$) lung day 6 cells after dual SMAD inhibition or dual SMAD inhibition followed by Wnt inhibition showing sorting gate for CPM(Hi) cells with CPM(hi) cell percentage quantified. **(b)** Expression of lung and liver genes in CPM(high) and CPM+ sorted populations after foregut patterning with dual-SMAD inhibition alone or with dual-SMAD inhibition preceding Wnt inhibition before differentiation to lung progenitor cells with an optimised protocol. Bars represent mean normalised expression to the housekeeping gene PBGD (ΔΔCt) ± SEM (N=1, 3 technical replicates).

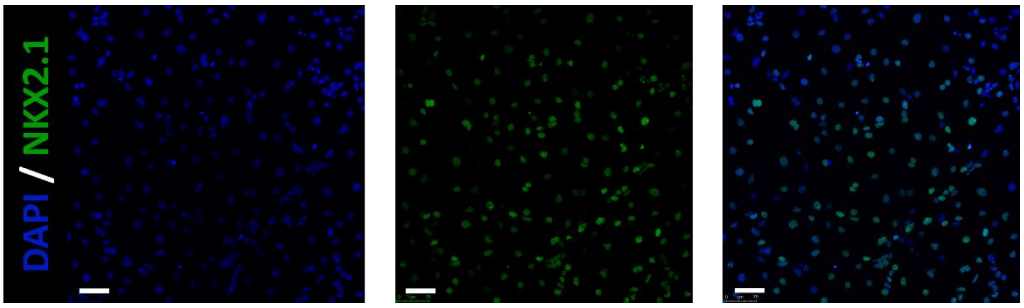


Figure 4-16. CPM(hi) cells differentiated in XFL medium express NKX2.1.

Immunofluorescence staining for NKX2.1 in FACS sorted BE.31 ($SFTPC^{WT/WT}$) CPM(hi) cells generated in XFL medium and seeded in 2D culture. Scale bars, 75 µm.

4.2.9. Optimised differentiation with recombinant WNT3a and R-spondin in XFL medium is reproducible

Once we had established our optimised directed differentiation parameters using commercially available WNT3a (**R&D Systems; Table 2-2**) and R-spondin (**PeproTech; Table 2-2**) in place of CHIR99021, we wanted to determine the optimal DE colony seeding ratio given the importance of culture density of differentiation efficiency. We observed robust induction of CPM(hi) cells on day 6 of lung differentiation for all three seeding ratios tested. DE cells split at 1:12, 1:15 and 1:18 led to induction of ~39%, ~32% and ~37% CPM(hi) cells, respectively (**Fig. 4-17a**). To maintain consistency between differentiations we aimed to perform FACS on day 6 of lung differentiation but this was not always possible. We showed that under our optimised conditions CPM(hi) cells could still be generated at earlier time points. On day 4 the mean percentage of CPM(hi) cells was ~12% and this increased to ~41% on day 5 (**Fig. 4-17b**). In the presence of CHIR99021 during lung differentiation, the mean percentage of CPM(hi) cells on day 6 was ~16% (**Fig. 4-17c**). We assessed the reproducibility of our optimised protocol and demonstrated the induction of CPM(hi) cells across 4 cell lines. For BE.31 ($SFTPC^{WT/WT}$) and ABE.27 ($SFTPC^{MUT/MUT}$) hiPSCs the mean percentage of CPM(hi) cells induced was ~31% and ~35%, and we could achieve a maximum percentage of ~48% and ~69% CPM(hi) cells, respectively (**Fig. 4-17d**). Across other cell lines analysed the efficiency was reduced and for LOPCK hiPSCs the mean percentage of CPM(hi) cells was only 4% (**Fig. 4-17d**). These findings suggest our optimise protocol

could specify lung progenitor cells across multiple cell lines, but some minor adjustments might be required for more resistant cell lines.

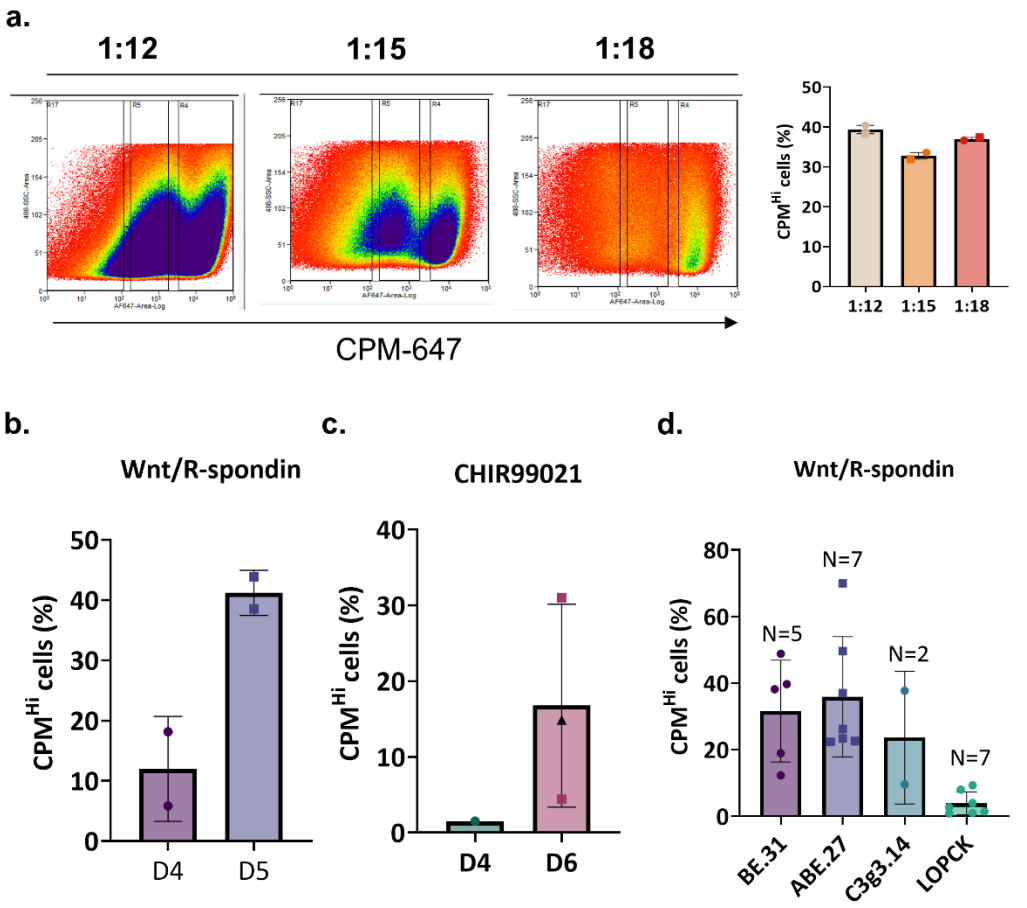


Figure 4-17. Specification of CPM cells on Matrigel and laminin in animal free medium.

(a) FACS analysis of BE.31 (SFTPC$^{WT/WT}$) lung progenitor cells for CPM after single cell splitting of DE at 3 increasing ratios and quantification of CPM(hi) cells. (Squares represent quantification on lung day 5 for one biological replicate, and circles represent quantification on lung day 4 for another biological replicate). **(b).** Quantification of BE.31 (SFTPC$^{WT/WT}$) CPM(hi) cells after lung differentiation with Wnt/R-spondin. **(c)** Quantification of CPM(hi) cells after differentiation with CHIR (Black data point = ABE.27 (SFTPC$^{MUT/MUT}$), all other data points represent BE.31 (SFTPC$^{WT/WT}$) hiPSCs). **(d)** Quantification of CPM(hi) cells for 4 cell lines on day 6 of lung differentiation after specification with Wnt/R-spondin. Bars represent mean ± SEM.

4.2.10. The percentage of CPM(hi) cells peaks early during lung differentiation

To interrogate the temporal kinetics of lung progenitor specification and CPM expression we repeated our optimised differentiation protocol and assessed CPM via flow cytometry across lung differentiation. Following foregut patterning with 3 days of dual-SMAD inhibition or dual-SMAD and WNT inhibition, there was robust generation of CPM(hi) cells on day 3 of lung specification (**Fig. 4-18a**). Comparison of CPM(hi) cell percentage every 2 days from lung day 2 to 8 revealed the onset of lung progenitor cells as early as day 2 of the lung stage of differentiation. Based on the flow cytometry profiles these cells shifted into a high-CPM expressing population and peaked on day 4 (**Fig. 4-18b**). On day 4 the mean percentage of CPM(hi) cells was ~43%, reduced to ~37% on day 6 and dropped below ~20% on day 8 (N=2) (**Fig. 4-18b, c**). Almost all cells expressed CPM on day 4 and 6, and the kinetics of CPM expression for BE.31 ($SFTPC^{WT/WT}$) and ABE.27 ($SFTPC^{MUT/MUT}$) hiPSCs were the same (**Fig. 4-18b, c**). Although the greatest percentage of CPM(hi) cells was induced on day 4, for enrichment and further culture a large number of cells are required, so the additional number of lung progenitor cells that can be collected after 2 more days of proliferation on day 6 negate the small difference in CPM(hi) cell percentage. For this reason, we continued with 6 days of lung specification.

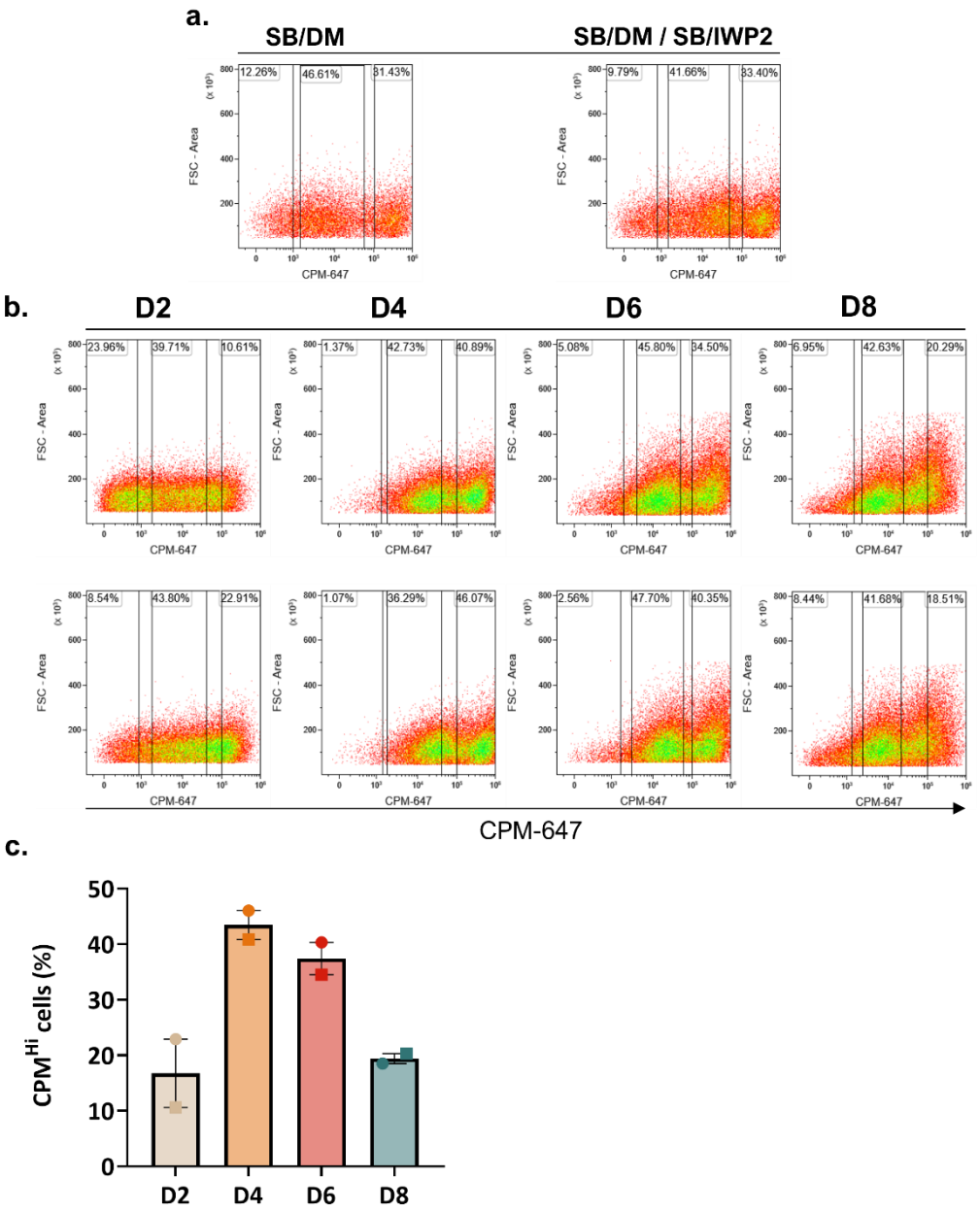


Figure 4-18. Induction of CPM(Hi) cells during early differentiation.

(a) Flow cytometry analysis of BE.31 ($SFTPC^{WT/WT}$) cells on day 3 of lung differentiation for CPM after treatment of DE cells with dual-SMAD inhibition or dual-SMAD inhibition followed by Wnt inhibition. Time course analysis of CPM expression during 8 days of lung differentiation via flow cytometry for **(b)** BE.31 ($SFTPC^{WT/WT}$) (top panel) ABE.27 ($SFTPC^{MUT/MUT}$) (bottom panel) hiPSCs, and **(c)** quantification of CPM(hi) cell numbers on each day with bars representing mean ± SEM (N=2) (Circles = ABE.27 ($SFTPC^{MUT/MUT}$) and squares = BE.31 ($SFTPC^{WT/WT}$) cell lines).

4.2.11. Defining the transcriptomic profile of hPSC-derived lung epithelial progenitor cells

Once we had established our optimal differentiation conditions, we sought to define the transcriptomic profile of cells during the specification of hPSC-DE to lung epithelial progenitor cells. We performed a time series analysis using RNA-seq of cells on lung differentiation day 2, 4, 6 and 8 and CPM-, CPM+ and CPM(hi) cell populations. Principal-component analysis revealed that hiPSCs and DE cells, and NaPi2B cells clustered separately from all other samples, and cells during lung differentiation and CPM populations clustered closely (**Fig. 4-19a**). This pattern of clustering was further confirmed by performing hierarchical clustering analysis on the same samples (**Fig. 4-19b**). We did not observe the enrichment of lung progenitor genes in CPM(hi) cells in our analysis of the top DEGs, although *TTR* and *COL2A1* were expressed across the entirety of lung differentiation (**Fig. 4-20a**). Profiling of our selected lung-related genes revealed the robust expression of lung epithelial progenitor transcripts such as *NKX2.1*, *CPM*, *FOXA2* and *ID2* in CPM(hi) cells compared to CPM+, CPM- and unsorted lung populations (**Fig. 4-20b**). The human distal lung bud tip markers *SOX2* and *SOX9* were expressed in all lung samples along with high expression of the epithelial cell marker *EPCAM*, as expected (**Fig. 4-20b**). During development, cells of the developing thyroid also emerge from ventral anterior foregut endoderm and express NKX2.1, but we did not observe expression of the thyroid marker *PAX8* at any stage of our differentiation. Transcripts related to type 2 AECs such as the surfactant proteins were also not expressed, as expected at this stage of differentiation (**Fig. 4-20b**). A volcano plot of DEGs in undifferentiated hiPSCs compared to CPM(hi) cells further confirmed the enrichment of *NKX2.1* and *TTR*, the lung bud endoderm marker *FOXP2* and signalling genes such as *SHH* in CPM(hi) cells (**Fig. 4-21a**). Together, these data confirm that our CPM(hi) population is enriched for lung epithelial progenitor cell transcripts.

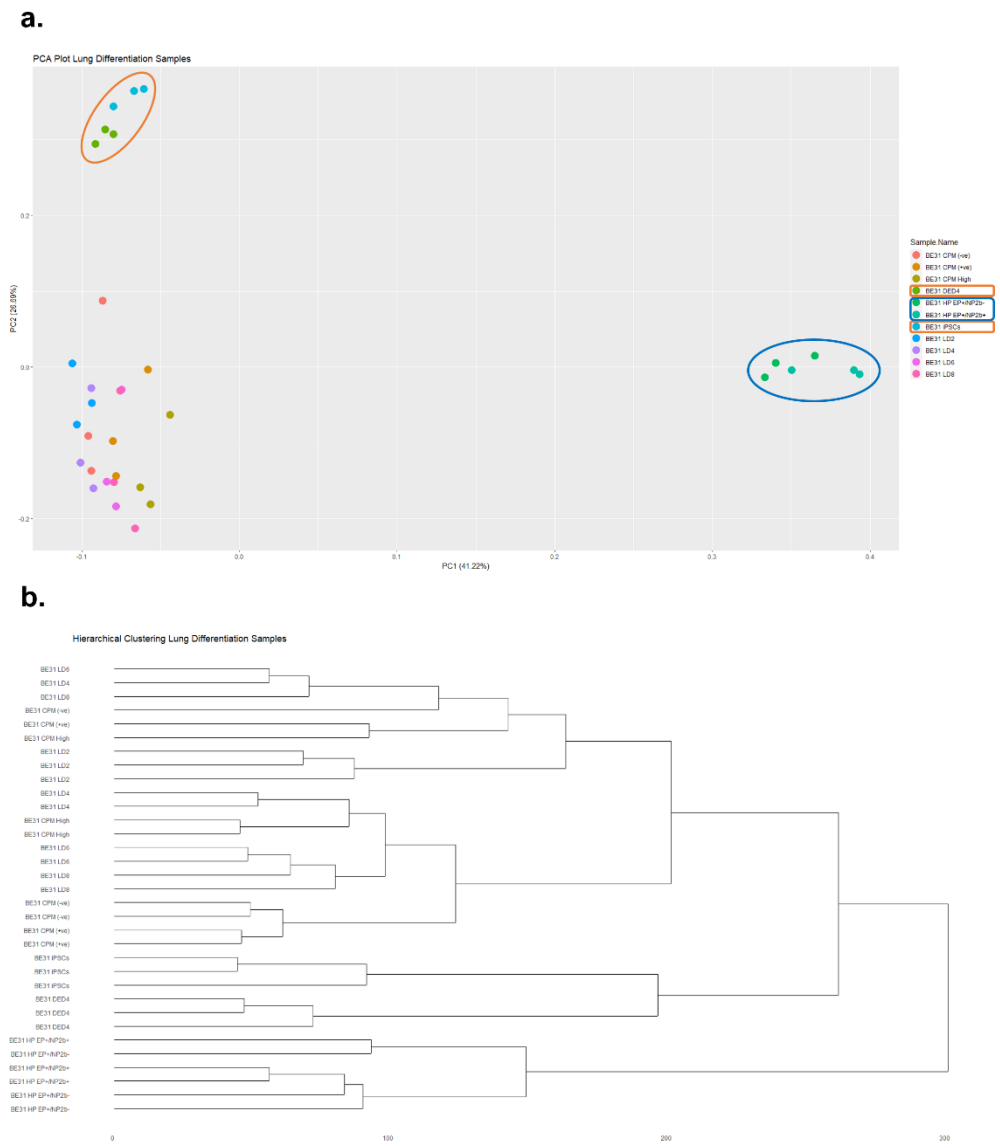


Figure 4-19. Transcriptomic profiling of hiPSC-derived cell types in at developmental major stages of differentiation to type 2 AECs.

(a) Principal component analysis of gene expression variance across all samples during lung differentiation. **(b)** Hierarchical clustering of cell types at critical timepoints of lung differentiation.

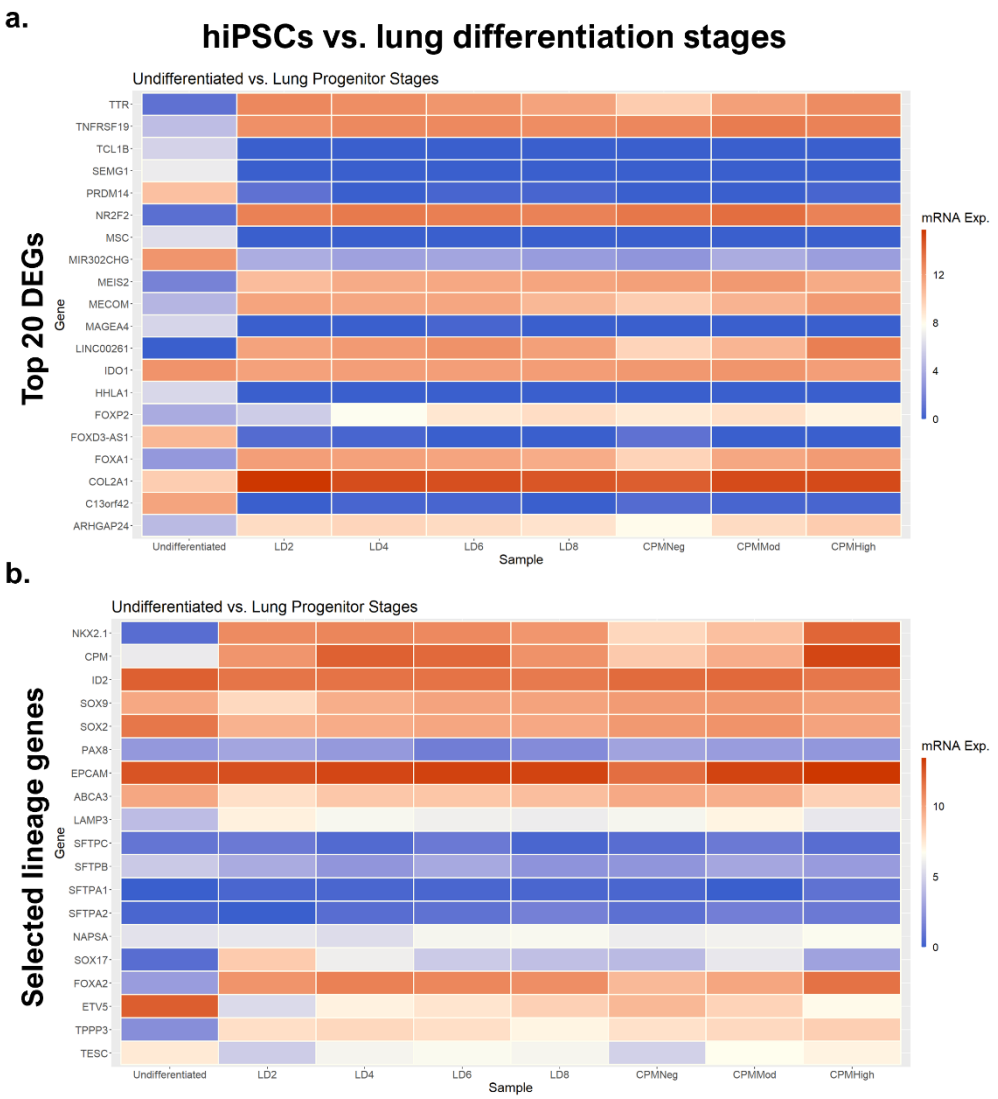


Figure 4-20. Gene expression profiling of cells during lung differentiation and CPM populations.

(a) Heatmap of the top 20 DEGs in undifferentiated hiPSCs versus cells on sequential days of lung progenitor differentiation and isolated CPM-, CPM+ and CPM(hi) populations. **(b)** Heatmap of the expression of selected lineage markers in undifferentiated hiPSCs versus cells on sequential days of lung progenitor differentiation and isolated CPM-, CPM+ and CPM(hi) populations. All DEGs displayed are statistically significant to $P < 0.05$.

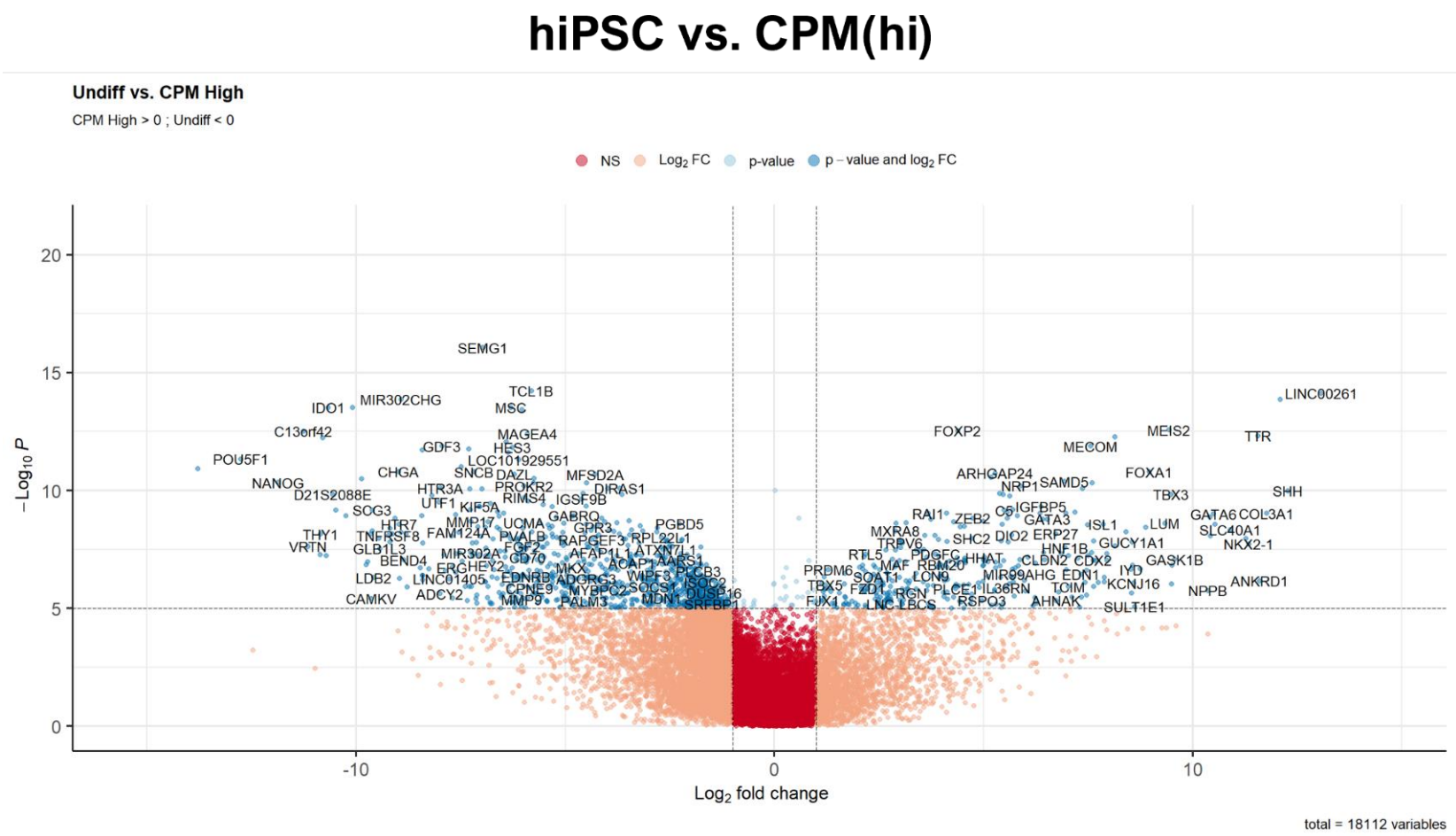


Figure 4-21. Transcriptomic analysis reveals enrichment of lung progenitor cell transcripts in CPM(hi) cells.

(a) Volcano plot of differentially expressed genes in undifferentiated hiPSCs versus isolated CPM(hi)cells. Statistically significant ($P \leq 0.05$) changes are highlighted.

4.2.12. CPM(hi) cells give rise to NAPI2B+/EPCAM+ type 2 AEC-like cells in XFL medium

We sought to determine whether XFL medium could support the differentiation and self-assembly of CPM(hi) cells into organoids containing NaPi2B+/EPCAM+ cells. In XFL medium, CPM(hi) cells resuspended in Matrigel™ droplets can self-assemble into 3D organoids and retain the capacity to form organoids from single cells over several passages (**Fig. 4-22a**). After growth in the presence of the lung maturation factors previously utilised for organoid culture, we observed the emergence of a NaPi2B+/EPCAM+ population (**Fig. 4-22b**). We showed NaPi2B+/EPCAM+ enriched organoids in XFL medium express and secrete SFTPC as determined by immunofluorescence imaging (**Fig. 4-22c**).

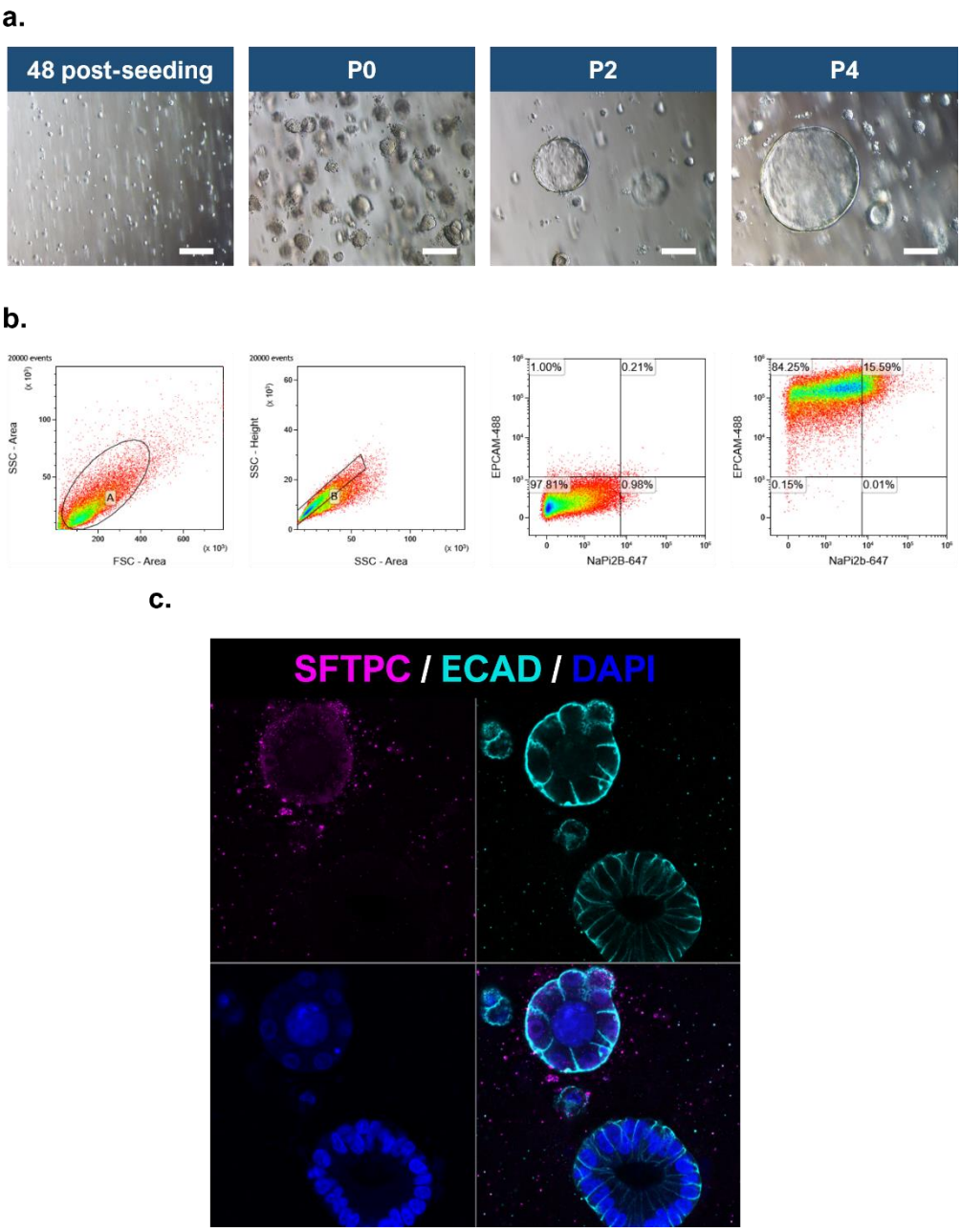


Figure 4-22. CPM(hi) cells can give rise to NaPi2B+/EPCAM+ cells in organoids maintained in XFL medium.

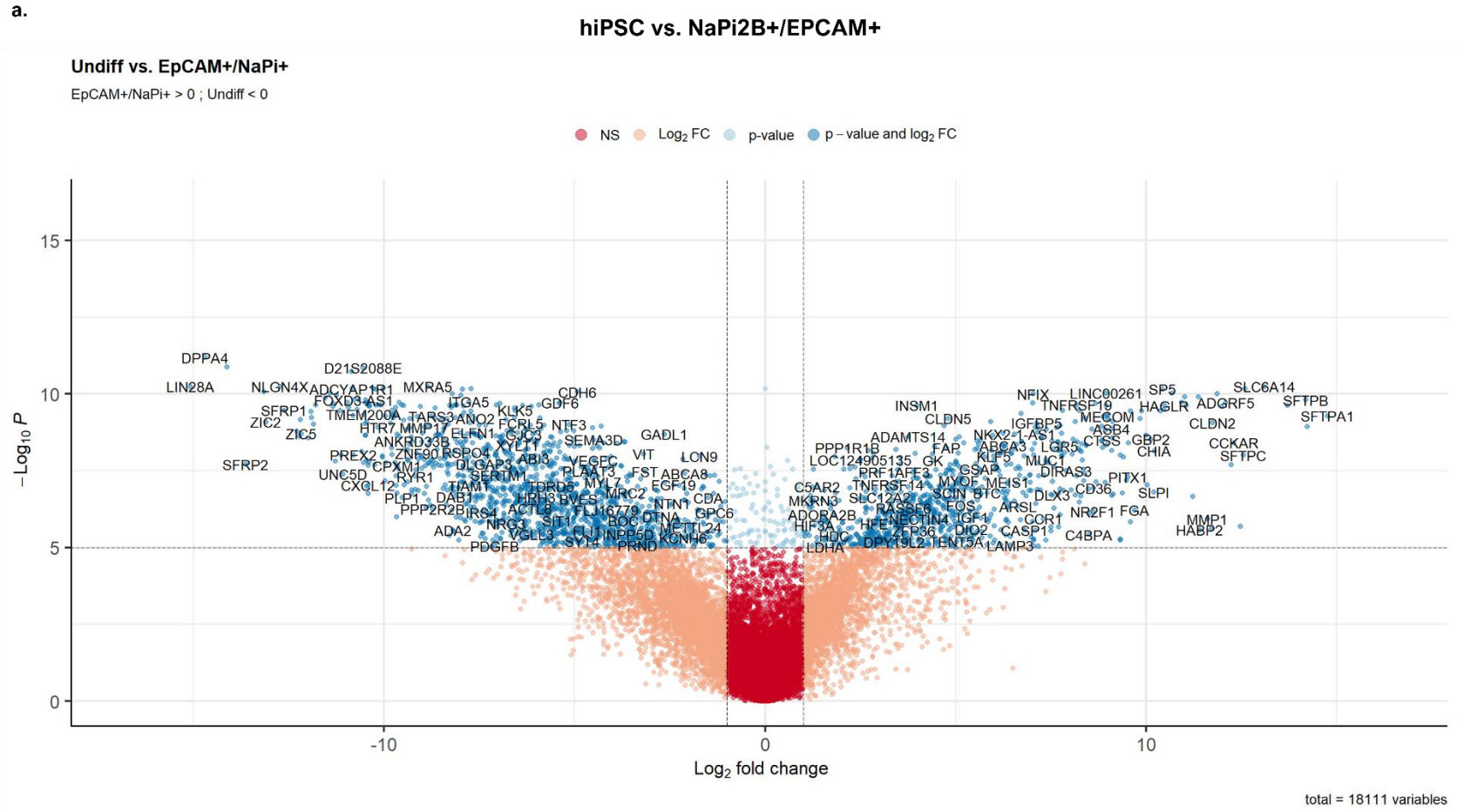
(a) Representative light microscope images of CPM(hi) cells seeded into Matrigel droplets and expanded as organoids in XFL medium and lung maturation growth factors. **(b)** Representative flow cytometry plots showing the strategy to isolate NaPi2B+/EPCAM+ cells. **(c)** Immunofluorescence staining for SFTPC and ECAD in BE.31 ($SFTPC^{WT/WT}$) type 2 AEC organoids generated in XFL medium. (Figure 22c was performed by Dr Sara Cuevas Ocaña).

4.2.13. Characterising the transcriptome of NaPi2B+/EPCAM+ cells

As with previous stages of our differentiation protocol, we performed RNA-seq to define the transcriptomic profile of late-stage NaPi2B-isolated populations that were maintained in XFL medium. Compared to hiPSCs, among the top DEGs in NaPi2B+ cells were pulmonary surfactant protein genes including *SFTPC*, *SFTPB*, *SFTPA1* and markers related to surfactant processing of type 2 AECs such as *LAMP3* and *ABCA3* (**Fig. 4-23a**). When compared to NaPi2B-/EPCAM+ cells, the type 2 AEC genes *SFTPB*, *SFTPD* and *SLC34A2*, and type 1 AEC marker *HOPX* were enriched in NaPi2B+/EPCAM+ cells (**Fig. 4-23b**). Differential gene expression analysis revealed an enrichment of *NKX2.1* and *SFTPB* in both NaPi2B populations with slightly higher *SFTPB* expression in NaPi2B+ cells (**Fig. 4-24a**).

We then compared DEGs in undifferentiated hiPSCs to CPM(hi) lung progenitor cells and NaPi2B populations. From our selected type 2 AEC gene panel, we observed upregulation of genes related to surfactant processing including SFTPB, SFTPC ABCA3, the lipid processing gene LPCAT1, and transcripts associated with maturation of type 2 AECs such as NAPSA, SLC34A2, SFTPA1 and SFTPA2 in both NaPi2B populations, but expression was enriched in NaPi2B+ cells (**Fig. 4-24b**). This suggests our NaPi2B- cells are still enriched for type 2 AECs but might be slightly less mature. Profiling of our selected type 1 AEC markers in the same populations showed no strong type 1 AEC gene signature in either NaPi2B cell fractions although HOPX was upregulated in NaPi2B+ cells. The type 1 AEC genes CAV1, PDPN, SPOCK2, HOPX and CLIC5 were upregulated in CPM(hi) cells indicating the potential of lung progenitor cells to differentiate into type 1 AECs. In CPM(hi) cells, KRT7 was also highly upregulated (**Fig. 4-24c**). To determine the potential expression of transcripts related to proximal airway cells, we generated a gene set for basal cells and goblet, club and multiciliated cells and performed analysis for DEGs in the same populations. Well-characterised basal cell markers such as TP63, KRT5 and KRT13 were not expressed in either NaPi2B populations, but KRT19, KRT17 and AQP3 were upregulated (**Fig. 4-25a**). We observed expression of TUBA1A, which encodes for acetylated α-tubulin, across all samples, and CFTR was also expressed in CPM(hi) and NaPi2B+ and NaPi2B- cells (**Fig. 4-25b**).

a.


hiPSC vs. NaPi2B+/EPCAM+
Undiff vs. EpCAM+/NaPi+
EpCAM+/NaPi+ > 0 ; Undiff < 0
NS
Log2 FC
p-value
p – value and log2 FC
−Log10 P
Log2 fold change
total = 18111 variables

b.

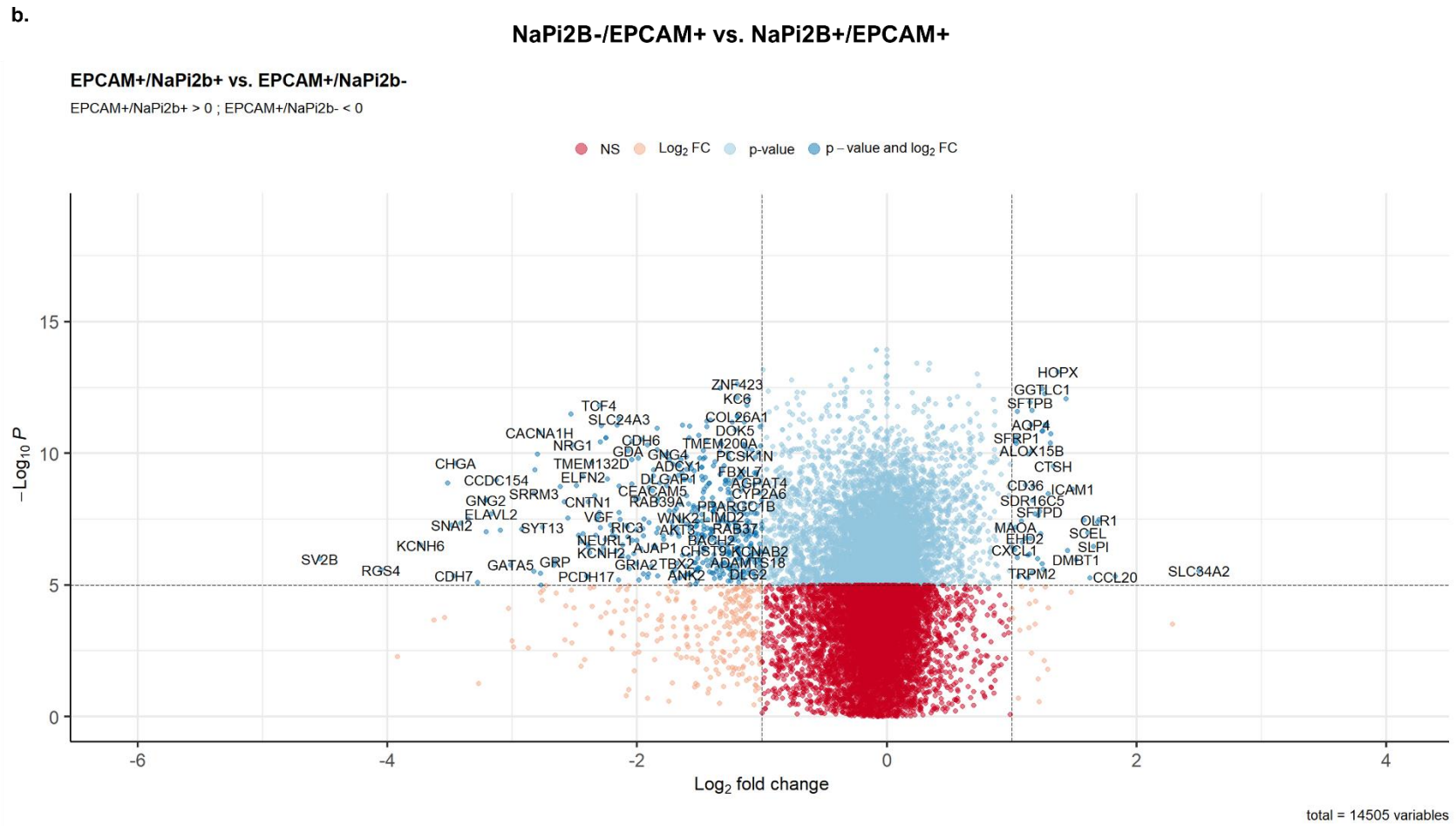

NaPi2B-/EPCAM+ vs. NaPi2B+/EPCAM+
EPCAM+/NaPi2b+ vs. EPCAM+/NaPi2b-
EPCAM+/NaPi2b+ > 0 ; EPCAM+/NaPi2b- < 0
NS
Log2 FC
p-value
p – value and log2 FC
−Log10 P
Log2 fold change
total = 14505 variables
HOPX
GGTLC1
SFTPB
AQP4
SFRP1
ALOX15B
CTSH
CD36
ICAM1
SDR16C5
SFTPD
OLR1
MAOA
SCEL
EHD2
SLPI
CXCL1
DMBT1
TRPM2
CCL20
SLC34A2
ZNF423
KC6
TOF4
SLC24A3
COL26A1
DOK5
CACNA1H
NRG1
CDH6
TMEM200A
GDA
GNG4
PCSK1N
CHGA
TMEM132D
ADCY1
FBXL7
CCDC154
ELFN2
DLGAP1
AGPAT4
SRRM3
CEACAM5
CYP2A6
GNG2
CNTN1
RAB39A
PPARGC1B
ELAVL2
VGF
WNK2
LIMD2
SNAI2
SYT13
RIC3
AKT3
RAB37
NEURL1
BACH2
KCNH6
AJAP1
CHST9
KCNAB2
SV2B
KCNH2
GRIA2
TBX2
ADAMTS18
RGS4
GATA5
GRP
ANK2
DLG2
CDH7
PCDH17

Figure 4-23. Transcriptomic profiling of isolated NaPi2B+/EPCAM+ cells.

Volcano plot of differentially expressed genes in **(a)** undifferentiated hiPSCs versus isolated NaPi2B+/EPCAM+ cells and **(b)** NaPi2B+/EPCAM+ cells versus NaPi2B-/EPCAM+ cells. Statistically significant ($P \leq 0.05$) changes are highlighted.

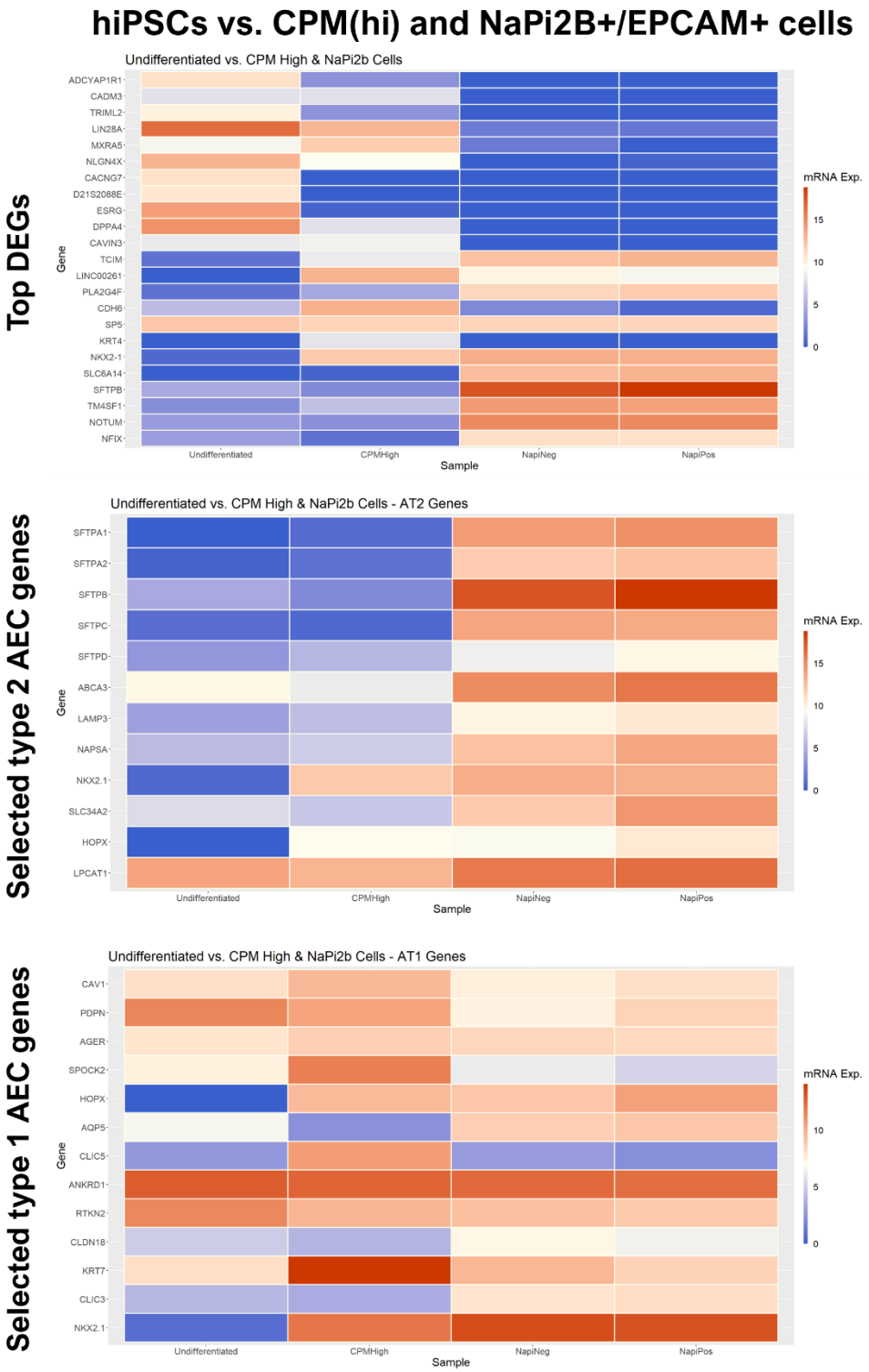


Figure 4-24. Transcriptomic analysis reveals enrichment of type 2 AEC genes in NaPi2B+/EPCAM+ cells.

Heatmaps for the **(a)** top 20 DEGs, **(b)** selected type 2 AEC and **(c)** selected type 1 AEC transcripts in undifferentiated hiPSCs versus CPM(hi) and NaPi2B+/EPCAM+ cells. All DEGs displayed are statistically significant to $P < 0.05$.

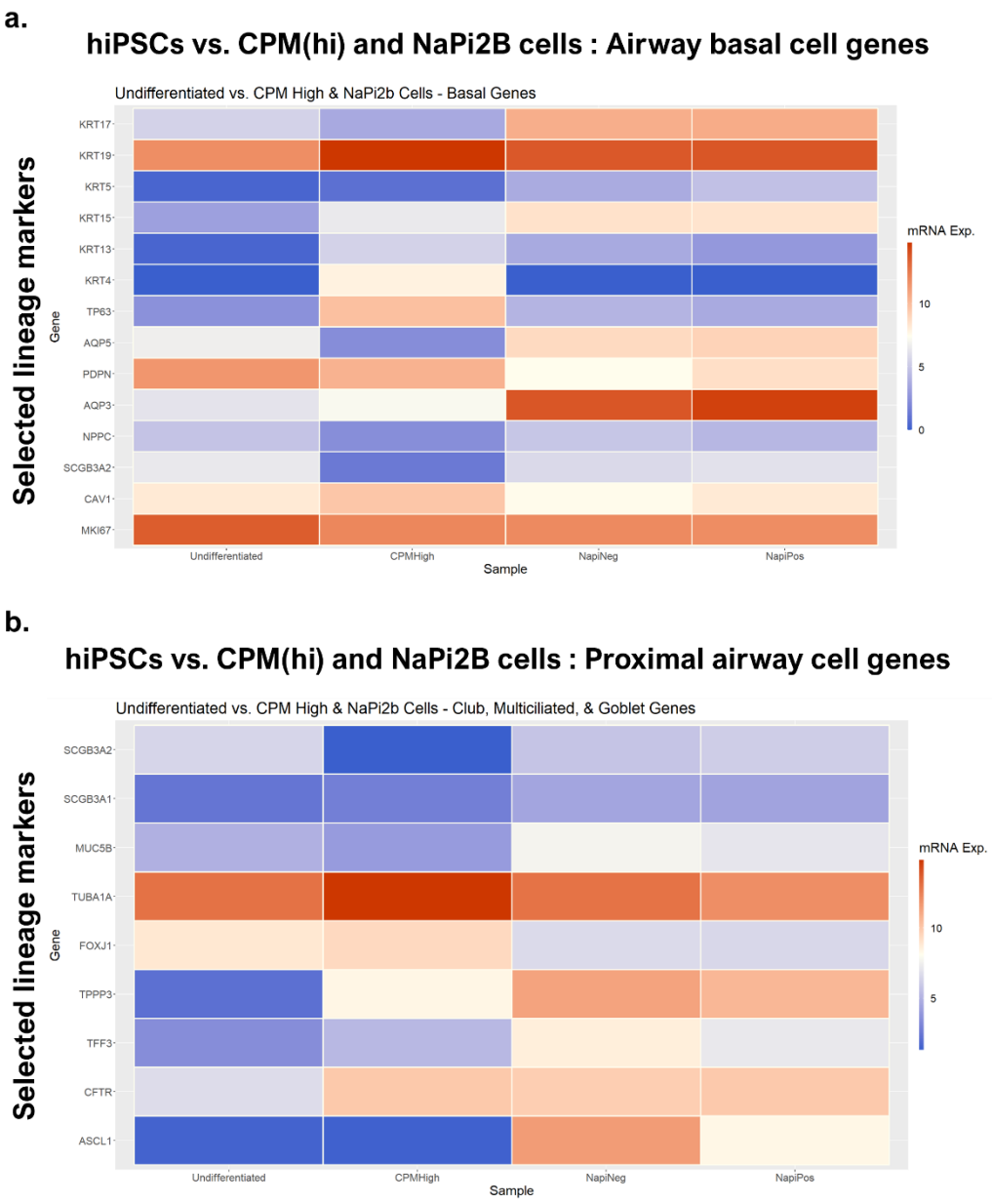


Figure 4-25. NaPi2B+/EPCAM+ cells do not strongly express transcripts related to proximal airway cells.

Heatmaps of the expression of selected **(a)** airway basal cell and **(b)** airway goblet, club and multiciliated cell transcripts. All DEGs displayed are statistically significant to $P < 0.05$.

4.2.14. Comparison of the transcriptomic profiles of NaPi2B populations to SFTPC-tdTomato enriched cells

As further validation of our chemically defined differentiation protocol, we benchmarked the transcriptomic profile of our isolated NaPi2B-isolated populations to previously published RNA-seq data sets of hPSC-derived type 2 AECs isolated using a SFTPC-tdTomato fluorescence reporter system (Alysandratos et al., 2021). Differential gene expression analysis revealed enrichment of *NKX2.1* across both SFTPC-tdTomato and NaPi2B populations, but the type 1 AEC gene *HOPX* was expressed higher in both SFTPC+ and SFTPC- cells (**Fig. 4-26a**). Profiling of our type 2 AEC gene set revealed the expression patterns between NaPi2B and SFTPC populations were comparable (**Fig. 4-26b**). The type 2 AEC transcripts *SFTPB*, *SFTPA2*, *SFTPC* and *LPCAT1* were expressed marginally higher in SFTPC+ cells, but in NaPi2B+ cells *SFTPA1*, *ABCA3* and *SLC34A2* exhibited higher expression (**Fig. 4-26b**).

Technologies like scRNA-seq are beginning to reveal the heterogeneity of cell types that exist in the lung. A population of Wnt-responsive 'signalling' type 2 AECs has recently been identified and has been posited as the progenitor population of the alveoli (Travaglini et al., 2020). We therefore determined gene sets for proliferating and transitional type 2 AECs and Wnt signalling and performed differential gene expression analysis on NaPi2B and SFTPC-tdTomato populations. Markers of proliferating type 2 AECs like *CDK1* and *PRC1* were enriched in SFTPC+ and SFTPC- cells and expressed at lower levels in NaPi2B cells (**Fig. 4-27a**). Both SFTPC populations exhibited upregulation of transitional genes including *RNASE1*, *KRT7* and *SCGB3A2* that were not highly expressed in NaPi2B cells (**Fig. 4-27b**). Several transcripts related to Wnt signalling were expressed in all populations but *WIF1* and *LRP5* were enriched in SFTPC+ and SFTPC- cells (**Fig. 4-27c).** Significant enrichment of type 1 AEC markers in both SFTPC populations compared to NaPi2B cells was also found (**Fig. 4-27d**). Together, these data suggest the transcriptomic profile of NaPi2B+ and SFTPC+ are comparable and both highly enriched for type 2 AEC genes, but SFTPC+ cells might be higher cycling or closer to a transitional state and display higher expression of type 1 AEC markers.

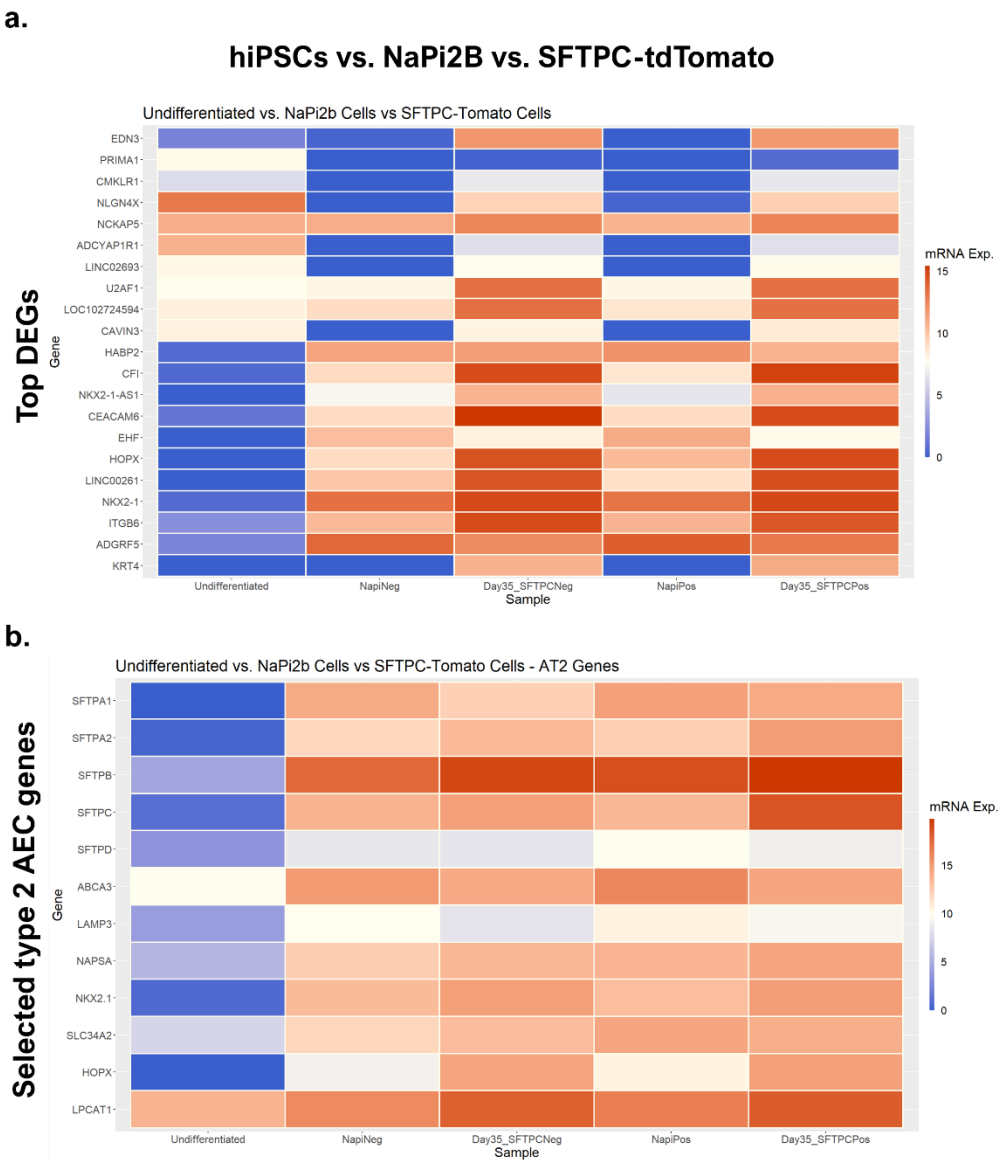


Figure 4-26. NaPi2B+/EPCAM+ cells have a similar transcriptomic profile to hPSC-SFTPC+ cells.

(a) Heatmap of the top DEGs in undifferentiated hiPSCs versus NaPi2B-positive and negative cells versus SFTPC-tdTomato-positive and negative sorted cells from previously published data. **(b)** Heatmap of the expression of selected type 2 AEC transcripts in undifferentiated hiPSCs versus NaPi2B-positive and negative cells versus SFTPC-tdTomato-positive and negative sorted cells. All DEGs displayed are statistically significant to $P < 0.05$.

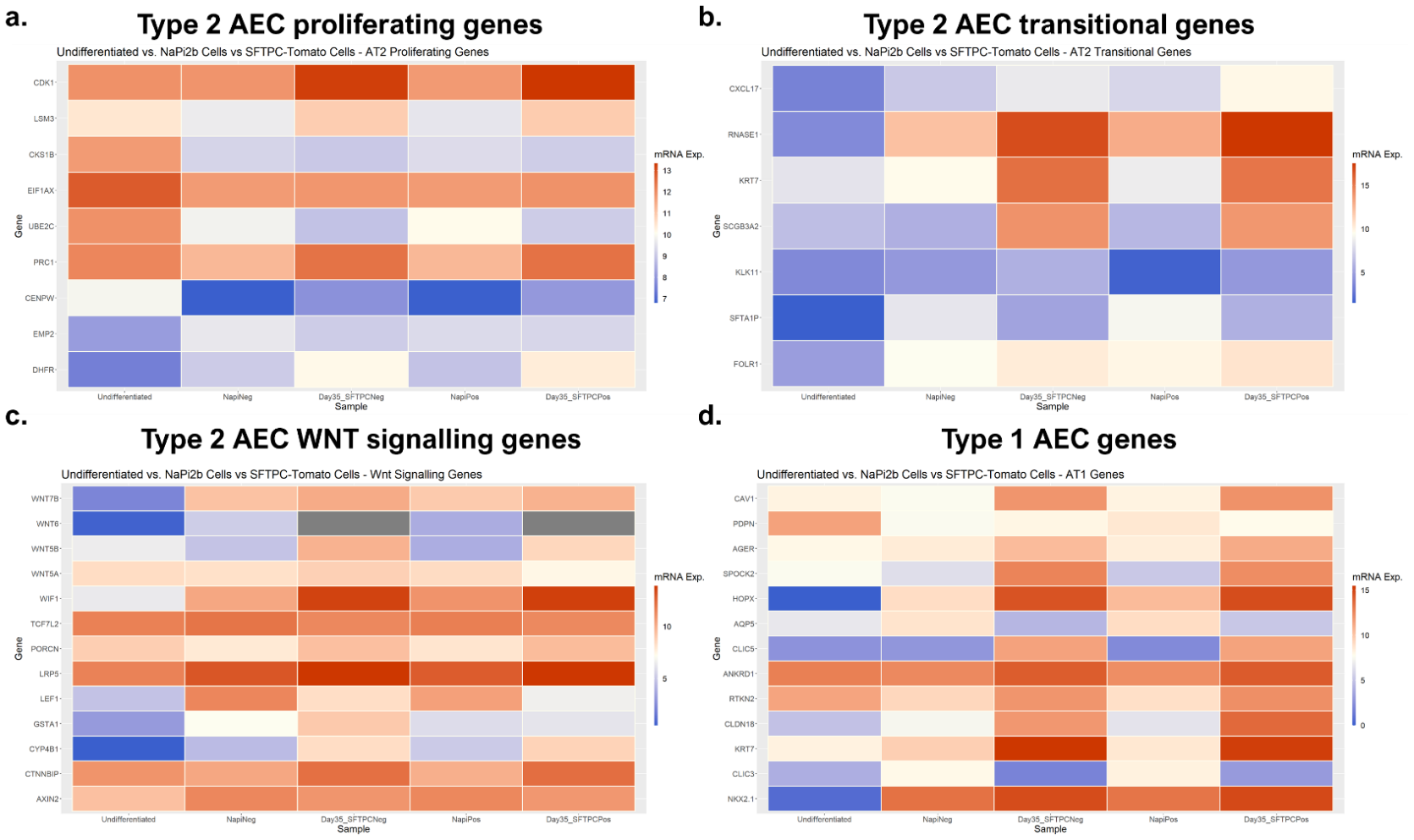

a.
Type 2 AEC proliferating genes
Undifferentiated vs. NaPi2b Cells vs SFTPC-Tomato Cells - AT2 Proliferating Genes
CDK1
LSM3
CKS1B
EIF1AX
UBE2C
PRC1
CENPW
EMP2
DHFR
Gene
mRNA Exp.
13
12
11
10
9
8
7
Undifferentiated
NapiNeg
Day35_SFTPCNeg
NapiPos
Day35_SFTPCPos
Sample
b.
Type 2 AEC transitional genes
Undifferentiated vs. NaPi2b Cells vs SFTPC-Tomato Cells - AT2 Transitional Genes
CXCL17
RNASE1
KRT7
SCGB3A2
KLK11
SFTA1P
FOLR1
Gene
mRNA Exp.
15
10
5
Undifferentiated
NapiNeg
Day35_SFTPCNeg
NapiPos
Day35_SFTPCPos
Sample
c.
Type 2 AEC WNT signalling genes
Undifferentiated vs. NaPi2b Cells vs SFTPC-Tomato Cells - Wnt Signalling Genes
WNT7B
WNT6
WNT5B
WNT5A
WIF1
TCF7L2
PORCN
LRP5
LEF1
GSTA1
CYP4B1
CTNNBIP
AXIN2
Gene
mRNA Exp.
10
5
Undifferentiated
NapiNeg
Day35_SFTPCNeg
NapiPos
Day35_SFTPCPos
Sample
d.
Type 1 AEC genes
Undifferentiated vs. NaPi2b Cells vs SFTPC-Tomato Cells - AT1 Genes
CAV1
PDPN
AGER
SPOCK2
HOPX
AQP5
CLIC5
ANKRD1
RTKN2
CLDN18
KRT7
CLIC3
NKX2.1
Gene
mRNA Exp.
15
10
5
0
Undifferentiated
NapiNeg
Day35_SFTPCNeg
NapiPos
Day35_SFTPCPos
Sample

Figure 4-27. Gene expression analysis of type 2 AEC cell states and type 1 AEC genes in NaPi2B- and SFTPCtdTomato-isolated populations.

Heatmap of the expression of selected type 2 AEC **(a)** proliferating, **(b)** transitional and **(c)** WNT signalling transcripts, and **(d)** type 1 AEC genes in undifferentiated hiPSCs versus NaPi2B-positive and negative cells versus SFTPC-tdTomato-positive and negative sorted cells. All DEGs displayed are statistically significant to $P < 0.05$.

4.2.15. Manipulation of TGF-β and BMP signalling impacts the specification of NAPI2B+ cells in organoid culture

There is currently only one condition that has been described for the differentiation and maintenance of hiPSC-type 2 AECs. It includes WNT and FGF signalling, and factors related to maturation of type 2 AECs. We demonstrated this condition could induce cells with a transcriptomic signature of type 2 AECs, but emerging roles for other signalling pathways including BMP and TGFβ pathways in type 2 AEC specification are now beginning to be identified. We reasoned that further manipulation of these major signalling factors involved in lung alveolar development might improve the efficiency of induction and maintenance of type 2 AECs in our organoid system. Inhibitors of SMAD signalling is routinely used to culture and maintain the phenotype of human adult primary type 2 AECs *in vitro*. We therefore tested 5 growth conditions for organoid culture: our standard control protocol (CHIR, FGF7, FGF10, Dexamethasone, IBMX, cAMP) and combinations of CHIR99021 and FGF7 with TGF-β inhibition alone (condition 2), CHIR99021 and FGF7 with dual-SMAD inhibition (condition 3), and CHIR99021 and FGF7 with TGF-β inhibition and BMP signalling activation (condition 4 and 5) in XFL medium. We seeded 20,000 CPM(hi) cells and generated 3D organoids that were maintained over several passages without enrichment for NaPi2B+/EPCAM+ cells but continued single cell splitting and re-seeding. We assessed the percentage of NAPI2B+/EPCAM+ cells at each passage via flow cytometry. Under control conditions, cells formed typical large, cystic organoids but for this particular experiment we observed a lack of proliferation and beyond passage 0 there were only enough cells to maintain the culture, not for flow cytometry assessment (**Fig. 4-28a**). In contrast, cells grown with CHIR, FGF7 and SB431542 exhibited a rounded, compacted morphology during earlier passages (**Fig. 4-28a**). By passage 2 they became cystic with a thickened epithelium and continued culture led to lobular-shaped organoids with branches that coincided with the loss of NAPI2B+EPCAM+ cells (**Fig. 4-28a, b**). In the presence of dual-SMAD inhibition (condition 3), organoids remained small and rounded across all passages. A similar morphology was observed as with BMP activation, however, in the presence of BMP4 cells proliferated very slowly and could not be maintained beyond passage 1 or passage 3 for 50 ng/ml BMP4 or 10 ng/ml BMP4, respectively (**Fig. 4-28a**). In contrast to other conditions, the percentage of NAPI2B+/EPCAM+ cells in condition

3 continued to increase up to passage 2 and subsequently decreased before peaking after 5 passages with >80% of total cells co-expressing NaPi2B/EPCAM (**Fig. 4-28b**). These findings indicated in the presence of Wnt and FGF signalling with dual-SMAD inhibition organoids are gradually purified to a near homogenous NAPI2B+/EPCAM+ population without enrichment via cell sorting.

We repeated this condition but added EGF to stimulate cell proliferation to promote the growth of organoids in XFL medium. This time we did not track the NaPi2B+/EPCAM+ but we observed via qRT-PCR the expression of lung and non-lung endoderm genes over several passages without NaPi2B+/EPCAM+ enrichment. The gene encoding NaPi2B (*SLC34A2*) was enriched at early passages, but its expression was lost with continued culture. The basal cell marker *KRT5* was similarly highly upregulated at P2 but its expression was lost beyond P3. The lung progenitor cell genes *NKX2.1* and *CPM*, and *SFTPC* were not expressed at all. Markers of hindgut/intestinal cells (*CDX2*, *Villin*) were highly upregulated across all passages (**Fig. 4-29a**). These findings suggest that under these conditions, cells expressing airway basal markers emerge but with continued culture without FACS purification this population is lost, and contaminating hindgut lineages emerge.

To interrogate these findings further we repeated the experiment but tracked and isolated the NaPi2B+/EPCAM+ population. The addition of EGF led to larger organoids with a compacted morphology which could be passaged every 8–11 days (**Fig. 4-30a, b**). We sorted NAPI2B+/EPCAM+ cells via FACS and re-seeded into the same condition at each passage over 40 days to determine the effect of cell enrichment on the purification and maintenance of the NAPI2B+ cell population (**Fig. 4-30b**). After a single passage, the mean percentage of NAPI2B+/EPCAM+ cells were 31.10% (N=3), and following two further enrichment steps, >93% of cells co-expressed NAPI2B and EPCAM at P3 (N=1) (**Fig. 4-30c**). This showed the purification and continued culture of cells rapidly led to a homogenous NAPI2B+/EPCAM+ cell population, suggesting cell enrichment is still required to maintain the desired phenotype. given we had manipulated major signalling pathways, we wanted to confirm the gene signature of the generated NAPI2B+ epithelial cells.

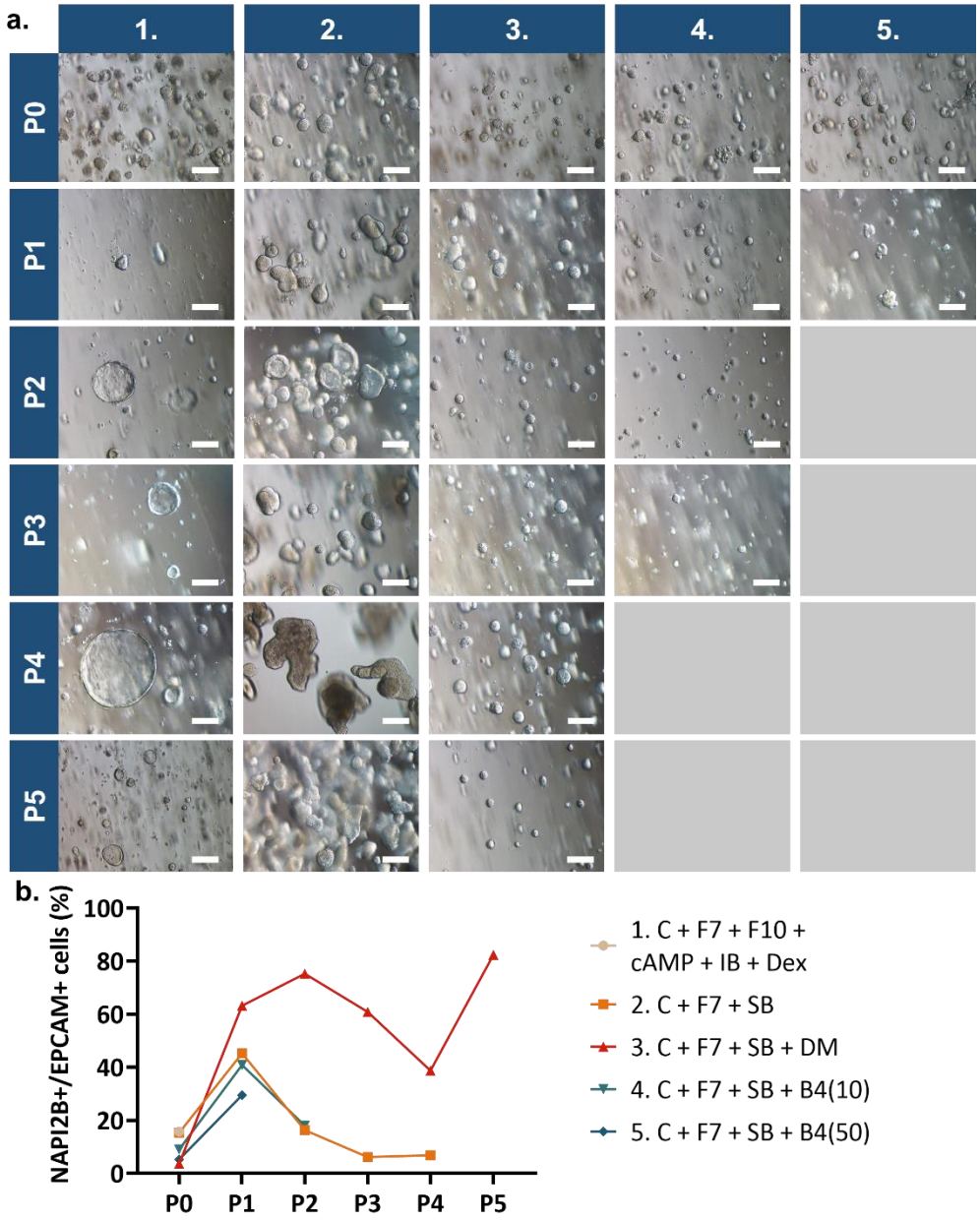


Figure 4-28. Manipulation of TGF-β and BMP signalling leads to variance in differentiation to NAPI2B+/EPCAM+ cells.

(a) Light microscope images of 3D organoids across 5 growth factor conditions from passage 0 to passage 5. Blank squares reflect conditions that were removed. Scale bars, 200 µM. **(b)** Quantification of NAPI2B+/EPCAM+ cells via flow cytometry analysis just prior to passaging of organoids (N=1). C = CHIR99021 (3 µM), F7 = FGF7 (10 ng/ml), F10 = FGF10 (10 ng/ml), cAMP (100 nM), IB = IBMX (100 nM), Dex = Dexamethasone (50 nM), SB = SB431542 (10 µM), DM = Dorsomorphin (2 µM), B4 = BMP4 (10 or 50 ng/ml where specified).

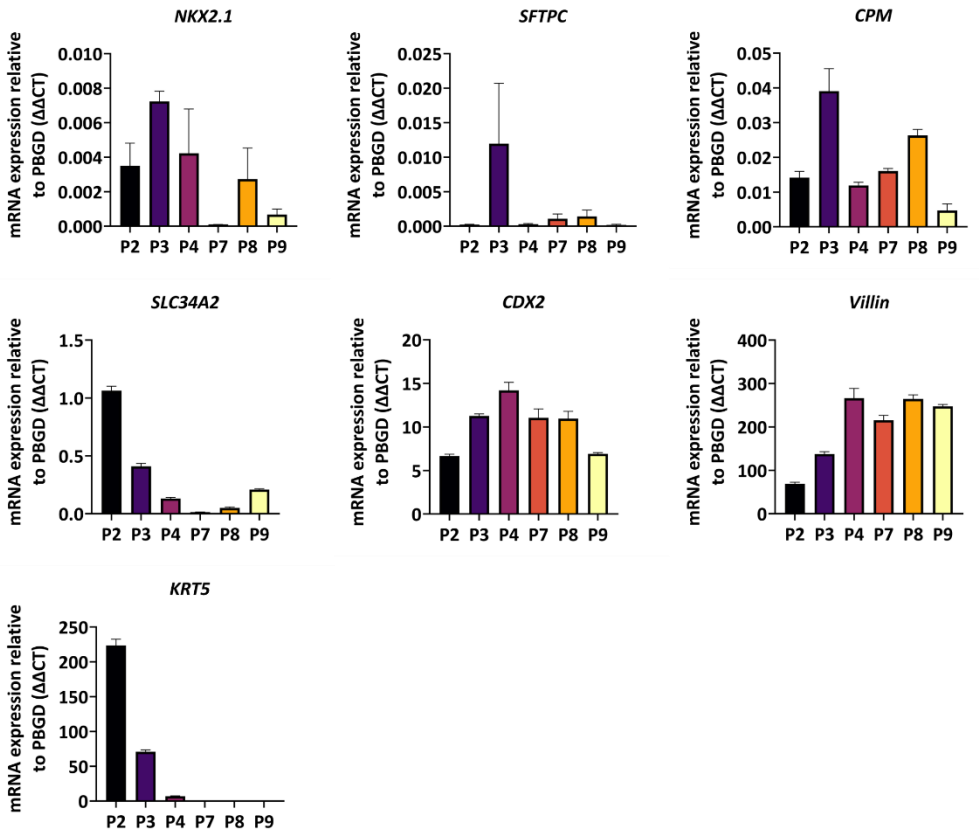


Figure 4-29. Organoids differentiated in the presence of CHIR99021, FGF7, EGF and dual-SMAD inhibition express non-lung endoderm markers without FACS enrichment.

qRT-PCR analysis of type 2 AEC, basal and non-lung endoderm cell markers at sequential passages for organoids maintained with CHIR99021, FGF7, EGF, SB431542 and dorsomorphin without enrichment for NaPi2B/EPCAM. Bars represent mean normalised expression to the housekeeping gene PBGD (ΔΔCt) ± SEM (N=1, 3 technical replicates).

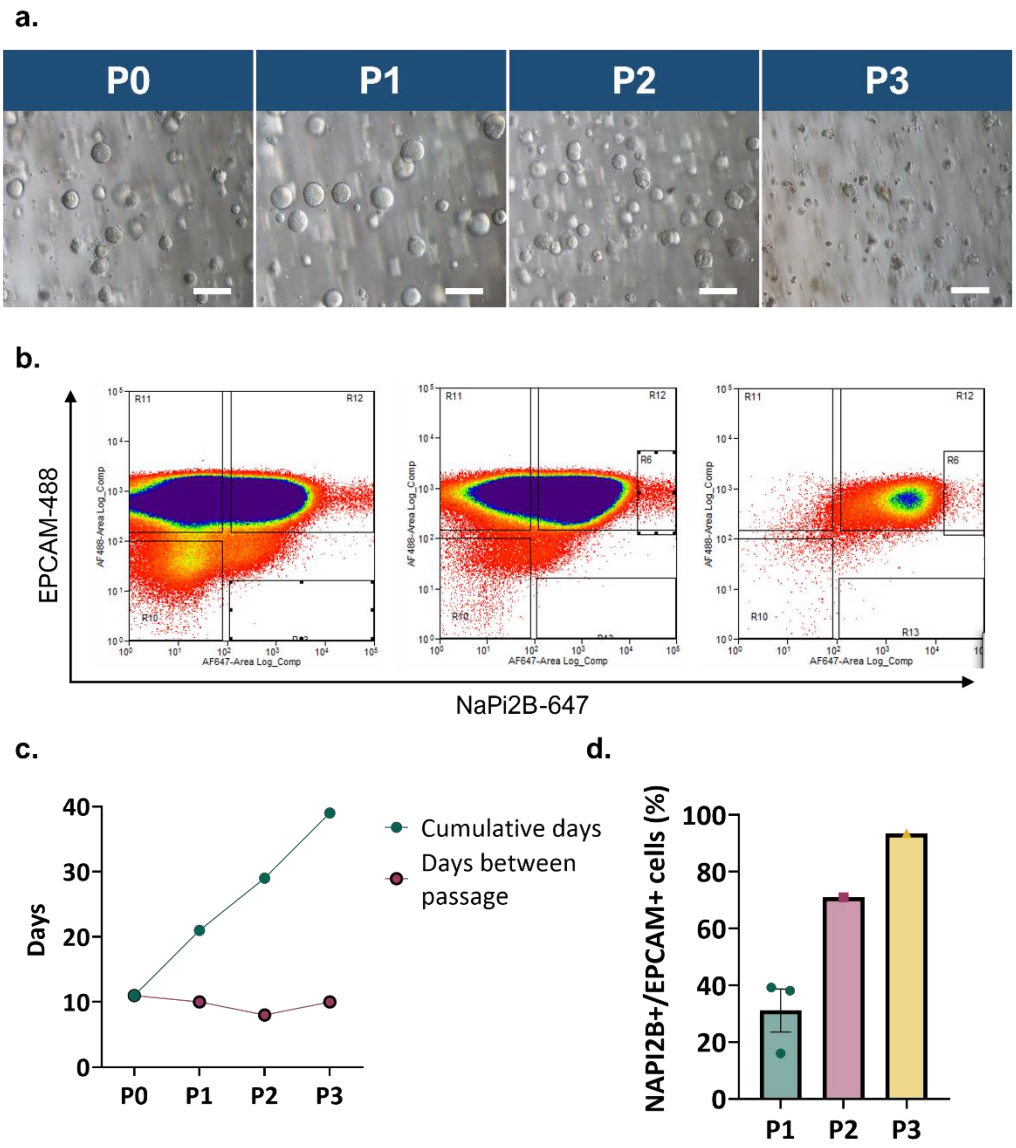


Figure 4-30. Dual-SMAD inhibition with Wnt and FGF signalling induces NAPI2B+ epithelial cells in organoid culture.

(a) Light microscope images of organoids grown in the presence of CHIR99021 and dual-SMAD inhibition in defined medium over 3 passages. Scale bars, 200 µM. **(b)** Representative FACS analysis of BE.31 ($SFTPC^{WT/WT}$) organoids for NaPi2B+/EPCAM+ cells at 3 sequential passages after continued NaPi2B+/EPCAM+ cell enrichment. **(c)** Representative graphic of the timing of the splitting and culture of organoids over 3 passages (N=1). **(d)** Quantification of the percentage of NAPI2B+/EPCAM+ cells at the end of each passage via FACS. Bars represent mean ± SEM where there is more than one biological replicate (P1 N=3, P2 and P3 N=1).

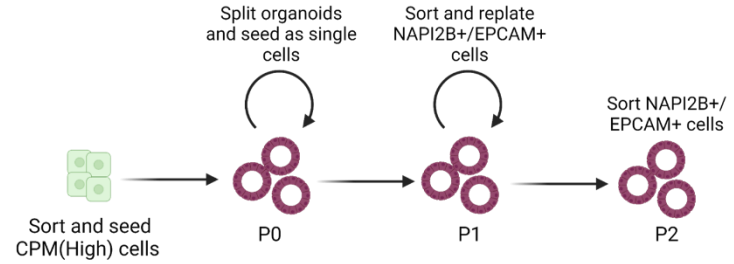
4.2.16. NAPI2B+ epithelial cells grown with WNT and FGF signalling and SMAD inhibition express multiple lung gene markers

Given we had manipulated major signalling pathways, we wanted to determine the gene signature of the NAPI2B+/EPCAM+ cells under this new condition. We performed a more in-depth qRT-PCR analysis on NAPI2B sorted populations at P1 and P2 of the same biological replicate for lung progenitor, airway, alveolar and non-lung endoderm genes (**Fig. 4-31a, b**). Expression of *SLC34A2* was significantly upregulated in the NaPi2B+ population compared to NaPi2B- cells as expected. Interestingly the lung progenitor genes *NKX2.1* and *CPM* and lung bud endoderm marker *FOXP2* were expressed higher in the NaPi2B- population but expression was lost by P2. Several markers related to type 1 AECs (*AGER*, *PDPN*, *AQP5*, *HOPX*), type 2 AECs (*SFTPA*, *SFTPB*) and lung progenitor cells were upregulated in NAPI2B- cells, although statistical significance was not achieved for some genes. There was no expression of *SFTPC* in either condition. In both NAPI2B+ and NAPI2B- populations, the airway basal cell genes *TP63*, *KRT17*, *KRT5*, *PDPN* and *CAV1* were highly upregulated, and expression was enriched at P2 (**Fig. 4-32b**). Markers of other airway cell types, including goblet cells (*MUC5A5*, *MUC5B*), club cells (*SCGB1A1*, *SCGB3A2*, *CCSP*), pulmonary neuroendocrine cells (*CGRP*, *synaptophysin*), ciliated cells (*FOXJ1*) and the airway marker *CFTR* were not significantly upregulated in either NAPI2B population (**Fig. 4-32b**). Non-lung endoderm markers such as *ALB* and *Villin* were upregulated in NAPI2B- at P1 cells but this was not statistically significant (**Fig. 4-32b**). These data suggest under these conditions NaPi2B+/EPCAM+ cells are enriched for genes related to airway basal cells, and purification by FACS and re-culture enhances this phenotype.

To further evaluate these organoids, under the same conditions we performed a comparison of the expression of the same genes between NAPI2B populations at P1 for two independent experiments. As before, *SLC34A2* expression was significantly upregulated in NAPI2B+ cells (**Fig. 4-33a**). The lung bud marker *FOXP2* was expressed significantly in NAPI2B- cells, but we observed no statistically significant differences between NAPI2B populations for other lung progenitor (*NKX2.1*, *CPM*) and type 2 AEC (*SFTPC*, *SFTPA*, *SFTPB*) markers, although these genes were enriched in NAPI2B+ cells of the second biological repeat. The airway basal cell genes *KRT5*, *KRT17* and *CAV1* were also significantly upregulated in NAPI2B+ cells, and other airway basal cell

markers, *TP63* and *PDPN* were upregulated in both NAPI2B populations, but in contrast to lung progenitor and type 2 AEC markers, expression was higher in cells of the first biological replicate. The ciliated cell marker *FOXJ1* was significantly upregulated in NAPI2B- cells, and although not statistically significant, *Villin* and *CFTR* expression were also higher in the NAPI2B- fraction in cells of the second biological replicate. Genes related to other cell types of the airway were not significantly expressed, but the hepatocyte marker *ALB* was significantly upregulated in NAPI2B- cells (**Fig. 4-33a**). Together these findings indicate that treatment with CHIR, FGF7 and EGF, and SMAD inhibition strongly induce the expression of airway basal genes but also other lung-related markers indicating heterogeneous cultures comprised of multiple cell types which varies between independent experiments. This requires further investigation to understand the emerging cell types under these conditions.

a.



b.

Gene	Cell type
NKX2.1	Lung epithelial progenitor
CPM	Lung epithelial progenitor
FOXP2	Lung bud endoderm
SFTPC	Type 2 AECs
ABCA3	Type 2 AECs
SFTPA	Type 2 AECs
LAMP3	Type 2 AECs
SFTPB	Type 2 AECs
SLC34A2	Type 2 AECs
AGER	Type 1 AECs
PDPN	Type 1 AECs / Airway basal
HOPX	Type 1 AECs / Type 2 AECs
KRT5	Airway basal
TP63	Airway basal
KRT17	Airway basal
CAV1	Airway basal / Type 1 AECs
AQP5	Airway basal / Type 1 AECs
FOXJ1	Multiciliate
MUC5AC	Goblet
MUC5B	Goblet
CCSP	Club
SCGB3A2	Club
SYN	Neuroendocrine
CGRP	Neuroendocrine
CFTR	Airway epithelium
ALB	Hepatocyte
Villin	Intestinal / epithelial cells with brush border

Figure 4-31. Airway, alveolar and non-lung endoderm genes and markers.

(a) Schematic showing the procedure for splitting and FACS purifying organoids.

(b) Table of lung and non-lung cell types and their respective markers.

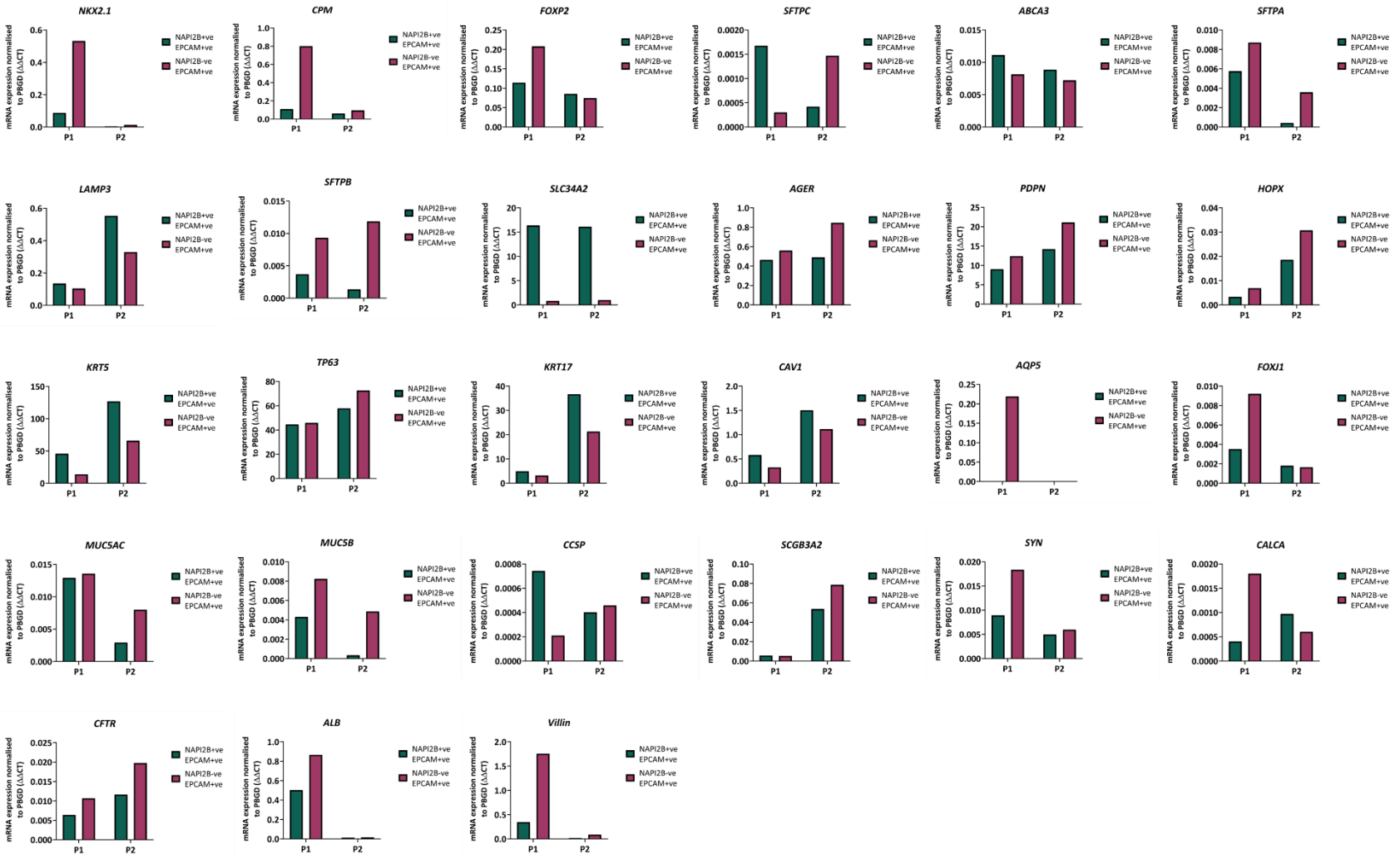

NKX2.1
CPM
FOXP2
SFTPC
ABCA3
SFTPA
LAMP3
SFTPB
SLC34A2
AGER
PDPN
HOPX
KRT5
TP63
KRT17
CAV1
AQP5
FOXJ1
MUC5AC
MUC5B
CCSP
SCGB3A2
SYN
CALCA
CFTR
ALB
Villin
mRNA expression normalised to PBGD (ΔΔCT)
NAPI2B+ve EPCAM+ve
NAPI2B-ve EPCAM+ve
P1
P2

Figure 4-32. Dual-SMAD inhibition specifies organoids with NAPI2B+ cells that express a mix of proximal and distal airway markers.

(a) RT-qPCR analysis of airway, alveolar and non-lung endoderm genes in NAPI2B+/EPCAM+ sorted cells at P1 and after enrichment and further culture at P2. Bars represent normalised expression to the housekeeping gene PBGD ($\Delta\Delta$Ct) ± SEM (N=1).

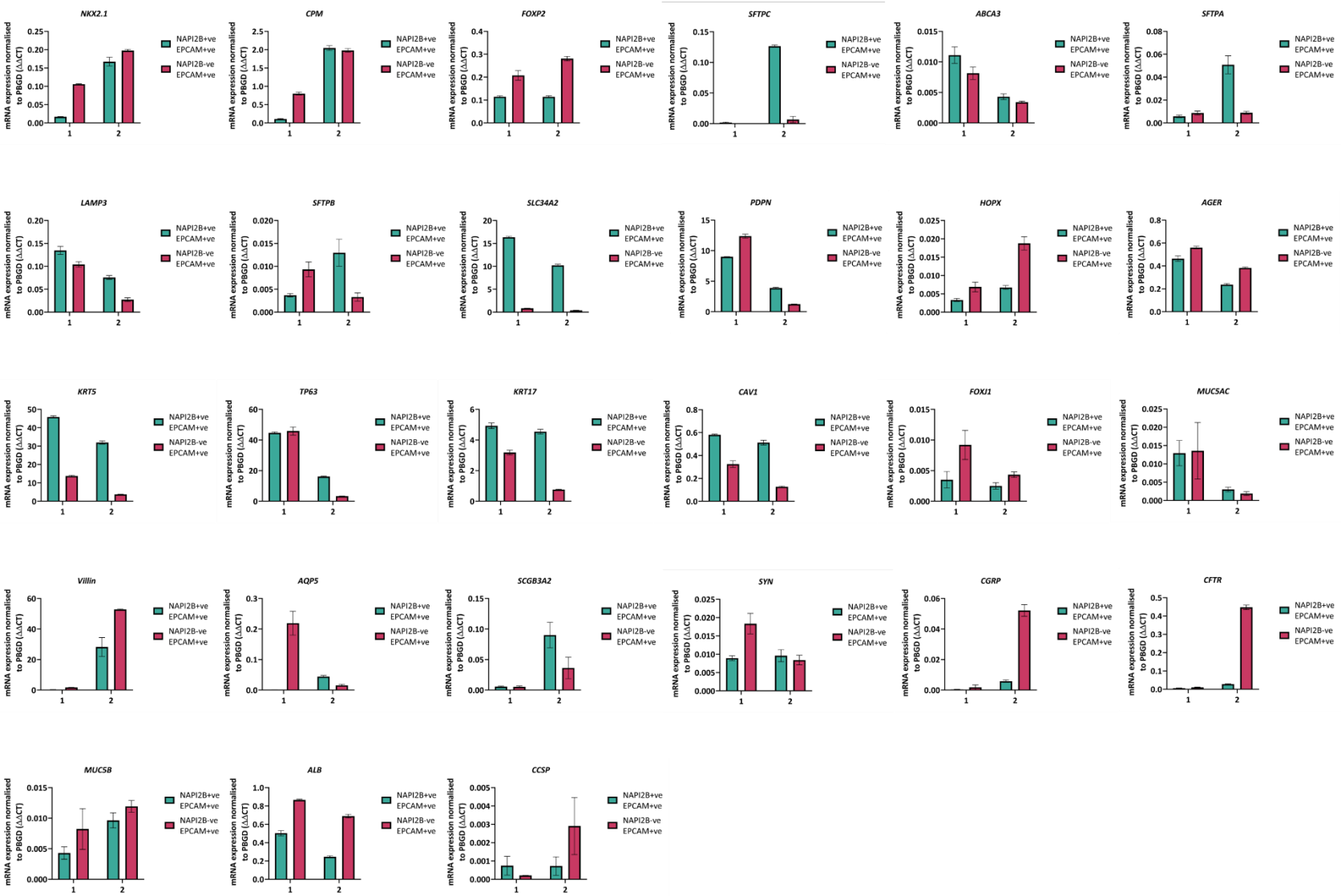

NKX2.1
CPM
FOXP2
SFTPC
ABCA3
SFTPA
LAMP3
SFTPB
SLC34A2
PDPN
HOPX
AGER
KRT5
TP63
KRT17
CAV1
FOXJ1
MUC5AC
Villin
AQP5
SCGB3A2
SYN
CGRP
CFTR
MUC5B
ALB
CCSP
mRNA expression normalised to PBGD (ΔΔCT)
NAPI2B+ve EPCAM+ve
NAPI2B-ve EPCAM+ve

Figure 4-33. NAPI2B+/EPCAM+ cells from organoids maintained with dual-SMAD inhibition express markers of alveolar and airway cell identity.

(a) qRT-PCR of alveolar, airway and non-lung endoderm lineage genes in passage 1 NAPI2B+ and NAPI2B- cell sorted fractions. Bars represent normalised expression to the housekeeping gene PBGD ($\Delta\Delta$Ct) ± SEM (N=1)

4.2.17. Benching the transcriptome of NaPi2B+/EPCAM+ cells to human primary fetal and adult lung

After showing our NaPi2B+ cells were enriched for type 2 AEC genes and were comparable to previously published *in vitro* derived alveolar cells, we compared to RNA-seq data sets of human adult and fetal primary lung samples to determine what developmental stage our hPSC-type 2 AECs reflect. Populations for comparison included NaPi2B+ and NaPi2B- cells isolated from organoids maintained in SFD (cSFDM) and XFL medium, day 35 SFTPC-tdTomato+ and SFTPC-tdTomato- cells, early and late human fetal lung (HFL), 21-week fetal lung and adult lung samples. By PCA of these populations, we showed all NaPi2B-isolated samples clustered together, and early HFL and fetal 21-week samples clustered closely, but all other samples clustered separately, and this was confirmed with hierarchical clustering analysis (**Fig. 4-34a, b**). We performed differential gene expression analysis across all samples and observed the expression pattern for all the primary lung samples were similar but not compared to NaPi2B+ cells (**Fig. 4-35a**). This is to be expected given the primary total lung samples contain a heterogeneous mix of lung cell types. When we compared the DEGs of our selected type 2 AEC markers we showed similar expression patterns between NaPi2B+ cells and early to late HFL (**Fig. 4-35b**). Transcripts related to maturation of type 2 AECs including *SFTPA1*, *SFTPA2* and *ABCA3* were expressed higher in NaPi2B+ cells in comparison to early and late HFL and *SFTPB*, *SLC34A2* and *LPACT1* were expressed across all samples (**Fig. 4-35b**). In contrast, *SFTPC* was expressed higher in late HFL compared to NaPi2B+ cells, and almost all type 2 AEC genes were enriched highest in primary adult lung (**Fig. 4-35b**). These findings suggest our NaPi2B+ are more closely related to late-stage fetal development. Lastly, we performed profiling of our selected gene sets for type 1 AECs and proximal airway cells and observed no significant enrichment of other cell types in NaPi2B+ cells in comparison to other primary lung samples, consistent with enrichment of type 2 AECs in a NaPi2B+ population (**4-36a, b, c**).

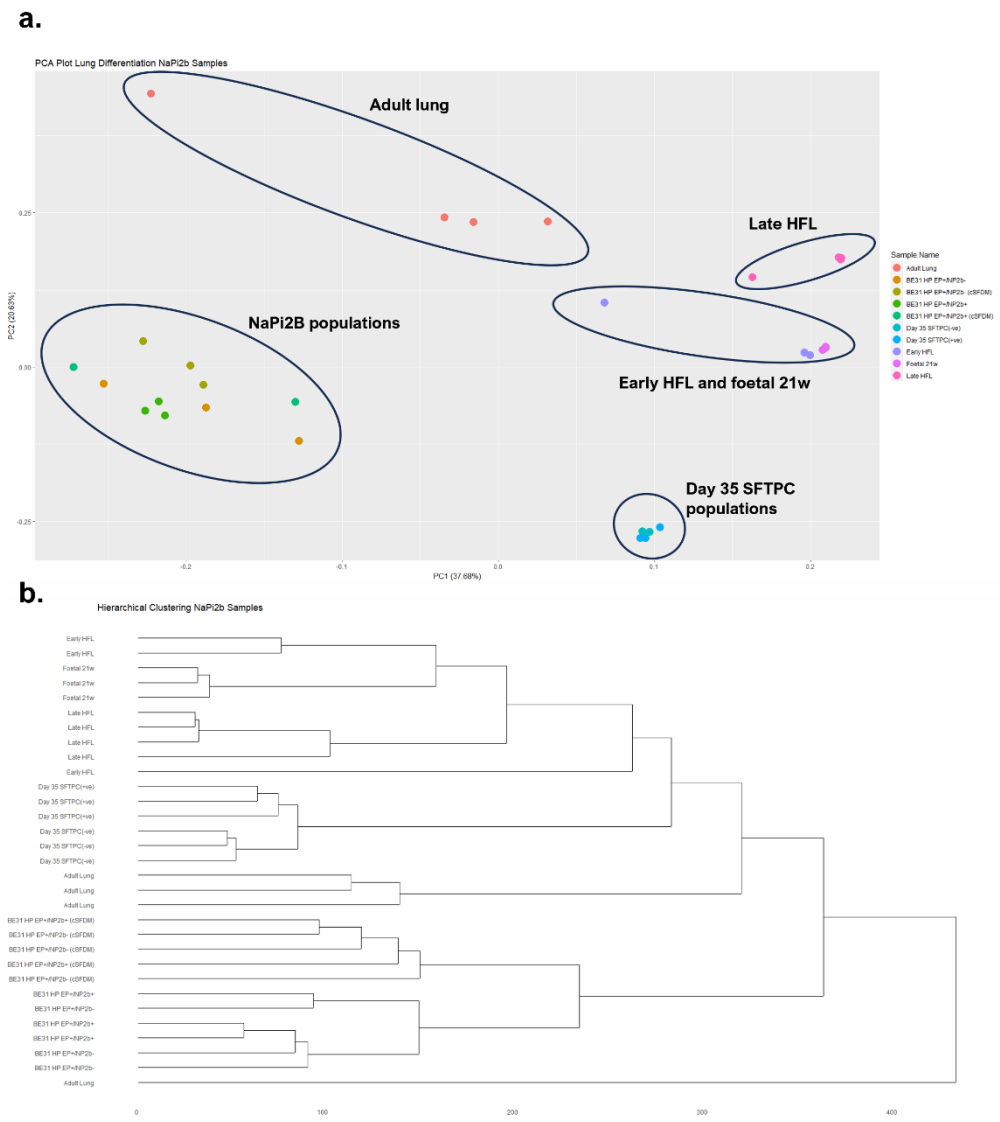


Figure 4-34. Comparison of transcriptomic profiles of NaPi2B cells and primary lung samples.

(a) Principal component analysis of gene expression variance across NaPi2B populations maintained in XFL medium and SFD medium and human fetal and adult primary lung samples. **(b)** Hierarchical clustering of NaPi2B populations and primary lung samples.

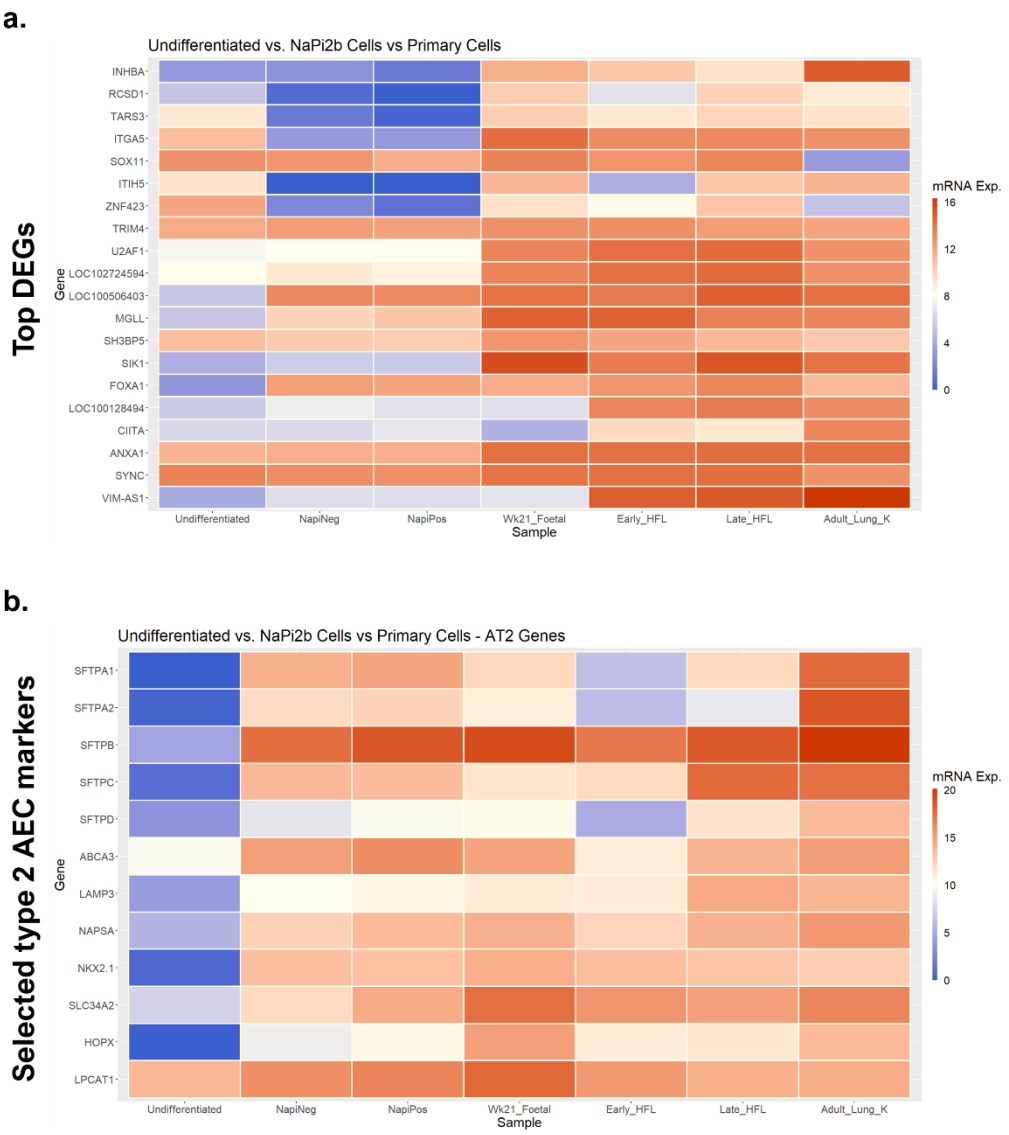


Figure 4-35. Benchmarking the transcriptomic profile of hiPSC-type 2 AECs to human fetal and adult lung samples.

(a) Heatmap of the top DEGs in undifferentiated hiPSCs versus NaPi2B populations versus human fetal and adult primary lung samples. **(b)** Heatmap of the expression of selected type 2 AEC markers in undifferentiated hiPSCs versus NaPi2B populations versus human fetal and adult primary lung samples.

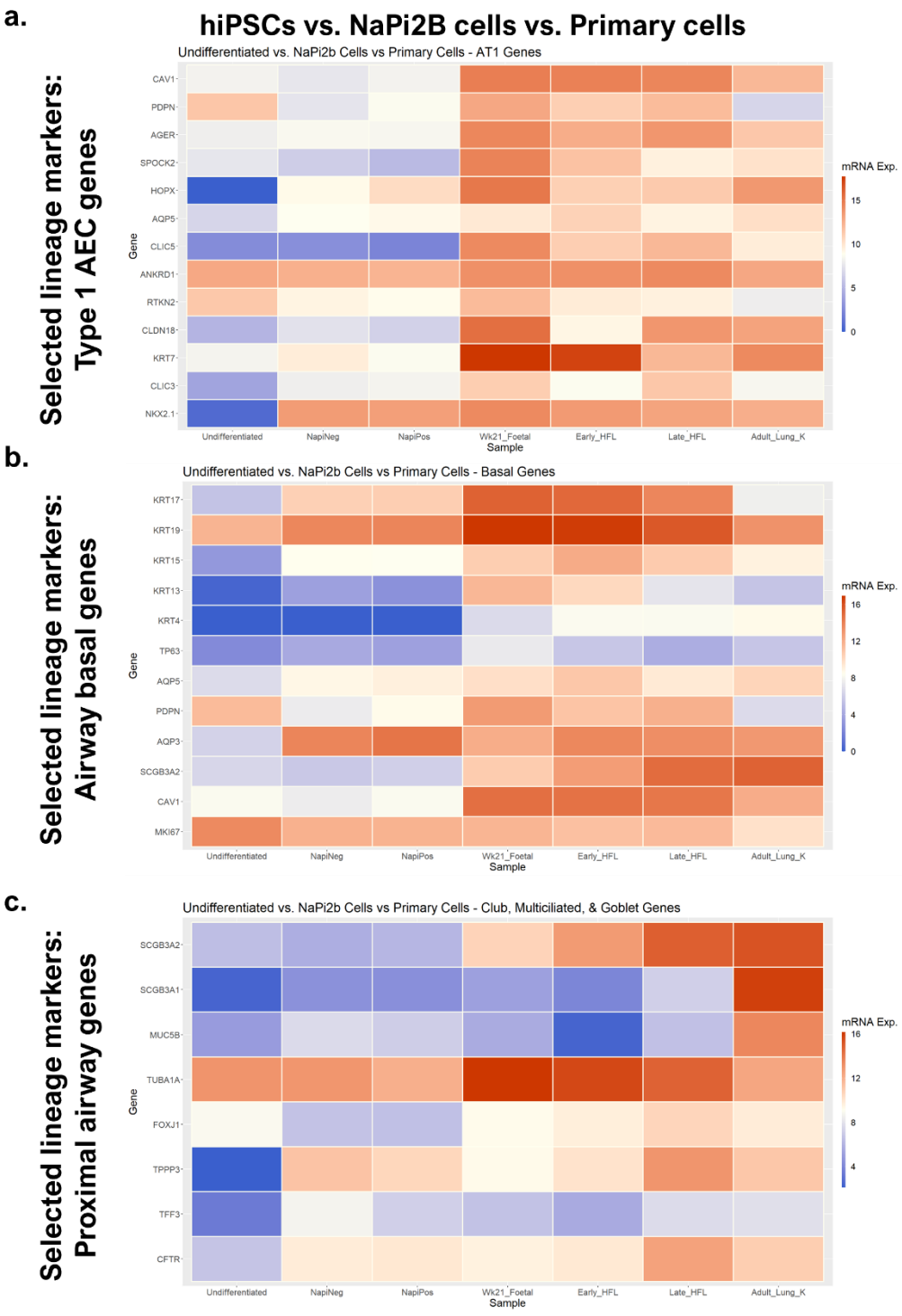


Figure 4-36. Gene expression analysis of non-type 2 AEC lung markers in NaPi2B cells and primary lung samples.

Heatmap of the expression of selected **(a)** type 1 AEC, **(b)** Airway basal and **(c)** goblet, club and multiciliated cell genes in undifferentiated hiPSCs versus NaPi2B populations and fetal and adult primary lung samples.

4.2.18. hiPSC-derived type 2 AEC organoids display mature functional characteristics and can recapitulate aspects of lung disease

To determine whether our derived type 2 AEC organoids display functional characteristics like lamellar bodies, we performed ultrastructural analysis by transmission electron microscopy. We observed the presence of lamellar body inclusions approximately 1-2 µm in diameter with concentric lamellae and a dense core in BE.31 ($SFTPC^{WT/WT}$) organoids (**Fig. 4-37a**). In ABE.27 ($SFTPC^{MUT/MUT}$) organoids, smaller, irregularly shaped inclusions with dense masses and less organised lamellae were observed (**Fig. 4-37b**). To detect secretion of surfactant products, an ELISA was performed to measure secreted SFTPC levels in BE.31 ($SFTPC^{WT/WT}$) and ABE.27 ($SFTPC^{MUT/MUT}$) type 2 AEC organoids after 14 days in culture. We detected significantly higher levels of SFTPC in the supernatant of BE.31 ($SFTPC^{WT/WT}$) organoids compared to ABE.27 ($SFTPC^{MUT/MUT}$) organoids (**Fig. 4-37b**). These data confirm our hiPSC-type 2 AEC organoids display functional lamellar bodies that can synthesise and secrete SFTPC. Measurement and quantification of factors related to ER stress would further validate the potential of this system for disease modelling.

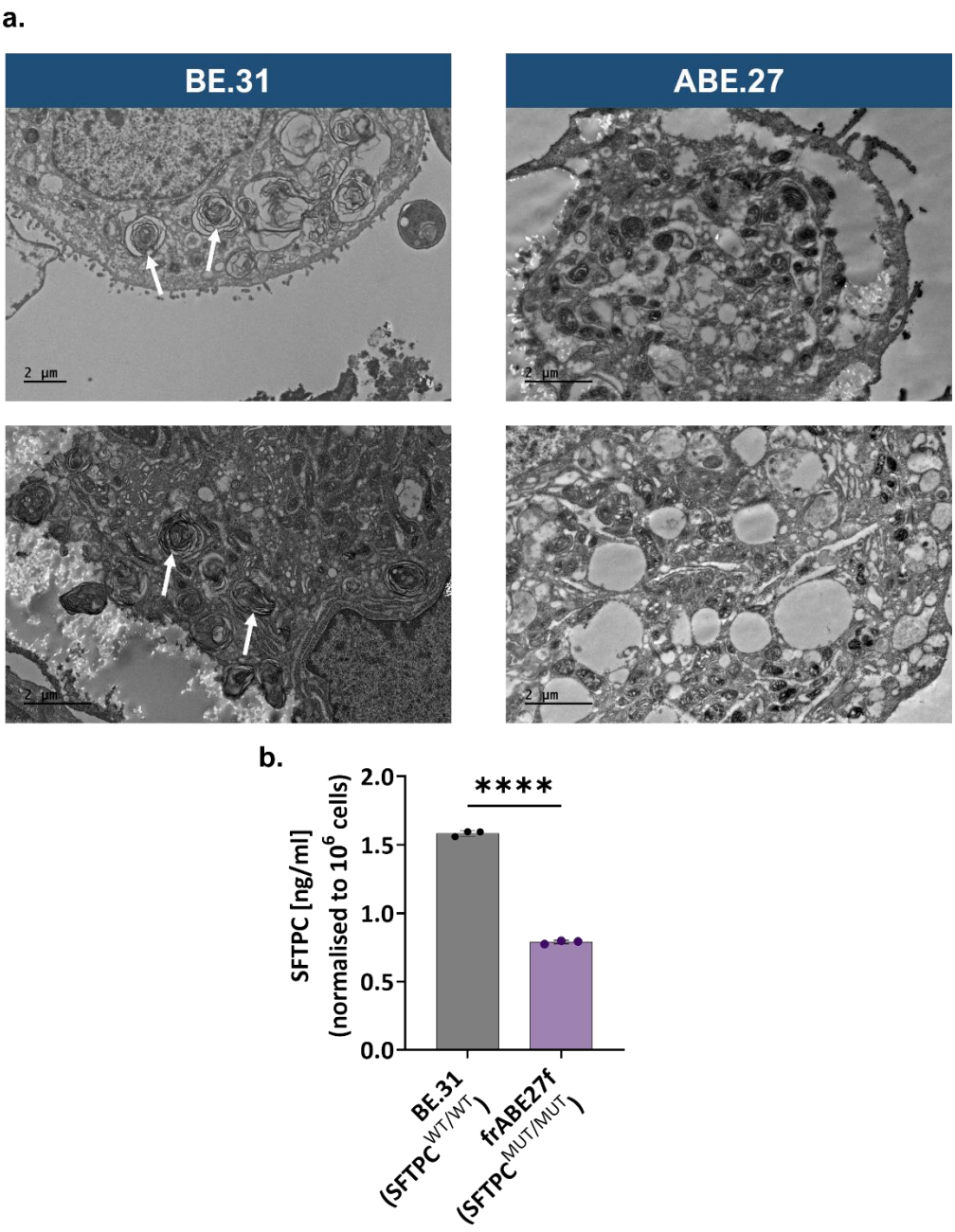


Figure 4-37. hiPSC-derived type 2 AECs are functionally mature and can recapitulate aspects of lung disease.

(a) Representative TEM images of BE.31 ($SFTPC^{WT/WT}$) and ABE.27 ($SFTPC^{MUT/MUT}$) hiPSC-derived type 2 AEC organoids. Arrows indicate lamellar body inclusions. Scale bar, 2 µm. **(b)** Quantification of SFTPC in cell culture medium by ELISA for BE.31 ($SFTPC^{WT/WT}$) and ABE.27 ($SFTPC^{MUT/MUT}$) organoids. Bars represent mean ± SEM (N=3). ****p<0.0001, unpaired two-tailed Student's t test (Figure 4-37 was performed in collaboration with Dr Sara Cuevas Ocaña).

4.3. Discussion

Here, our findings provide an optimised method to efficiently direct the differentiation of hiPSCs to type 2 AECs in fully defined medium with high reproducibility. We show that several combinations of small molecule inhibitors of TGFβ, BMP and WNT signalling are able to induce anterior patterning of DE but do not significantly impact further lung differentiation. By controlling culture density and directly activating WNT/β catenin with recombinant proteins we illustrate enhanced efficiency of lung progenitor cell specification. We performed in-depth transcriptomic profiling at major stages of lung differentiation including in lung progenitor cells and type 2 AECs isolated using cell surface markers we had previously utilised. This confirmed the induction of lung fated cells that emerge in our protocol and validated our platform using defined medium for differentiation. This technology enabled us to benchmark our *in vitro* derived cells at varying differentiation stages to human fetal and adult lung data sets.

In fully defined medium, we observed low efficiency of differentiation to DE cells using a combination of recombinant WNT3a and activin A, but this was significantly improved with PI3K inhibition, as we previously demonstrated. Consistent with our findings and other reports, maximal DE induction required one-day treatment with CHIR99021, transient PI3K inhibition and continued activin A signalling (Naujok et al., 2014; Teo et al., 2014). This was highly reproducible across multiple cell lines and led to a near homogenous population of DE cells which also expressed SOX17 and FOXA2. Previous findings have shown that in the absence of serum or albumin, the activity of CHIR is enhanced and can improve the efficiency of differentiation, but higher doses can also induce toxicity (Lian et al., 2015). We found that 24 h treatment with CHIR at a concentration of 3 μM led to efficient DE differentiation without inducing toxicity, similar to our observations in medium containing albumin. However, given the findings above, it is important to consider the CHIR concentration when differentiating new cell lines. Furthermore, signalling of insulin through PI3K is known to be refractory to DE induction, and differentiation efficiency is enhanced in insulin-free medium, although supplements such as BSA are required for cell survival (Siller et al., 2016). In our medium we used a xeno-free insulin-selenium-transferrin (ITS) supplement that likely accounted for cell survival, but we observed no adverse effect on DE differentiation efficiency, and although we used a PI3K inhibitor, we showed in the

absence of PI3K inhibition DE induction was also highly efficient. This suggests transient CHIR stimulation and activin A treatment enables DE induction in insulin-containing medium, and PI3K can further enhance efficiency with no adverse effects on cell viability.

During development, DE and mesoderm emerge from precursors in the anterior primitive streak and posterior primitive streak, respectively (Lawson et al., 1991). hPSCs have been suggested to transition through a similar primitive streak-like phase during *in vitro* differentiation, identified by expression of genes such as *MIXL1*, *T* and *GSC*. Combinations of WNT, Activin, BMP and FGF signalling are then able to drive endoderm and mesoderm identity (Loh et al., 2014; Mendjan et al., 2014). Previous findings have demonstrated that depending on the duration and concentration of CHIR treatment in combination with bulk cell density, within 24 hours of differentiation, paracrine factors such as NODAL antagonists are released and induce anterior-posterior patterning similar to that in the primitive streak, and prime hPSCs to develop into three alternative fates; DE, presomitic mesoderm and precardiac mesoderm (Kempf et al., 2016). Our findings indicate differentiation in the presence of CHIR, activin A and LY294002 was sufficient to drive DE induction, but we did not analyse early transcripts to identify if cells transitioned through a primitive streak state with capacity to generate mesodermal cells. Transcriptomic profiling did reveal expression of *MIXL1* after 4 days of DE differentiation indicative of a primitive-streak like population, but we also demonstrated the enrichment of putative DE genes including *SOX17*, *FOXA2*, *GATA4*, *GATA6* and *CXCR4* and transcripts related to early DE development like *EOMES* and *TBX3*. *EOMES* has been shown to govern the transcriptional network dictating DE development and cooperates with *TBX3* (Kartikasari et al., 2013; Teo et al., 2011). We also observed expression of *HNF4α* which plays an important role in hepatic development and dictates the specification of liver progenitor cells from hPSC-DE, which could explain the capacity for hepatic lineages to emerge in our culture (DeLaForest et al., 2018; Parviz et al., 2003).

We highlighted the importance of the density of DE cells prior to foregut patterning on the efficiency of lung differentiation in defined medium. Without dissociating and re-plating DE cells, only a very small CPM(hi) population was induced. Like before, we tested single cell and colony passaging of DE cells and found both methods could lead

to specification of CPM(hi) cells, although dissociation to single cells induced more cell death and was harder to reproduce. Endoderm cells that were split as colonies promoted cell survival and could give rise to CPM(hi) cells that express lung progenitor genes like *NKX2.1*, *CPM* and the distal lung marker *SOX9*, and NKX2.1 protein. Cell culture density is known to affect factors such as cell junction proteins and apical-basal polarity. These cues are transduced by transcriptional regulators such as YAP/TAZ to regulate many cellular functions like proliferation and differentiation during development and in PSC culture (Hsiao et al., 2016; Totaro et al., 2018). YAP/TAZ integrates with other signalling pathways such as WNT and can interact with other transcription factors related to lung development like NKX2.1 (Park et al., 2004; Piersma et al., 2015). Cell density can also lead to changes in paracrine signalling which can alter cell fate specification despite the presence of exogenous signalling factors (Kempf et al., 2016). YAP is also known to drive type 1 AEC differentiation with cofactors such as NKX2.1 (Gokey et al., 2021). Although these findings are beyond the scope of this work, understanding whether similar mechanisms occur during earlier timepoints of lung epithelial differentiation could lead to further manipulation of cell density and signalling for greater control of lineage specification.

Our data show that dual-SMAD inhibition alone or in combination with a period of WNT inhibition can pattern DE to an anterior fate and lead to induction of CPM(hi) cells. We found no additional benefit to using more potent small molecule inhibitors and there were no significant differences in differentiation efficiency, although we showed 3 days of dual-SMAD inhibition to be optimal. Other groups have reported high efficiencies generating NKX2.1+ cells from hESCs after dual-SAMD inhibition followed by WNT inhibition, although quantification in these studies was based on immunofluorescence images which are not always specific enough to accurately reflect reported efficiencies. The same findings also reported efficiencies of <40% NKX2.1+ cells for two hiPSC lines indicating variability in differentiation potential of hPSC lines (Huang et al., 2014). We observed higher expression of lung-related transcripts *NKX2.1*, *CPM*, *HOPX*, *AGER* and *SLC34A2* in CPM(hi) cells after dual-SMAD inhibition compared to with combined WNT inhibition. Interestingly, dual-SMAD inhibition with the pharmacological inhibitor LDN193189 led to complete loss of the CPM(hi) population and was only partially recovered in conditions with combined WNT signalling inhibition. LDN193189 is a derivative of dorsomorphin but exhibits higher

potency and specificity for BMP receptors (Yu et al., 2008). LDN193189 has shown to induce AFE broadly expressing PAX9 *in vitro* which is a precursor for several tissues such as the oesophagus, parathyroid glands and thymus (Kearns et al., 2013). Treatment with LDN193189 might bias toward AFE derivatives other than lung, but pharmacological inhibitors of BMP signalling have several reported effects on other intracellular pathways which could explain the differences we observed (Boergermann et al., 2010; Horbelt et al., 2015).

Transient RA signalling for 2 days was sufficient for AFE to gain competence to respond to WNT and BMP ligands for differentiation to lung epithelial progenitor cells, consistent with prior findings (Hein et al., 2022; Rankin et al., 2016). However, continued treatment with RA was optimal for lung induction. We found that FGF signalling was refractory to lung differentiation, consistent with reports that FGF signalling is dispensable, and the essential factors needed for lung specification are WNT, BMP and RA (Rankin et al., 2016; Serra et al., 2017). Combined WNT3a and BMP4 signalling in hPSC-AFE have been shown to favour a lung fate, but addition of FGF2 drives thyroid specification (Serra et al., 2017). Although expression of FGF ligands and receptors has been confirmed across several developmental stages in human fetal lungs, their role in human lung development is yet to be determined (Danopoulos et al., 2019). Treatment of human lung explant cultures with FGF7 and FGF10 for 48 h fails to stimulate epithelial proliferation, but FGF7 has a mitogenic effect on mouse lung explants (Cardoso et al., 1997; Danopoulos et al., 2019). These studies were performed on tissue that would be at a later stage of development than our experiments, but determining whether FGF ligands induce proliferation in hPSC-derived epithelial cells might explain the loss of lung progenitor identity owing to changes in culture density. Nonetheless, FGF signalling was not required in our system for lung differentiation.

An important observation from our study was the enhanced efficiency of CPM(hi) lung progenitor cell specification in the presence of recombinant WNT3a and R-spondin compared to CHIR99021. We also observed significant morphological differences, as treatment with CHIR99021 led the aggregation of cell clusters across the monolayer indicative of over-proliferation, although we did not quantify this. R-spondin proteins function to prevent the degradation of WNT signalling receptors to increase the

availability for binding by WNT3a ligands, which directly stimulates the canonical WNT pathway (ter Steege & Bakker, 2021). On the other hand, CHIR99021 inhibits GSK3 to prevent degradation of β-catenin to indirectly activate WNT signalling, but is known to inhibit more than 20 other kinases, so lacks specificity (Hu et al., 2021). Further to this, the importance of mitochondrial function in determining cell fate is now being realised, and CHIR99021 has shown to impact mitochondrial biogenesis during differentiation to definitive endoderm *in vitro* (Ma et al., 2019; Zhang et al., 2018). Furthermore, other during hPSC differentiation to cardiac cells, CHIR99021 and culture density were shown to significantly alter cell cycle regulation and impact differentiation (Laco et al., 2018). Despite similar findings not yet being reported during lung differentiation, it highlights an important consideration for use of small molecule inhibitors and perhaps favours direct signalling pathway activation to reduce off-target effects, leading to greater reproducibility.

Using this protocol, we showed multiple hiPSC lines could generate CPM(hi) cells, but one hiPSC line displayed resistance to differentiation. It is well-understood that hiPSC lines can exhibit extensive variability in DNA methylation and gene expression as a result of donor or cell-type variation, and can significantly influence propensity for differentiation (Liu & Zheng, 2019). In differentiation-resistant hiPSC lines, screening growth factor concentrations might be necessary to determine optimal conditions for cell specification, as has been reported for lung progenitor cells (Gotoh et al., 2014). Unlike other protocols that perform lung differentiation for 9-11 days, we observed the onset of CPM(hi) cells only 2 days following AFE patterning, and by day 4 there was a robust induction of a distinct CPM(hi) lung progenitor cell population. This shows in defined medium lung progenitor cells emerge at early timepoints and extended specification periods are not necessary. Dependent on the downstream application, cells can be grown until day 6 and still maintain a high percentage of CPM(hi) cells if isolation of more cells is required. We adopted a 6-day differentiation to ensure we isolated enough cells for further differentiation.

Interestingly, when we performed whole transcriptome analysis, unsorted cells on sequential days of lung differentiation and CPM populations clustered together, but separately from hiPSCs, DE cells and NaPi2B-sorted populations. All of these cells are derived from the same DE population so may be transcriptionally similar, but at

progressive stages of acquiring a lung epithelial progenitor fate. CPM(hi) cells were enriched for *NKX2.1* which is regarded as a master transcription factor for lung development. *NKX2.1* binds to and regulates the expression of several type 2 AEC genes including *SFTPA1*, *SFTPA2*, *SFTPB* and *LAMP3* (Stuart et al., 2021). During development, prospective thyroid cells emerge from the same NKX2.1+ region as lung derivatives, and can arise during *in vitro* differentiation to lung, but we did not observe expression of the thyroid marker PAX8 in any sample analysed confirming the lung identity of CPM(hi) cells (Longmire et al., 2012). Similar to our qRT-PCR data, we also detected the expression of *TTR* across all lung differentiation samples and in CPM(hi) populations, but further investigation is required to determine the significance of this finding. We also observed significant expression of transcripts related to WNT signalling including *FZD1* and *RSPO3* in CPM(hi) cells, supporting the understanding that WNT signalling is key for lung specification. Similarly, *SHH* was significantly expression. During development, SHH signals from the ventral foregut epithelium to the adjacent mesenchyme to upregulate the expression of WNT and BMP to promote the onset of respiratory-related genes during development (Rankin et al., 2016). Improvements in differentiation efficiency of hPSCs to NKX2.1+ cells have been reported with recombinant SHH and small molecule enhancers of SHH signalling, albeit under different conditions to our protocol (Dye et al., 2015; Wong et al., 2012) However, activation of SHH signalling in our differentiation might promote lung differentiation efficiency.

We showed CPM(hi) lung progenitors were able to give rise to 3D organoids and differentiate into cells expressing SFTPC in defined medium. Whole transcriptome analysis revealed the enrichment of several type 2 AEC transcripts in NaPi2B+/EPCAM+ isolated cells, including genes related to maturation. We found NaPi2B+ and NaPi2B- populations clustered together and their gene expression profiles were highly comparable, with slightly higher expression of type 2 AEC markers in the NaPi2B+ fraction. This might reflect a continuum of cell maturation whereby NaPi2B- cells have not fully differentiated to type 2 AECs expressing NaPi2B. Similar findings have been reported in hPSC-derived SFTPC+ and SFTPC- cells isolated using a tdTomato fluorescence reporter (Jacob et al., 2017). The expression of transcripts of other lung cell types was not observed. To validate our strategy, we performed transcriptomic comparison of NaPi2B+ cells to SFTPC-tdTomato+ cells, and found similar gene

expression profiles, although some minor differences were observed. Our NaPi2B+ cells had less *SFTPC* expression, but higher expression of *ABCA3*. The *ABCA3* gene encodes for a transporter that mediates movement of phospholipids into lamellar bodies and is regarded as a mature marker of type 2 AECs (Beers & Mulugeta, 2017). During *in vitro* lung differentiation, compared to *SFTPC* fluorescence reporters, *ABCA3* reporters can enrich for a more mature type 2 AEC (Sun et al., 2021). This same study confirmed NaPi2B could enrich for *ABCA3-GFP*+ type 2 AECs. These findings suggest our sorting approach isolates type 2 AECs with a more mature phenotype. When we profiled our selected gene lists for type 2 AEC proliferation and transitional markers, and Wnt responsive genes, we found enrichment of these transcripts in SFTPC+ cells, indicating a more proliferative population than NaPi2B+, further supporting our hypothesis. Comparison to publicly available datasets of human fetal and adult lung showed our NaPi2B+ cells had similar gene expression profiles to early and late human fetal lung. Transcripts related to type 2 AEC maturation were enriched in adult lung, as expected, suggesting our type 2 AECs were mor fetal-like. However, our TEM analysis revealed the presence of functional lamellar body inclusions in BE.31 ($SFTPC^{WT/WT}$) organoids that could secrete SFTPC as identified by an ELISA. These findings demonstrate our type 2 AECs display functionally mature characteristics. In ABE.27 ($SFTPC^{MUT/MUT}$) organoids, the presence of small, poorly formed lamellar body inclusions and reduced SFTPC secretion confirmed that our platform could recapitulate features of human disease.

To date, only condition has been described for specification of hPSC-type 2 AECs (Jacob et al., 2017; Yamamoto et al., 2017). This combination can maintain type 2 AECs as a self-renewing population but repeated enrichment via SFTPC fluorescence reporters is required to maintain type 2 AEC phenotype and prevent off-target non-lung cell types emerging (Jacob et al., 2017; Sun et al., 2021; Yamamoto et al., 2017). To determine if we could induce further maturation and maintain the type 2 AEC phenotype, we subjected our isolated CPM(hi) cells to factors including WNT, FGF and EGF proteins and TGFβ and BMP inhibitors, which is more similar to conditions used to maintain human adult type 2 AECs in culture. We showed the rapid emergence of NaPi2B+/EPCAM+ cells in organoids (Katsura et al., 2020; Youk et al., 2020). After only 2 enrichments a near homogenous NaPi2B+/EPCAM+ population was specified. Interestingly, we observed a strong airway basal cell gene signature in isolated

NaPi2B+/EPCAM+ cells which was enhanced with enrichment and further culture under these conditions. When we repeated this condition, lung progenitor genes like *NKX2.1* and *CPM* were expressed in both NaPi2B-positive and -negative populations, and type 1 and 2 AEC genes were expressed in the NaPi2B+ fraction along with markers of basal cells, suggesting a heterogeneous population.

Both human adult- and hPSC-derived airway basal cells can be maintained *in vitro* as a self-renewing epithelial population with dual-SMAD signalling inhibition (Hawkins et al., 2021; Mou et al., 2016). It has also been shown that isolated human fetal lung bud tips give rise to organoids with alveolar characteristics such as SFTPC expression in the presence of CHIR99021 and dual-SMAD inhibition, smaller proliferative organoids expressing the basal marker TP63 with FGF7 and dual-SMAD inhibition, and a mix of both organoid types when CHIR99021, FGF7 and dual-SMAD inhibition is combined, which is similar to our gene expression analysis (Lim et al., 2023). Recent findings have identified a role for BMP signalling and TGFβ inhibition in promoting the expression of mature type 2 AEC markers in lung explants, and in combination with CHIR99021 and FGF7 signalling, can strongly promote the differentiation of fetal bud tip progenitor organoids to a mature type 2 AEC fate (Frum et al., 2023). In our system, this same condition could support the formation of small, rounded organoids from hPSC-CPM(hi) cells, but they did not readily expand and were not maintained beyond 3 passages. Although we observed the onset of NaPi2B+/EPCAM+ cells, we did not perform any characterisation to determine the identity of this population. Human type 2 AECs also exhibit plasticity, and can transdifferentiate to KRT5+ basal cells in culture, which is promoted by TGF-β signalling and BMP inhibition (Kathiriya et al., 2022). Further analysis by immunofluorescence staining and scRNA-seq is needed to confirm the identity of the cell populations present in our organoids cultured with WNT, EGF and FGF signalling, and BMP and TGFβ inhibition. However, our gene expression analysis indicates a heterogeneous mix of airway and alveolar cells. Our understanding of conditions required to facilitate alveolar cell development, maintenance and maturation is still incomplete. Furthermore, these findings highlight an important consideration when using cell surface proteins to enrich for target cells, as it appears NaPi2B also enriches for basal-like cells, which has not previously been reported.

5. Differentiation of hiPSCs to type 2 AECs in fully defined, xeno-free condition

5.1. Introduction

Induced pluripotent stem cell technology offers the potential to generate an unlimited source of therapeutically-relevant cell types for replacement therapy while circumventing problems surrounding immune rejection (Yamanaka, 2020). Significant advances have been made toward translating cell therapies to the clinic and several clinical trials have already been conducted utilising hPSC-derived cell types to treat diseases such as type 1 diabetes mellitus, Parkinson's disease and age-related macular degeneration (de Klerk & Hebrok, 2021; Mandai et al., 2017; Morizane, 2023). The development of stem-cell products for clinical application must adhere strictly to GMP guidelines for advanced therapy medicinal products. One major consideration is the exposure of the cell product to animal-derived materials during the manufacturing process. The presence of xenogeneic materials increases the risk of pathogens such as zoonotic viruses being transferred to patients or eliciting immune activation in response to animal-derived proteins or glycans in derived cells (Günzburg & Salmons, 2000; Jang et al., 2013). For this reason, development of fully defined, xeno-free differentiation protocols are necessary to facilitate clinical translation of hPSC-derived cells.

Laminins are large, trimeric glycoproteins that constitute a major component of the basement membrane which bind to integrins to anchor cells to the ECM and provide structural support for tissues (Colognato & Yurchenco, 2000). Laminins are composed of one α, one β and one γ polypeptide chain liked by disulphide bonds. The human genome encodes for five α, three β and three γ chains and multiple laminin isoforms have been identified according to the composition of polypeptides. For example, the subunit composition of laminin-521 is α5β2γ1, laminin-511 is α5β1γ1 and laminin-121 is α1β2γ1 (Hamill et al., 2009). Laminin-511 and -521 contain domains that can bind with high affinity to α3β1 and α6β1 integrins (Nishiuchi et al., 2006). During early embryonic development, β1, γ1, α1, and α5 laminin chains are expressed, and laminin-511 is found abundantly in the basement membrane and inner cell mass of the developing blastocyst (Klaffky et al., 2001; Miner et al., 2004). hESCs express β1, β2,

γ1, α1, and α5 laminin chains and laminin-511 and -521 can support the survival and self-renewal of hPSCs long-term in culture (Rodin et al., 2014; Rodin et al., 2010). Detachment of PSCs from the ECM causes heightened actin-myosin contraction which is regulated by phosphorylation of myosin light chain via the Rho associated kinase (ROCK), eventually resulting in apoptosis (Kaneko-Kawano et al., 2012). Therefore, inhibitors of ROCK such as Y-27632 are used in cell culture to prevent actin-myosin-mediated cell death during single-cell passaging of hPSCs (G. Chen et al., 2010). It has been proposed that laminin-521 can support survival and self-renewal of hPSCs independent of ROCK inhibitors through binding of α6 and β1 integrins, increased motility dependent on β1 integrins to rapidly form cell-cell connections and activation of the PI3K/AKT pathway (Rodin et al., 2014). Laminins are therefore highly suitable substrates for maintenance of hPSCs *in vitro*.

Laminins are also widely expressed in the developing lung, including all five α chains. The α5 chain is expressed during all stages of development of the lung and into adulthood (Pierce et al., 2000). In the mouse lung, the laminin α5 chain is found in the basement membrane of the alveoli and is essential for alveolar cell differentiation (Nguyen et al., 2005). Laminin-111 is restricted to early fetal lung but is critical for branching morphogenesis (Nguyen & Senior, 2006). Recombinant laminins could therefore provide a more defined, physiologically relevant substrate for *in vitro* specification of alveolar cells. Differentiation of hPSCs to cell types including that of the pancreas and liver have been previously reported (Yap et al., 2019).

Chemically defined, xeno-free differentiation protocols to DE have been established. However, they rely on 3D suspension culture which can be technically challenging and more difficult to reproduce (Diekmann et al., 2019). Currently, published methods to differentiate hPSCs to type 2 AECs rely on several animal-derived products. This includes Matrigel™, feeder cell layers, bovine serum albumin (BSA) and defined supplements such as B-27™ that also contain BSA and other supplements of animal origin (Jacob et al., 2017; Yamamoto et al., 2017). Matrigel™ matrix is an undefined, complex mix of ECM proteins and signalling molecules from the secretions of sarcoma cells of Engelbreth-Holm-Swarm mice and is widely used in stem cell and organoid culture (Hughes et al., 2010). These undefined factors introduce significant physiological variability and make it increasingly challenging to understand the

signalling events both during differentiation and in organoid maintenance and function. Variability of Matrigel™ batches further hinders the reproducibility of cell culture experiments (Aisenbrey & Murphy, 2020; Kozlowski et al., 2021). Therefore, beyond enabling clinical translation of hPSC-derived cells, defined conditions free of xenogeneic contaminants are desired to reduce biological variability, as discussed in chapter 4, and would have broader ethical implications regarding the reduction of animals used for research (refer to the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs)).

To the best of our understanding, there are currently no protocols for chemically defined, xeno-free differentiation of hPSCs to type 2 AECs. Therefore, our aim in the following work was to utilise our optimised differentiation protocol in chemically defined medium to generate type 2 AECs under xeno-free conditions by using recombinant human laminin as a basement membrane substrate. The ability to produce therapeutically relevant alveolar cells for future use in replacement cell therapy would be a major advancement in regenerative medicine.

5.2. Results

5.2.1. hiPSCs can be propagated on recombinant laminin

To develop a completely xeno-free type 2 AEC differentiation protocol, we first sought to determine whether hiPSCs which are routinely maintained on animal derived Matrigel™ could be propagated on human recombinant laminin in HB E8 medium. In the following chapter all experiments were performed using BE.31 ($SFTPC^{WT/WT}$) hiPSCs unless stated otherwise. We demonstrated that BE.31 ($SFTPC^{WT/WT}$) hiPSCs could be thawed onto laminin-521, be passaged as single cells and were able to form compacted, pluripotent cell colonies with well-defined edges after 72-96 h of culture (**Fig. 5-1a**). hiPSCs maintained on laminin-521 co-expressed the markers NANOG and OCT4 (**Fig. 5-1b**) and NANOG and SOX2 (**Fig. 5-1c**), demonstrating their maintenance of pluripotency. Given the suitability of laminin-521 for hiPSC maintenance, we wanted to determine whether this substrate could support differentiation.

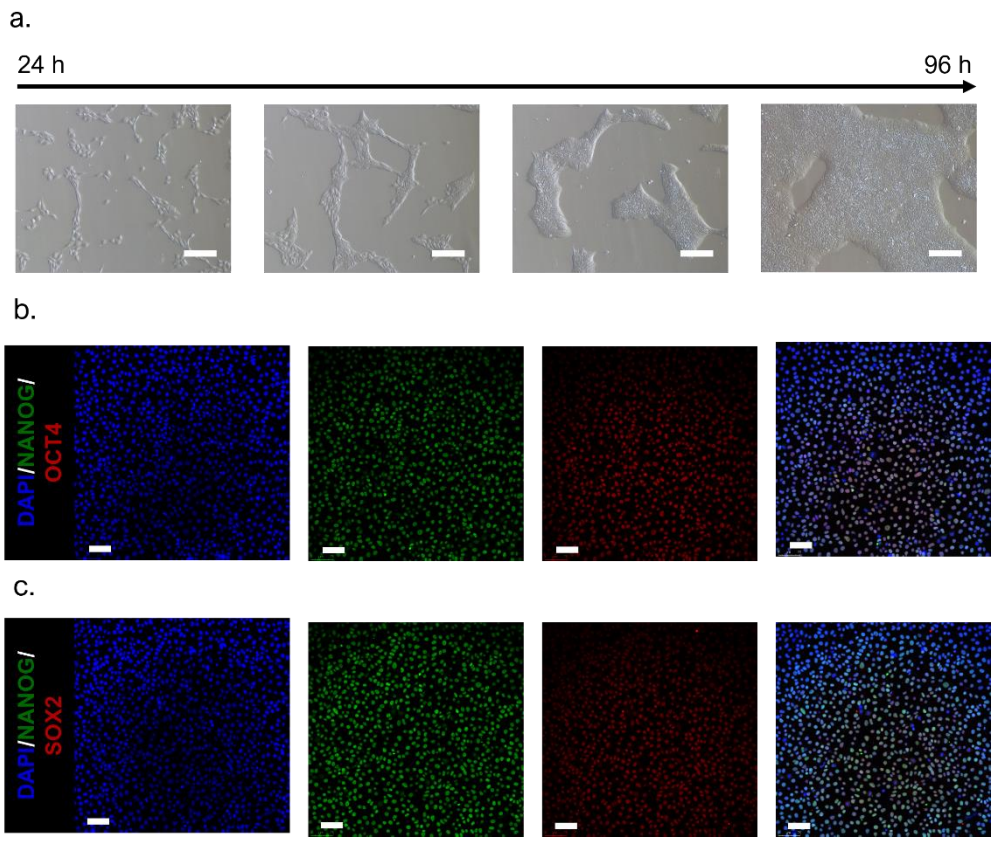


Figure 5-1. hiPSCs maintained on recombinant laminin express pluripotency markers.

(a) Representative light microscope images of BE.31 ($SFTPC^{WT/WT}$) hiPSCs maintained in E8 medium on recombinant laminin over 4 days. Scale bars, 200 µm. **(b)** Immunofluorescence staining of NANOG and OCT4 and **(c)** NANOG and SOX2 in hiPSCs maintained on recombinant laminin in defined E8 medium (Nuclei stained with DAPI, blue). Scale bars, 75 µm.

5.2.2. Highly reproducible differentiation of hiPSCs to DE in xeno-fee conditions

To determine whether hiPSCs could be specified to DE on laminin-521, we first tested a series of hiPSC seeding densities and performed our optimised DE differentiation for 4 days (**Fig. 5-2**). At each seeding density BE.31 ($SFTPC^{WT/WT}$) hiPSCs readily migrated out from colonies and formed a complete monolayer 2 days after differentiation induction (**Fig. 5-3a**). On day 4, there was a near-pure population of CXCR4+/CKIT+ DE cells across all seeding densities tested as assessed by flow cytometry and so we selected 20,000 cells/cm^2 as the density for future experiments (**Fig. 5-3b**). Our optimised differentiation on laminin led to >98% CXCR4+/CKIT+ cells for BE.31 ($SFTPC^{WT/WT}$) hiPSCs (N=8) and >99% CXCR4+/CKIT+ cells for ABE.27 ($SFTPC^{MUT/MUT}$) (N=3) and C3g3.14 (N=3) (**Fig. 5-4a, b**).

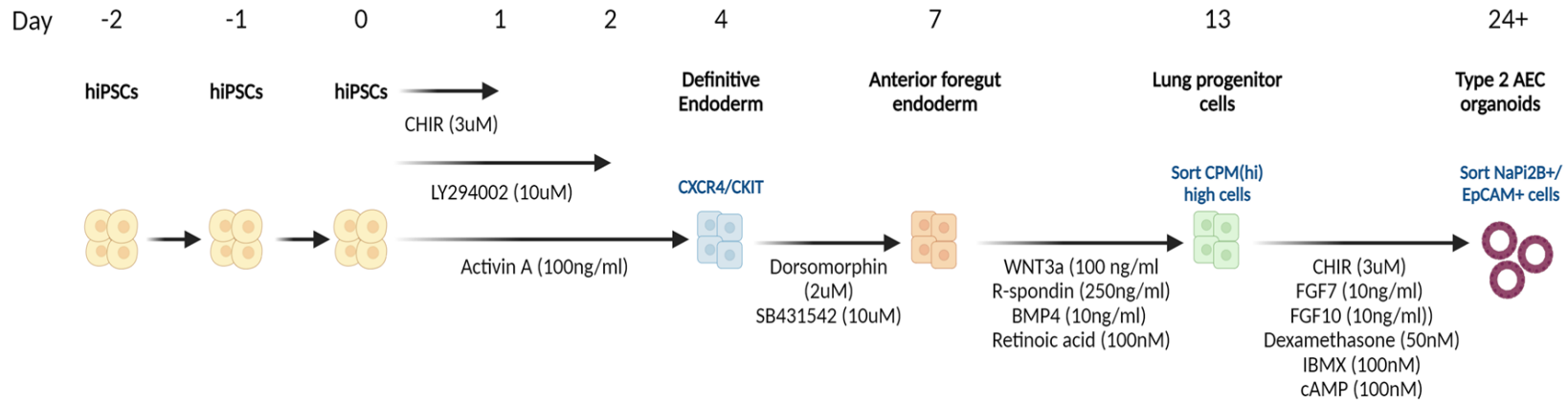


Figure 5-2. Schematic of the fully optimised differentiation of hiPSCs to type 2 AECs in defined medium.

The optimised protocol consists of a 4 day DE treatment with CHIR99021 (D0-D1), LY294002 (D0-D2) and activin A (D0-D4). Following endoderm differentiation, cells are treated with SB431542 and dorsomorphin to inhibit TGFβ and BMP signalling, respectively, and promote patterning to AFE. Cells are then differentiated toward lung epithelial progenitor cells through activation of WNT signalling via the recombinant proteins WNT3a and R-spondin in combination with BMP and RA signalling. Once CPM(hi) cells are isolated through cell sorting they are differentiated towards a type 2 AEC fate in the presence of CHIR99021, FGF7, FGF10 and factors to promote maturation including dexamethasone, IBMX and cAMP as described previously. Figure made using Biorender.

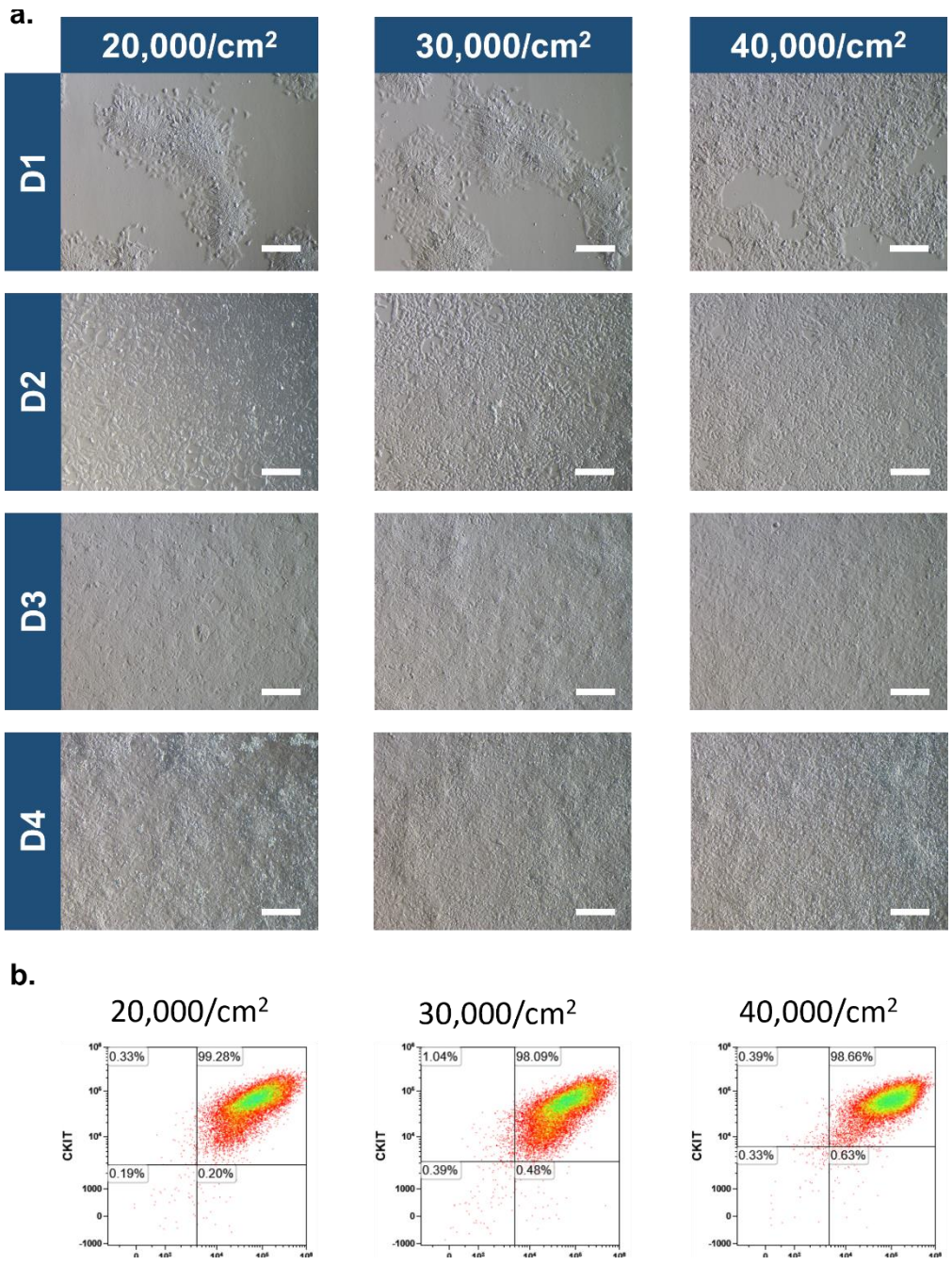


Figure 5-3. Different hiPSC seeding densities lead to robust induction of CXCR4+/CKIT+ cells in XFL medium on recombinant laminin.

(a) Representative light microscope images of cells across 4 days of DE differentiation on recombinant laminin in animal-free medium. Scale bars, 200 µm. **(b)** Flow cytometry analysis of endoderm cells on day 4 of differentiation for each seeding density. (N=1).

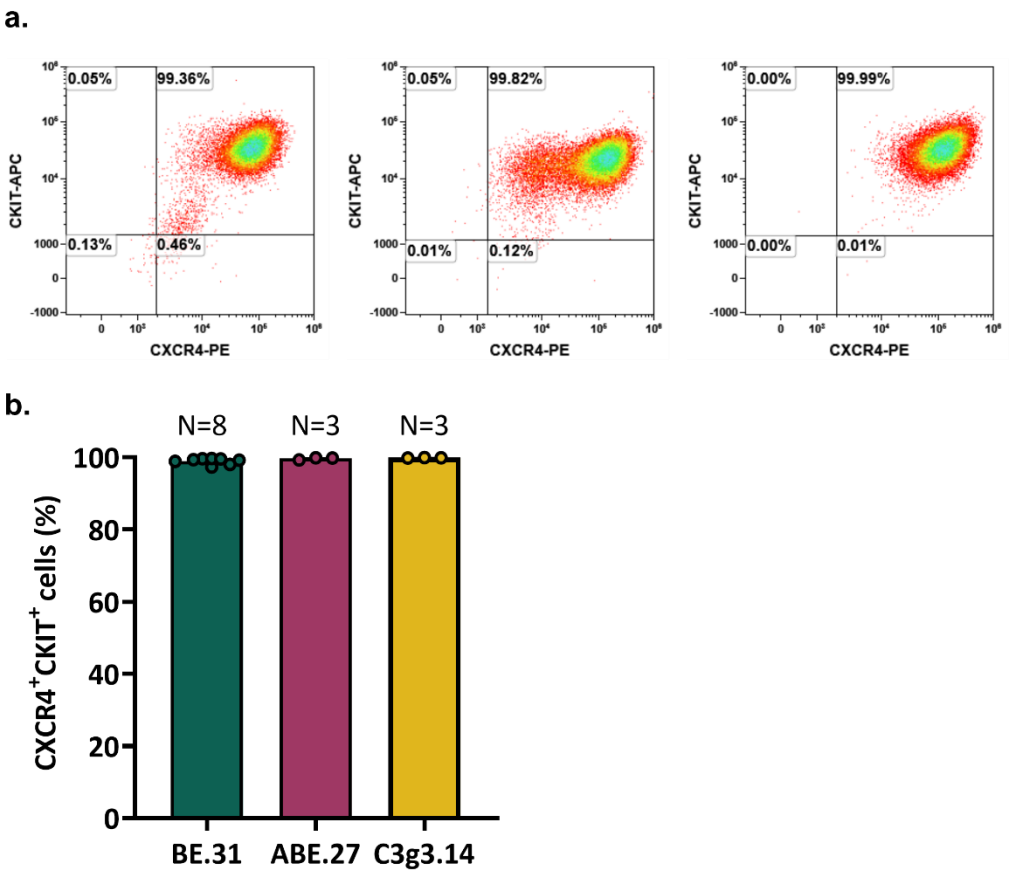


Figure 5-4. Differentiation of hiPSCs to DE is reproducible in XFL medium on recombinant laminin.

(a) Representative flow cytometry analysis of BE.31 ($SFTPC^{WT/WT}$), ABE.27 ($SFTPC^{MUT/MUT}$) and C3g3.14 D4 DE cells for CXCR4 and CKIT. **(b)** Quantification of CXCR4+/CKIT+ cells after 4 days of DE differentiation. Bars represent mean ± SEM.

5.2.3. Transcriptomic profiling of hiPSC-derived DE in xeno-free conditions

We next sought to characterise DE cells derived on laminin-521 by performing transcriptomic analysis using RNA-seq. For comparison, we included DE cells derived using our optimised protocol on Matrigel™. To identify genes differentially expressed after DE differentiation, we compared day DE with undifferentiated hiPSCs. We observed that the most differentially expressed genes in day 4 DE were enriched for DE genes such as *SOX17*, *GATA6*, *HHEX*, *CER1* and *FOXA2*, and transcripts associated with the DE developmental program including *FGF17*, *EOMES* and *TBX3* (**Fig. 5-5a**). The expression profiles of DE cells derived on Matrigel™ or laminin-521 were highly comparable and DE transcripts including *SOX17*, *EOMES* and *TBX3* were among the top 20 DEGs compared to undifferentiated hiPSCs (**Fig. 5-5b**). There were very few DEGs Between DE cells differentiated from hiPSCs on Matrigel™ compared to laminin-521 (**Fig. 5-6a**). Profiling of our DE gene set list showed both DE populations had highly similar transcriptomes and were enriched for DE genes (**Fig. 5-6b**). We did observe lower but continued expression of pluripotency transcripts including *NANOG* and *SOX2* in both DE samples compared to hiPSCs, but pluripotency genes were not significantly expressed in DE cells in our differential gene expression analysis (**Fig. 5-6a, b**).

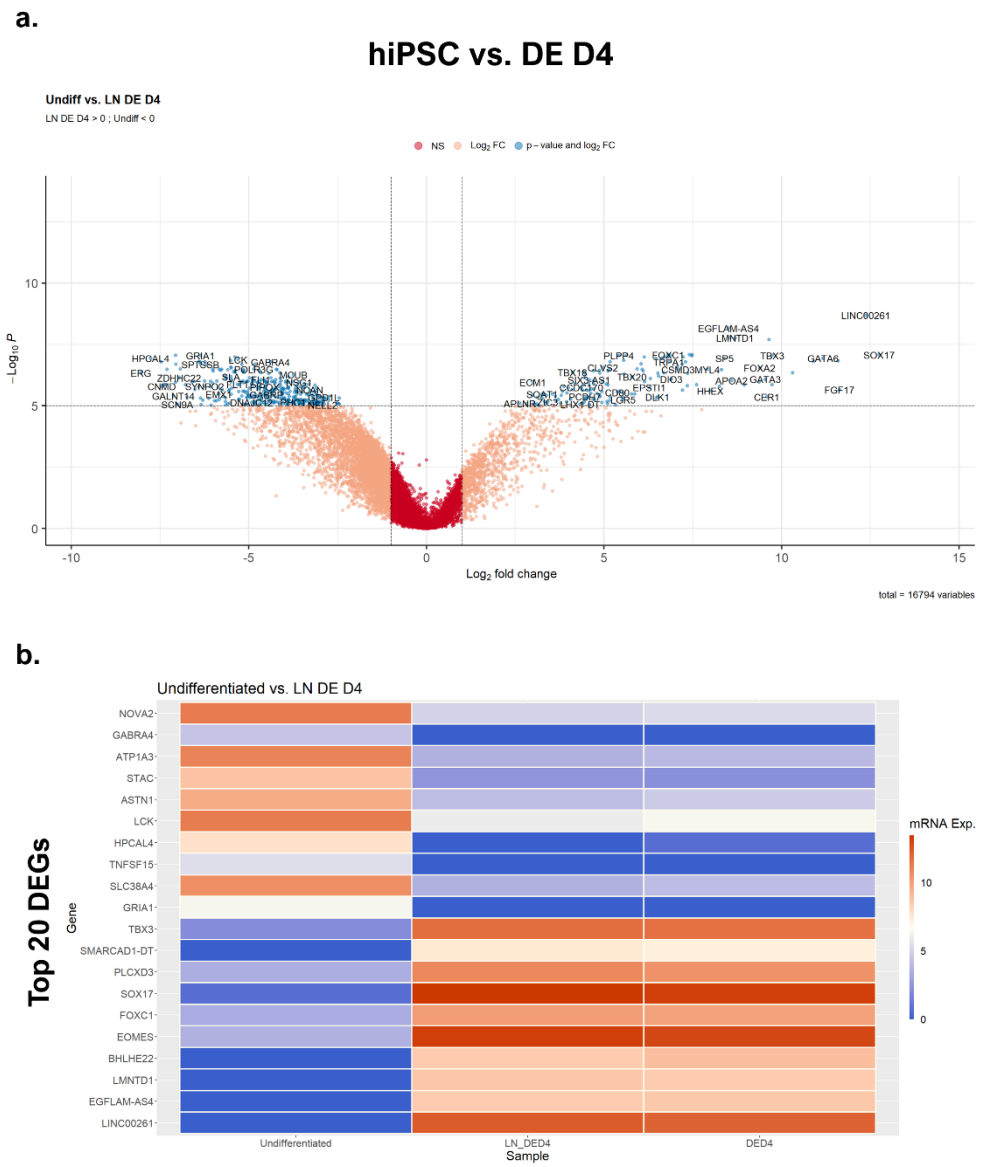


Figure 5-5. Transcriptomic profiling of hPSC-derived DE cells in xeno-free conditions.

(a) Volcano plot of differentially expressed genes in undifferentiated hiPSCs versus day 4 DE cells. Statistically significant ($P \leq 0.05$) changes are highlighted. **(b)** Heatmap of the top 20 DEGs in undifferentiated hiPSCs versus DE D4 cells differentiated in XFL medium on laminin-521 (LN) and Matrigel™.

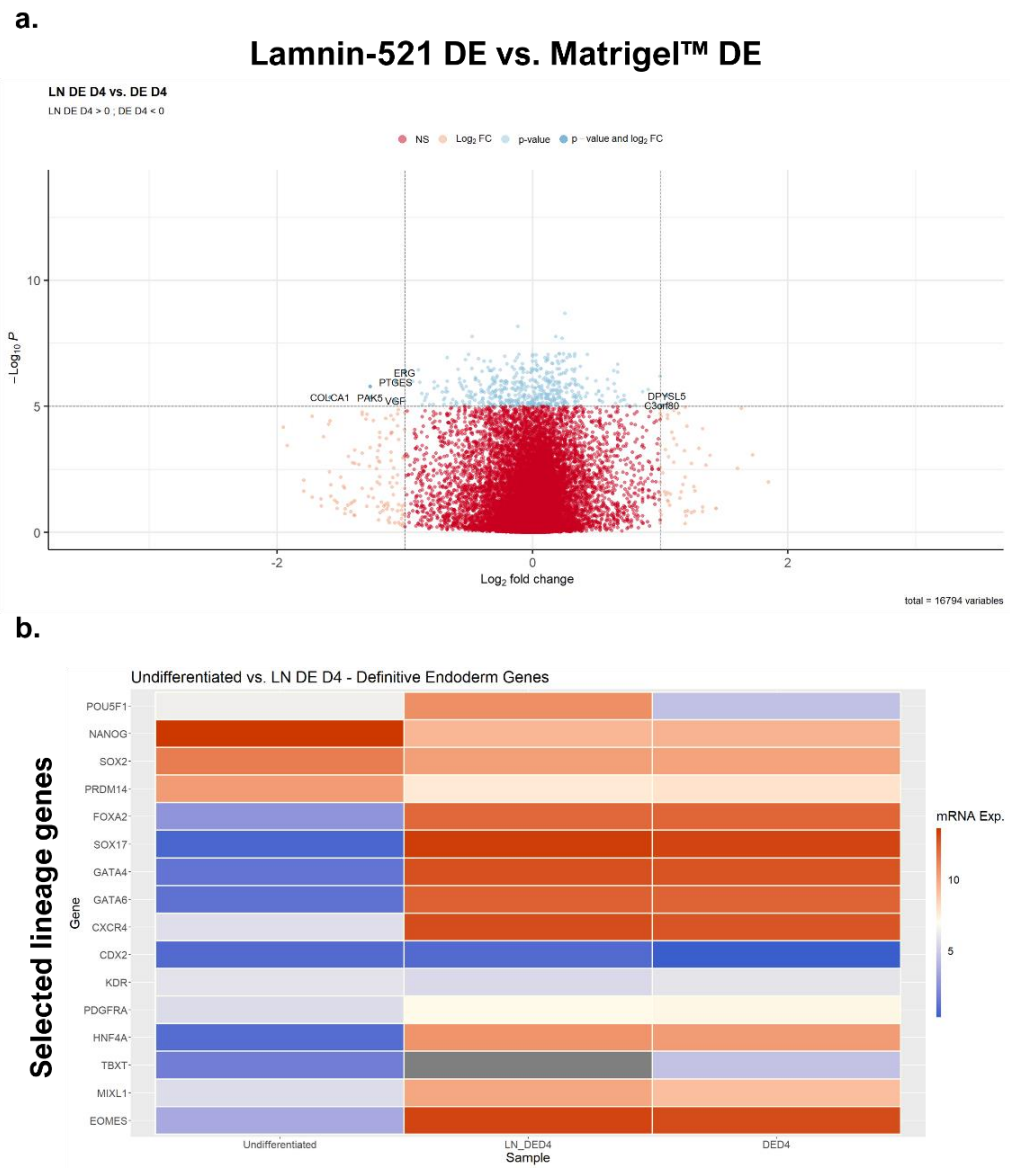


Figure 5-6. Transcriptome profiles of DE cells derived from hiPSCs in XFL medium on Matrigel and laminin-521 are comparable.

(a) Volcano plot of differentially expressed genes in DE cells differentiated from hiPSCs on laminin-521 versus DE cells differentiated from hiPSCs on Matrigel™. Statistically significant ($P \leq 0.05$) changes are highlighted. **(b)** Heatmap displaying the expression of selected lineage markers in undifferentiated hiPSCs versus DE cells derived on laminin-521 and Matrigel™.

5.2.4. Laminin-521 is a suitable substrate for induction of CPM(hi) cells from hPSC-derived DE

After confirming the optimal foregut and lung progenitor differentiation conditions in XFL medium on Matrigel™, we wanted to determine if we could differentiate hiPSCs to lung progenitor cells in completely defined, xeno-free conditions. We performed our lung differentiation protocol with BE.31 ($SFTPC^{WT/WT}$) hiPSCs on three different laminin isoforms, laminin-521, -121 and -511 and showed comparable percentages of CPM(hi) cells on lung specification day 4 by flow cytometry (**Fig. 5-7a**). Given we could maintain hiPSCs on laminin-521 we decided to continue with this isoform for subsequent experiments to maintain consistency. We repeated our lung differentiation and performed FACS for CPM on lung day 6. The percentage of CPM(hi) cells obtained was 48.8.% ± 23.7% (mean ± SEM) (N=3) with a maximum percentage of ~74% CPM(hi) cells on laminin-521 (**Fig. 5-7b, c**). On Matrigel™, the percentage of CPM(hi) cells was 35.8% ± 15.31% (mean ± SEM) (N=3) and a maximum of ~48% CPM(hi) cells was obtained (**Fig. 5-7c**). More than 90% of cells on both substrates expressed CPM, but the percentage of CPM+ cells was greater on Matrigel™ than laminin-521 (**Fig. 5-7c**). These data demonstrate that using our optimised differentiation protocol on laminin-521 in XFL medium, we can robustly generate CPM(hi)-expressing cells with greater efficiency than widely used matrices like Matrigel™.

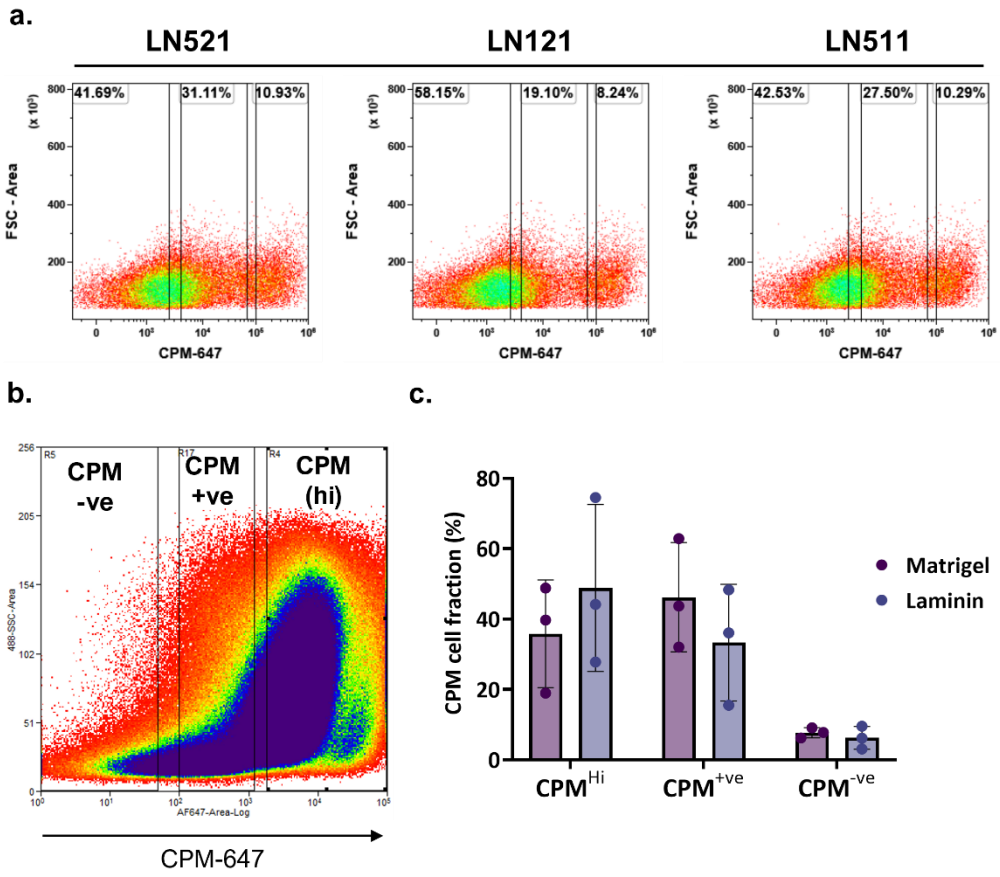


Figure 5-7. CPM(hi) cells can be specified from hiPSCs in fully defined, xeno-free conditions.

(a) Flow cytometry analysis of BE.31 ($SFTPC^{WT/WT}$) lung day 4 cells for CPM following differentiation in XFL medium on three different isoforms of recombinant laminin. **(b)** Representative FACS plot showing gating to isolate CPM fractions on day 6 of lung differentiation on laminin-521 with **(c)** quantification of cells differentiated in XFL medium on laminin-521 or Matrigel™ for BE.31 ($SFTPC^{WT/WT}$) hiPSCs. Bars represent mean ± SEM (N=3).

5.2.5. Defining the transcriptome of CPM(hi) lung progenitor cells derived in xeno-free conditions

We sought to determine the transcriptomic changes during our differentiation on laminin-521 so we performed RNA-seq on cells at sequential time points during lung specification and FACS isolated CPM populations. By PCA we found hiPSCs and DE cells clustered separately, but all other lung and CPM samples clustered together, similar to our findings for differentiated cells on Matrigel™ (**Fig. 5-8a**). We first compared gene expression profiles of undifferentiated hiPSCs with cells during lung specification and CPM populations. Among the top DEGs, transcripts related to pluripotency such as *NANOG*, *ESRG* and *PRDM14* were enriched in hiPSCs as expected, but we did not observe significant upregulation of lung progenitor genes during differentiation (**Fig. 5-9a**). When we performed differential gene expression analysis of the same populations using our selected lung gene set, we found lung progenitor markers including *NKX2.1* and *CPM* expressed as early as lung day 2 which aligns with our previous findings that CPM(hi) lung progenitor cells emerge after 2 days of lung specification, but expression was reduced on lung day 8 but enriched in CPM(hi) cells sorted on day 6 (**Fig. 5-9b**). We also showed the thyroid marker *PAX8* was not expressed across any differentiation stage, or transcripts related to surfactant protein processing, which we would not expect to be upregulated at this developmental timepoint (**Fig. 5-9b**).

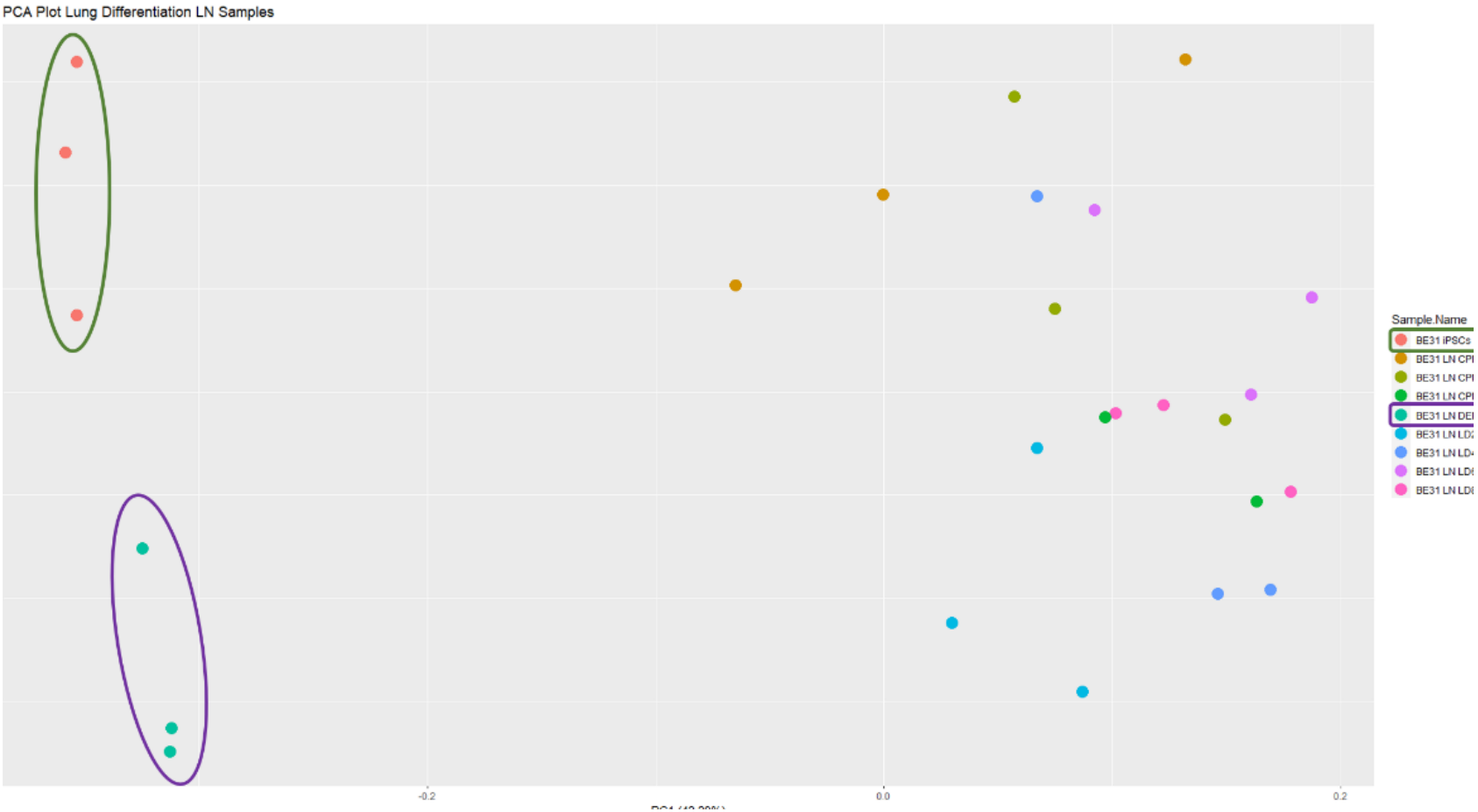


Figure 5-8. The transcriptome of lung cells and CPM populations differentiated in xeno-free conditions cluster together.

(a) PCA of gene expression variance across cells at major developmental stages during differentiation of hiPSCs to lung progenitor cells on laminin-521 in XFL medium.

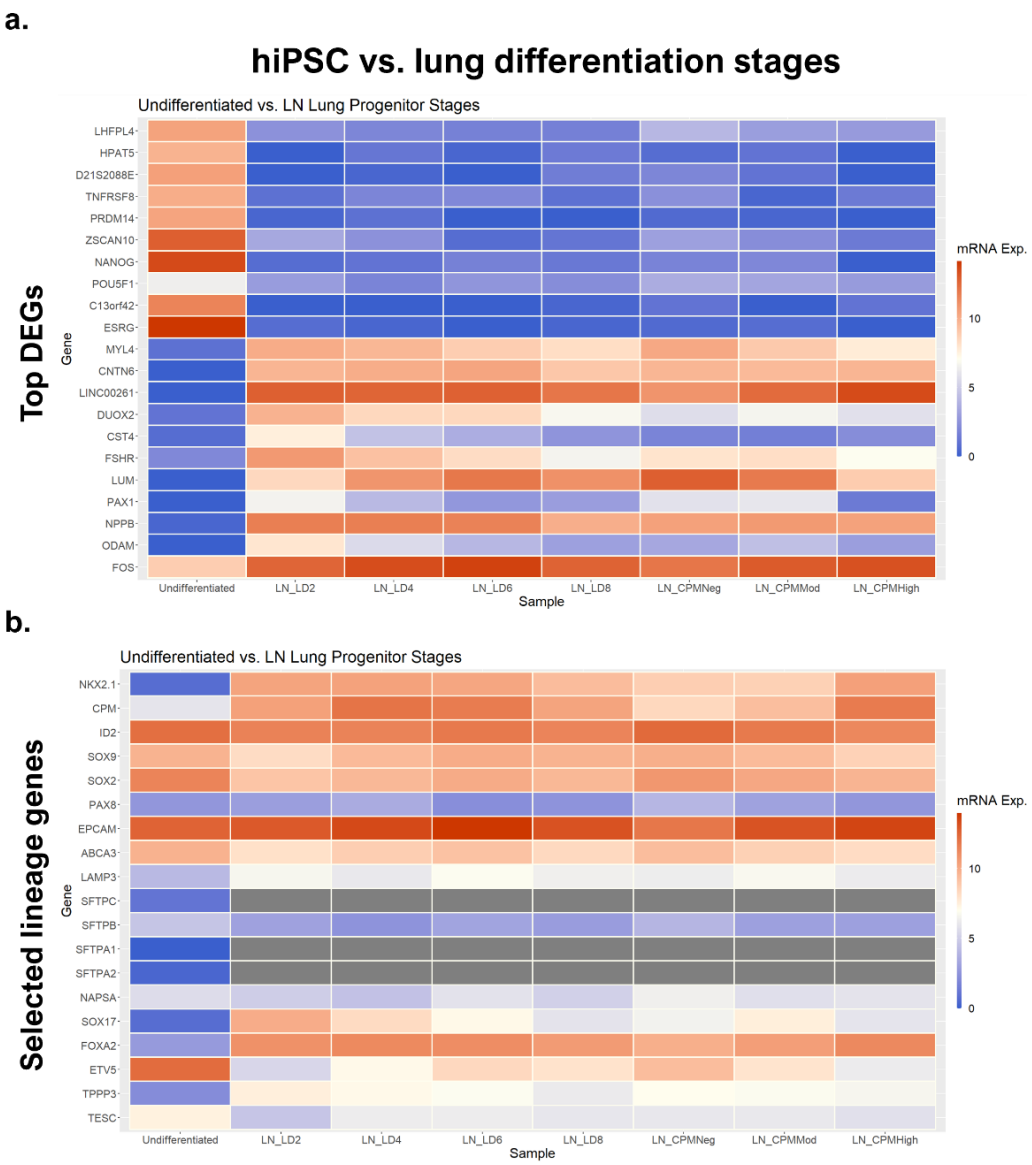


Figure 5-9. Transcriptional profiling of differentiated lung cells and CPM isolated populations.

(a) Heatmap of the top DEGs in undifferentiated hiPSCs versus cells on every 2 days of an 8-day lung differentiation and isolated CPM-. CPM+ and CPM(hi) populations. **(b)** Heatmap showing the expression of selected lineage markers in undifferentiated hiPSCs versus cells on every 2 days of an 8-day lung differentiation and isolated CPM-, CPM+ and CPM(hi) populations.

5.2.6. Comparable transcriptomic profiles of CPM(hi) cells differentiated on laminin-521 or Matrigel™

Next, we wanted to identify if any transcriptomic changes were induced by the basement membrane used to support cell attachment and differentiation. We analysed DEGs between CPM(hi) cells differentiated from hiPSCs on laminin-521 to Matrigel™ and observed very few significant changes in gene expression, although *NANOG* was among the top DEGs enriched in CPM(hi) cells on Matrigel™ compared to laminin-521 (**Fig. 5-10a**). Differential gene expression analysis revealed highly similar gene expression profiles between CPM(hi) populations on Laminin-521 and Matrigel™ when compared to undifferentiated hiPSCs. Among the top DEGs, *SHH* was enriched in both CPM(hi) populations as well as *GATA6*, and *SPINK1,* a gene related to the EGF signalling pathway (**Fig. 5-11a**). Profiling of our lung progenitor gene list showed *NKX2.1* and *CPM* were marginally upregulated in CPM(hi) cells on Matrigel™ compared to CPM(hi) cells on laminin-521, and *FOXA2* and *EPCAM* were enriched in both CPM(hi) cells (**Fig. 5-11b**). However, despite these marginal differences, expression profiles of both CPM(hi) populations were highly comparable. This demonstrates that laminin-521 is a suitable substrate for differentiation and CPM(hi) cells enriched for lung epithelial progenitor cell transcripts can be generated in xeno-free conditions.

a.

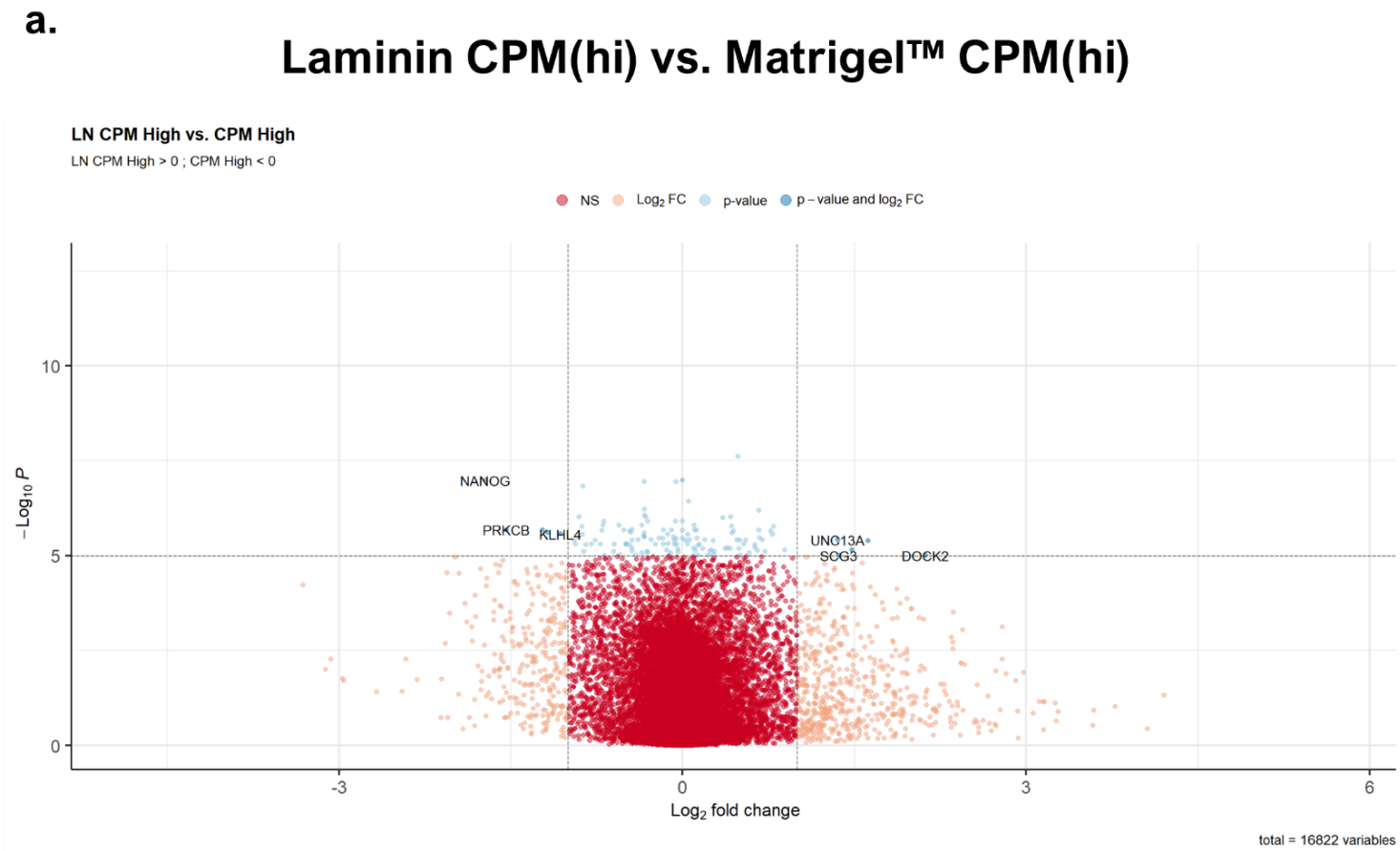


Figure 5-10. CPM(hi) cells derived from hiPSCs on laminin-521 and Matrigel™ are not associated with significantly different transcriptomes.

(a) Volcano plot of DEGs comparing CPM(hi) cells differentiated from hiPSCs on laminin-521 versus Matrigel™. Statistically significant ($P \leq 0.05$) changes are highlighted.

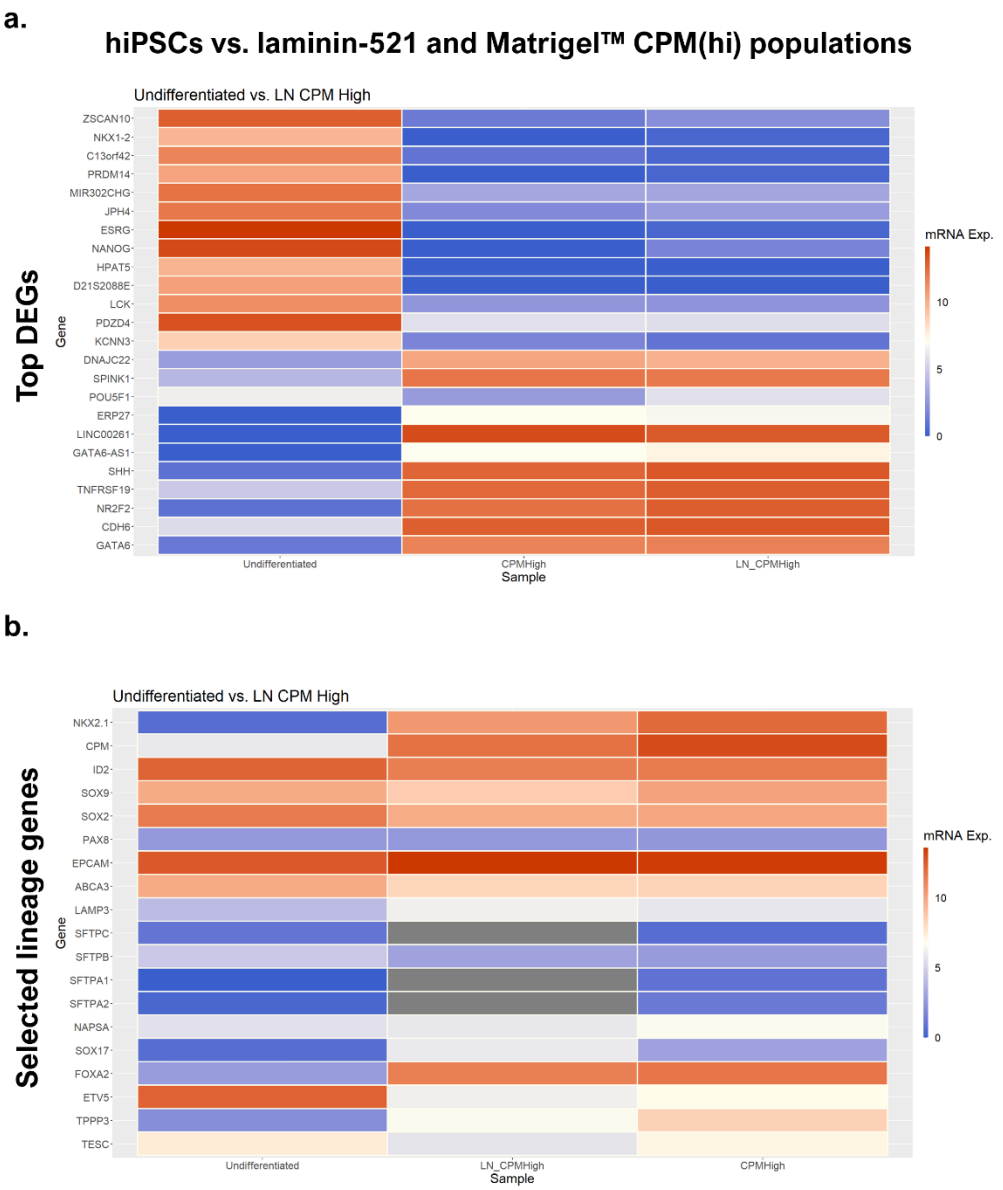


Figure 5-11. The transcriptome of CPM(hi) cells differentiated on laminin-521 is enriched for lung epithelial progenitor transcripts.

(a) Heatmap of the top DEGs in undifferentiated hiPSCs versus CPM(hi) cells derived on laminin-521 (LN) and Matrigel™. **(b)** Heatmap displaying the expression of selected lung progenitor lineage markers in undifferentiated hiPSCs versus CPM(hi) cells differentiated on laminin-521 (LN) and Matrigel™.

5.2.7. Generation of NaPi2B+/EPCAM+ cells in xeno-free conditions using recombinant 3D biosilk fibres

In order to demonstrate the capacity to generate type 2 AECs from CPM(hi) cells under xeno-free conditions, we had to replace Matrigel™ which we use to provide a 3D matrix for the formation of organoids to facilitate maturation to alveolar cells. We utilised a commercially available, defined, animal-free biosilk substrate that is composed of recombinant spider silk proteins with RGD-containing motifs to provide a 3D environment for cell attachment and growth. CPM(hi) cells were first cultured in an aggrewell to promote the formation of 3D structures and were then seeded in the biosilk substrate mixed with laminin-111. Cells aggregated and formed strands of 3D tissue along the silk fibres and could be maintained for 20 days after seeding (**Fig. 5-12a**). After 20 days in culture, we observed the emergence of a NaPi2B+/EPCAM+ population by flow cytometry, and the percentage of NaPi2B/EPCAM+ cells was 17.33 ± 2.35% (mean ± SEM) (N=2) (**Fig. 5-12b**). This confirmed we could differentiate hiPSCs to NaPi2B+/EPCAM+ cells in defined, xeno-free conditions.

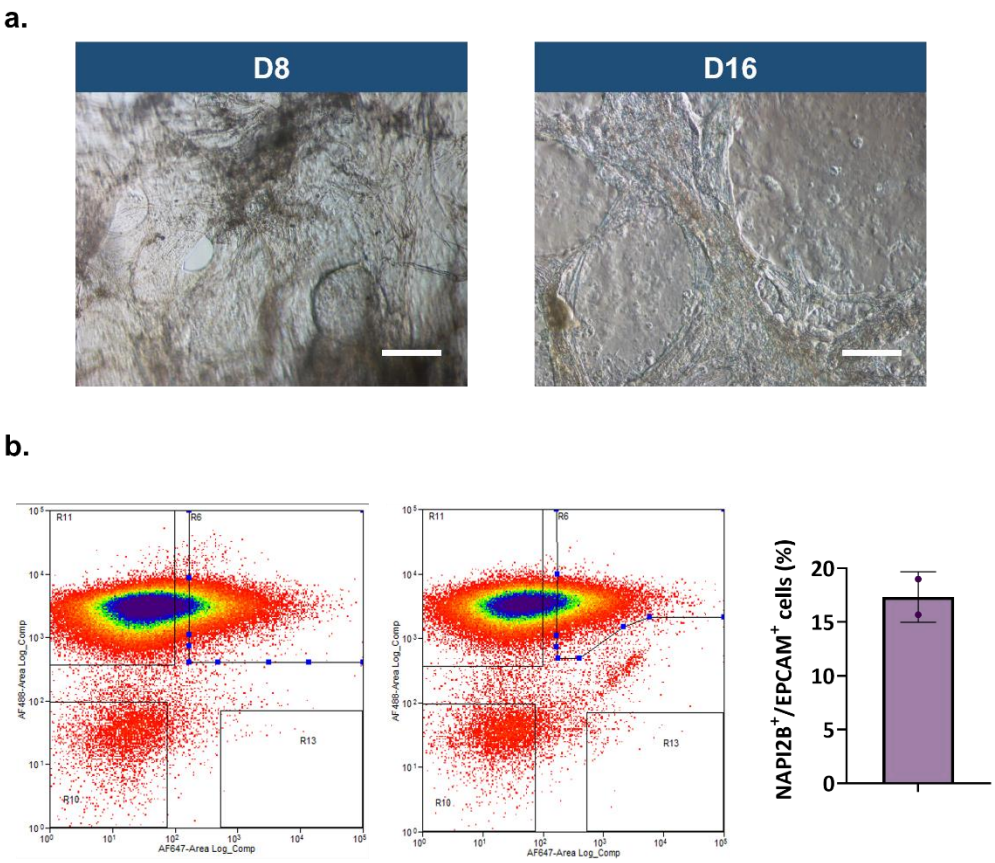


Figure 5-12. NaPi2B+/EPCAM+ cells can be generated in xeno-free conditions using recombinant biosilk.

(a) Representative light microscope images of CPM(hi) cells at day 8 and 16 in XFL medium in 3D recombinant biosilk fibres. **(b)** FACS analysis and quantification of BE.31 ($SFTPC^{WT/WT}$) CPM(hi) for NaPi2B and EPCAM following differentiation in XFL medium in 3D biosilk in the presence of CHIR, FGF7, FGF10, Dexamethasone, IBMX and cAMP (N=2). Bar represents mean ± SEM (Figure 12 was performed by Carlos Sainz Zuniga as previously described).

5.3. Discussion

Major advances are being made toward translating cell therapies to the clinic. Therefore, the development of defined, xeno-free platforms to derive adult cell types from hPSCs is becoming of considerable importance. To date, most differentiation protocols rely on materials from animal sources such as Matrigel™ and BSA to support cell survival, but this has raised concerns surrounding transmission of pathogens and causing xenogeneic immune responses in patients (Günzburg & Salmons, 2000; Jang et al., 2013). As such, the application of these protocols is limited as they do not adhere to GMP guidelines for stem cell-derived therapeutics. The presence of undefined factors in cell culture medium can also lead to reduced reproducibility and makes protocol standardisation difficult. Therefore, to overcome these limitations, we used our optimised protocol in defined medium on human recombinant laminin to develop a novel, xeno-free *in vitro* platform to differentiate hiPSCs to type 2 AECs.

We first demonstrated that hiPSCs could be propagated on recombinant laminin-521 and expressed core pluripotency genes, consistent with previous findings (Rodin et al., 2014). We used a ROCK inhibitor to promote the survival of dissociated single hiPSCs, although laminin-521 can support cell survival in absence of ROCK inhibitors, but we did not test this (Rodin et al., 2014). Laminin-521 was able to support the differentiation to DE with the same efficiency as Matrigel™ using our optimised protocol of CHIR, activin A, and PI3K inhibition via LY294002 across several cell densities tested. We showed the differentiation was robust across three cell lines, including genetically modified hiPSCs harbouring a SFTPC mutation. A short pulse of WNT signalling by CHIR-mediated inactivation of GSK3 is commonly used for maximal induction of DE and outperforms recombinant Wnt proteins (Diekmann et al., 2015; Diekmann et al., 2019; Loh et al., 2014; Teo et al., 2014). Interestingly, it has been shown that substrates that engage in integrin binding such as Matrigel™ are refractory to DE differentiation by activation of integrin-linked kinase and heightened Akt signalling (Wrighton et al., 2014). Although laminin-521 also engages in integrin binding, we showed this did not inhibit DE specification, perhaps owing to PI3K inhibition which diminishes Akt signalling (Rodin et al., 2014).

Differentiation to DE on laminin-521 was associated with significant transcriptomic changes and we observed the expression of putative DE genes and transcripts related

to early embryonic patterning such as *SOX17*, *FOXA2*, *GATA6* and *HHEX*. Other endoderm related transcripts such as *FGF17* were also upregulated . (Fisher et al., 2017)We also identified the expression of *CER1*, an antagonist of NODAL that plays major roles in early anterior-posterior patterning of the developing mouse embryo (Yamamoto et al., 2004). CER1 is expressed in hPSCs and has been proposed to cooperate with other NODAL antagonists, WNT and TGFβ pathways to influence lineage specification of pluripotent cells (Besser, 2004). Importantly, we showed the gene expression profiles of DE cells differentiated on laminin-521 and Matrigel™ were highly similar. Differentiation of hPSCs to DE under xeno-free conditions in suspension culture and bioreactors has been reported previously, but this approach is technically more challenging and variables like cell density, which is critical for endoderm development, are harder to control (Diekmann et al., 2019; Sahabian et al., 2019). Therefore, 2D monolayer cultures are easier to handle and reproduce and provide a cost-effective way to generate near-pure DE populations on physiologically relevant substrates for downstream applications.

Laminins are important constituents of the basement membrane of the developing lung, and in particular the laminin α5 chain is required for normal lung development and alveologenesis (Nguyen & Senior, 2006). We demonstrated CPM(hi) lung epithelial progenitor cells could be induced on laminin isoforms-521, -121 and -511 with similar efficiency, although the α5 chain containing isoforms led to a slightly more distinct population. We observed that differentiation on laminin-521 led to more efficient induction of CPM(hi) cells compared to Matrigel™. Matrigel™ is the most widely used substrate for propagation and differentiation of hPSCs owing to its rich composition of ECM components to facilitate cell attachment and growth. Although the major constituents are collagen, laminin and entactin, more than 1800 other proteins have been identified in Matrigel™ including structural proteins, growth factors and soluble cytokines (Hughes et al., 2010; Talbot & Caperna, 2015). Growth-factor reduced Matrigel has been developed as an alternative substrate with lower growth factor concentrations. However, there is still considerable batch variability both in type and concentration of proteins present. Comparison of growth factor reduced Matrigel™ batches between two manufacturers revealed only ~53% similarity, and comparison to standard Matrigel identified 424 unique peptides (Aisenbrey & Murphy, 2020; Hughes et al., 2010). ECM components are known to significantly influence stem cell

behaviour and fate, and Matrigel™-derived factors have been shown to impact lineage specification (Murphy et al., 2014; Wrighton et al., 2014). It is likely owing to these factors why our differentiation on growth factor reduced Matrigel™ resulted in less efficient differentiation to CPM(hi) lung progenitor cells. Laminin-521 is therefore a suitable substrate for differentiation. However, further understanding of the ECM niche which facilitates lung progenitor development *in vivo* will lead to more physiologically relevant basement membrane complexes for *in vitro* differentiation. Multitude synthetic materials that mimic constituents of the basement membrane have been developed, though none have been utilised for PSC differentiation to the lung (Aisenbrey & Murphy, 2020).

Transcriptome analysis revealed that hiPSCs clustered together as did DE cells, but interestingly we observed broad clustering of all cells sequenced at sequential stages of lung specification and CPM populations derived on laminin-521. This was consistent with our findings from differentiations performed on Matrigel™. These cells are likely very similar as they derive from a common progenitor and might reflect progressive states shifting toward a lung epithelial progenitor cell fate, associated with small changes in the transcriptome. This would support our previous results where the onset of CPM(hi) cells was as early as 2 days after the onset of differentiation. Similar to this, isolated SFTPC+ and SFTPC- populations were shown to cluster together, and it was proposed the SFTPC- cells reflected a cell state that had not fully committed to the type 2 AEC programme (Jacob et al., 2017). Surprisingly, we did not observe transcripts related to lung progenitor cells among the top differentially expressed genes during differentiation or in our CPM isolated cells, but analysis of select lineage genes revealed the expression of *NKX2.1* and *CPM* during lung specification and enrichment in CPM(hi) cells. *NKX2.1* expression diminished from day 6 to 8 of differentiation confirming day 6 is a suitable timepoint to perform FACS purification. There was also no expression of the thyroid marker *PAX8* at any stages, further confirming the lung-identity of our cells. Our observation of *FOXA2* enrichment in CPM(hi) cells is consistent with other proposed roles of FOXA2 in regulating lung morphogenesis and maturation of the alveolar epithelium (Wan et al., 2005; Wan et al., 2004). The highly comparable gene expression profiles of CPM(hi) cells specified on laminin-521 and Matrigel™ add further support to laminin-521 being a suitable substrate for xeno-free lung differentiation. In both populations, *SHH* was enriched which is known to

coordinate signalling between the developing mesenchyme and adjacent epithelium to drive lung fate (Rankin et al., 2016). Previous findings have shown expression of FoxA1 and FoxA2 can regulate the expression of Shh in the respiratory epithelium to influence mesenchyme-derived signals during branching morphogenesis (Wan et al., 2005).

One interesting observation from our transcriptomic analysis was the enrichment of *GATA6* in our CPM(hi) populations. *GATA6* is a transcriptional regulator and a putative marker of DE cells, but it has been identified to play diverse roles in the development of lung, among other organs. GATA6 has been implicated in alveolar maturation in the mouse and can bind to and activate NKX2.1 to drive expression of respiratory-related genes (Shaw-White et al., 1999; Yang et al., 2002). However, overexpression of GATA6 can impair distal lung differentiation (Koutsourakis et al., 2001). Consistent with this, Liao et al. (2018) reported an inverse relationship between *GATA6* expression and percentage of NKX2.1+ cells, such that hiPSC lines with lower endogenous levels of *GATA6* led to more efficient induction of NKX2.1+ cells. This finding was confirmed with short-interfering RNA knockdown of GATA6 and complete GATA6 knockout hiPSCs displayed enhanced capacity to generate NKX2.1+ cells, although reported efficiencies in DE induction were low and cell viability was compromised. It is possible that other GATA transcription factors like GATA4 acted in a compensatory manner. They showed GATA6 was necessary for maturation to alveolar identity. Furthermore, they identified a role for GATA4 and GATA6 expression during DE patterning in favouring hepatic over lung differentiation. This suggests a different role for GATA transcription factors in human lung development. These findings were corroborated in another study showing a SOX2+GATA6(low) AFE population is enriched for lung epithelial progenitor cells (Brafman et al., 2013). During our differentiation protocol we do observe emergence of off-target hepatic cells. Further transcriptomic profiling of our hiPSC lines might reveal higher levels of endogenous *GATA6* expression that encodes a bias to hepatic fates during differentiation. However, our efficiency of differentiation to lung progenitor cells is higher than those reported in Liao et al. (2018), and the enrichment of *GATA6* in our CPM(hi) populations suggest it might play a role in early human lung specification, but this required further investigation. We also observed higher expression of *NKX2.1* and *CPM* in CPM(hi) cells differentiated on

Matrigel™ compared to laminin-521, but we would need to investigate this further to determine the functional significance, if any.

The importance of mimicking a physiologically relevant 3D environment to promote the maturation of progenitor cells to mature distal lung cells has been demonstrated by us and others (Chen et al., 2017; Jacob et al., 2017; Miller et al., 2019; Yamamoto et al., 2017). Most of these protocols rely on generating self-assembled 3D organoids in droplets of Matrigel™ which is not a physiologically relevant ECM. We have shown here that we can induce maturation of CPM(hi) lung progenitor cells to cells co-expressing NaPi2B and EPCAM, which we previously identified as a population enriched for markers of type 2 AECs, in completely xeno-free 3D matrices composed of recombinant silk proteins and laminin-111. The formation of 3D clusters of cells and the induction of NaPi2B expression suggests the biosilk could support the differentiation of alveolar cells. There have been few other reports of 3D Matrigel™-free systems to support alveolar differentiation. Hoffman et al. (2023) described the use of hydrogels made from decellularized alveolar-rich ECM from human lung which could support proliferation of hPSC-type 2 AECs. Another study developed a core-shell microfibres comprised of an alginate gel and showed they could support the aggregation of lung progenitor cells into 3D clusters that could differentiate into airway and alveolar fates (Ikeo et al., 2021). Further findings have reported synthetic hyaluronic acid-based hydrogels fabricated with small cavities could support the assembly of hPSC-derived type 2 AECs into 3D alveolar epithelial organoids that retain functional characteristics of type 2 AECs in absence of Matrigel™ or animal-derived laminin/entactin (Loebel et al., 2021). These studies show there are several approaches being taken to generate defined ECM niches for type 2 AECs. However, these systems still rely on the presence of animal-derived components in media and Matrigel™ to generate hPSC-derived type 2 AECs. Although we need to further characterise our platform, we have shown the ability to differentiate hiPSCs to NaPi2B+ lung cells in completely defined, xeno-free conditions.

6. General Discussion

In this work, we have developed a novel platform for the directed differentiation of hiPSCs to type 2 AECs of the lung in completely defined, xeno-free conditions by recapitulating major stages of embryonic lung development. This work is highly important for several reasons: there is an increasing need for human-relevant models of lung disease to facilitate the discovery of therapeutics for diseases including IPF, COPD, cystic fibrosis and lung cancer; we require models to understand late fetal and early neonatal periods of lung development that would be otherwise inaccessible to researchers due restrictions on human embryo research and limited access to fetal tissue; protocols with defined conditions will allow us to faithfully interrogate signalling mechanisms in development and disease without the presence of confounding variables, and finally, to provide a method to generate a therapeutically-relevant source of cells for clinical translation.

We first described the reprogramming of dermal fibroblasts using non-integrating Sendai virus from a patient diagnosed with familial-linked IPF resulting from a heterozygous missense mutation in the BRICHOS domain of the SFTPC gene. The mutation was caused by a single base pair substitution replacing alanine with guanine (A>G). Therefore, this mutation was amenable to genetic editing using newly developed base editing technology to catalyse the conversion of G>A or A>G to generate homozygous-corrected or -mutant isogenic control cell lines, respectively (Anzalone et al., 2020). Using isogenic cell lines that share the same genetic background is a powerful method to understand the impact of pathogenic mutations. In this study we focussed on a single *SFTPC* mutation as proof-of-principle, but this approach is applicable to many other single nucleotide variants in *SFTPC*, and in other genes including *SFTPA2*, *ABCA3* and *MUC5B* associated with IPF (Kropski et al., 2015).

Considerable progress has been made in developing protocols to direct the differentiation of hPSCs to cell types of interest, although many challenges remain. One of the main considerations in the use of hiPSCs is accounting for heterogeneity between cell lines. Deriving hiPSC lines from different individuals, and sometimes from different cells of origin, results in significant variability in DNA methylation patterns and gene and protein expression which impact differentiation propensity and

morphological characteristics (Kilpinen et al., 2017). These findings have prompted the stem cell community to define a set of standards for characterising hiPSC lines for use in research. This includes detailed characterisation of pluripotency, routine screening for genetic changes, rigorous quantitative assessment of differentiation potential and functional analysis ("Setting standards for stem cells," 2023). This will inevitably lead to greater reproducibility of published protocols across hiPSC lines and laboratories (Volpato & Webber, 2020). We ensured the full characterisation of newly derived hiPSC lines by flow cytometry, immunofluorescence imaging and karyotype analysis.

When we applied a previously established differentiation protocol to our newly derived hiPSC lines, we observed resistance to differentiation and poor induction of relevant gene transcripts. This led to us optimising the differentiation protocol from the starting point. Beginning with a homogeneous population of endoderm cells is critical to enable efficient differentiation to lung epithelium. Our findings showing the incomplete suppression of the pluripotency program with recombinant WNT3a and activin A, and the synergistic effect of BMP4, FGF2 and PI3K on DE induction efficiency, was consistent with previous reports (Teo et al., 2014; Teo et al., 2012). We found 1-day treatment with CHIR alongside 2-day treatment with a PI3K inhibitor and continued activin A signalling was optimal for DE differentiation. This was consistent in defined medium and on recombinant laminin, and we observed no adverse effect on cell viability under these conditions. To the best of our knowledge, this is also the first fully defined, xeno-free 2D DE differentiation protocol described. Other xeno-free 3D platforms to generate DE from hPSCs have been reported (Diekmann et al., 2019; Sahabian et al., 2019). We did not find 3D culture was necessary at this stage, and we demonstrated that 2D monolayer differentiation was highly reproducible and could lead to a near homogeneous population of DE cells. Whole transcriptome analysis revealed our DE population derived under xeno-free conditions was enriched for putative DE genes, validating our protocol.

Restricting the potential of DE cells to an anterior foregut fate is a highly important stage during differentiation, but it remains poorly characterised. Our findings showed 3-day inhibition of TGFβ and BMP signalling to inhibit SMAD signalling generated a population of cells that was competent to become lung epithelium in response to WNT and BMP signals, consistent with findings by Green et al. (2011). Combination of

transient dual-SMAD inhibition followed by TGFβ and WNT inhibition could pattern DE to AFE, but showed no further benefit for lung induction efficiency, in contrast to other findings (Huang et al., 2014). One of the major limitations of current lung differentiation protocols is the emergence of non-lung endoderm lineages (Gotoh et al., 2014; Hein et al., 2022; Jacob et al., 2017). In our system, we consistently observed differentiation to the hepatic lineage, suggesting a precursor population was specified after patterning to AFE, or there was a fate switch during lung specification. Previous findings have observed by scRNA-seq a CDX2+ hindgut population that can give rise to intestinal organoids after dual-SMAD inhibition for 3 days (Mithal et al., 2020). Similar findings using scRNA-seq to map the fate trajectory of hPSCs during lung differentiation revealed early hepatic precursors emerge during DE specification, with lung progenitor cells arising later during foregut patterning, along distinct trajectories. On day 15 of differentiation there was co-existence of an NKX2.1+ lung progenitor population, and a population of NKX2.1- cells enriched for early fetal liver genes. They implicated several transcription factors with known roles in foregut and lung development, including *FOXA2*, *GATA4* and *GATA6*, in hepatic specification (Ori et al., 2021). *FOXA2* has been shown to coordinate activation of hepatic transcription factors, and genes such as *HHEX*, *TBX3* and *GATA6* and *HNF4A* are required for hepatic development (Feng et al., 2023; Lemaigre, 2009). Furthermore, other findings have reported *GATA6* is refractory to differentiation of hPSCs to NKX2.1+ cells (Liao et al., 2018). Our transcriptome analysis did reveal the expression of *TBX3*, *HHEX* and *HNF4A*, along with DE markers including *SOX17*, *FOXA2*, *GATA4*, *GATA6* and *CXCR4* in our DE population, which suggests the presence of early liver precursor cells. This would explain the expression of fetal liver genes like *ALB* and *AFP* after lung differentiation, consistent with the findings above. This suggests signalling networks could be further refined to restrict to potential to lung identity. However, this might prove difficult given the transient nature of foregut morphogenesis and specification and the shared signalling pathways and transcription factors governing cell fate (Zorn & Wells, 2009).

We found other variables at this stage of the protocol were also critical and showed modulation of cell density prior to foregut patterning could significantly improve the efficiency of lung specification. This was further enhanced by replacing the small molecule CHIR with recombinant WNT3a and R-spondin for WNT pathway activation and using recombinant laminin during lung differentiation. These findings were

aligned with the known off-target effects of CHIR, and the undefined proteins and signalling molecules found in Matrigel™ that can significantly influence differentiation (Hu et al., 2021; Hughes et al., 2010; Laco et al., 2018; Wrighton et al., 2014). We did not observe a synergistic effect of FGF signalling on lung specification but did show RA signalling was necessary to permit lung induction, supporting previous reports (Rankin et al., 2018). Nonetheless, we still reported the presence of non-lung endoderm types which justified our adoption of a strategy to enrich for lung progenitor cells.

Based on previous findings we selected CPM as a cell surface marker to purify lung epithelial progenitor cells (Gotoh et al., 2014). On both Matrigel™ and under xeno-free conditions, our protocol led to the induction of a high-expressing CPM population that was shown to be enriched for lung progenitor markers such as *NKX2.1* by RNA-seq, immunofluorescence staining and qRT-PCR. qRT-PCR analysis identified enrichment of non-lung endoderm markers in CPM+ and CPM- fractions, consistent with findings discussed above (Ori et al., 2021). Our transcriptomic analysis also highlighted significant expression of *ISL1* and *TBX3* in CPM(hi) cells specified on Matrigel™, and *SHH* and *FOXA2* CPM(hi) cells specified on Matrigel™ and laminin-521. *FoxA2* can directly activate *Isl1*, and previous findings have shown *Isl1* can directly regulate *Nkx2.1* in a population of anterior foregut cells that is required for tracheal and oesophageal septation and normal lung lobation in the mouse (Kang et al., 2009; Kim et al., 2019). Further to this, reports have identified a role for *Isl1* in regulating branching morphogenesis of the mouse lung by activating Shh signalling between the epithelium and mesenchyme (Huang et al., 2023). This study found the ablation of *Isl1* led to downregulation of genes including *Sox9*, *Tbx2*, *Tbx3*, *Irx1* and *Irx2* in the lungs. The *Irx* genes are expressed in the lung epithelium during development and are important for lung morphogenesis in the rat (Tuyl et al., 2006). Furthermore, it was demonstrated using scRNA-seq that during *in vitro* differentiation to lung, *ISL1* was upregulated during foregut specification preceding *NKX2.1* expression in the lung progenitor cell trajectory (Ori et al., 2021). These findings suggest a transcription factor cascade of FOXA2 > ISL1 > NKX2.1 and SHH (and other TFs such as TBX3) in regulating specification of the respiratory domain in the foregut and lung epithelial development. These TFs were significantly expressed in our CPM(hi) cells, and *FOXA2* and *SHH* were upregulated across the entire differentiation, suggesting the onset and maintenance of a lung-specific genetic program. It is well understood that SHH plays important roles

during respiratory specification and lung development, and it would be interesting to see if supplementing exogenous SHH would improve the induction of lung cells in our system (Kugler et al., 2015; Rankin et al., 2016). WNT signalling related genes such as *FZD1* and *RSPO3* were also upregulated in our CPM(hi) population, consistent with the role of the WNT/β-catenin pathway in lung epithelium development (Aros et al., 2021; Hussain et al., 2017; Jacob et al., 2017).

Although our CPM(hi) population is enriched for lung progenitor transcripts, we cannot ignore the significant upregulation of the hepatic transcripts *TTR* and *HNF1B*, highlighting a potential subpopulation of cells with hepatic identity. When we induced maturation of CPM(hi) cells to type 2 AECs, we observe differentiation to off-target cell types expressing markers of non-lung lineages, similar to earlier differentiation timepoints. Hurley et al. (2020) has showed hPSC-derived NKX2.1+ cells, which are equivalent to our CPM(hi) cells, could differentiate along diverse trajectories to lung and non-lung lineages, including to CDX2+ hindgut precursors, possibly by WNT signalling overstimulation. Similarly, Hein et al. (2022) showed after repeated NKX2.1+ cell purification, CDX2+ cells could still arise. This highlights the inherent plasticity of hPSC-derived cells. We propose to perform scRNA-seq on our AFE population after 3 days of dual-SMAD inhibition, and on day 6 of lung differentiation so we can understand what populations are present at these timepoints, understand where the heterogeneity in our system arises and map the trajectory of lung progenitor cell specification to further understand the transcriptional networks that govern this. Nonetheless, this further supports our implementation of enrichment steps.

Early stages of differentiation are supported in 2D format, but this does not reflect the complex 3D architecture of native tissues. Therefore, to facilitate differentiation and maturation of hPSC-derived cells, 3D platforms are favoured (Chen et al., 2017; Jacob et al., 2017; Miller et al., 2019; Miller et al., 2018; Rodrigues Toste de Carvalho et al., 2021; Yamamoto et al., 2017). We showed CPM(hi) cells could readily self-assemble into 3D organoids and give rise to type 2 AECs expressing SFTPC, and TEM revealed the presence of lamellar bodies, the organelles responsible for surfactant storage and secretion. Moreover, the detection of SFTPC in the culture medium by ELISA indicated functional maturity.

We found it necessary to include a second enrichment step to maintain the type 2 AEC phenotype. Based on the expression of NaPi2B and EPCAM, our transcriptomic analysis revealed this population was enriched for type 2 AEC transcripts including *SFTPC*, *SFTPB*, *LAMP3* and markers associated with maturation like *ABCA3* and *SFTPA1* (Korogi et al., 2019). Interestingly, *SFTPB* and *SFTPD* were significantly expressed in NaPi2B+/EPCAM+ cells compared to NaPi2B-/EPCAM+ cells, but PCA showed these populations clustered together, and analysis of our selected lineage genes revealed highly similar gene expression profiles. One potential explanation is the NaPi2B+ cells reflect a more mature state, as we did observe marginal increases in gene expression of mature type 2 AEC genes compared to NaPi2B- cells. A similar finding was reported previously where SFTPC+ and SFTPC- populations isolated by a tdTomato reporter line clustered together, reflecting progressive states of cell maturation (Jacob et al., 2017). We performed a transcriptomic comparison of our NaPi2B populations to *SFTPC*-tdTomato+ and *SFTPC*-tdTomato- datasets made publicly available, and we observed some small differences. *SFTPC+* cells displayed higher expression of *SFTPC*, *SFTPB*, *NKX2.1* and *SFTPA2*, but the NaPi2B+ cells expressed higher levels of *ABCA3* and *SFTPA1*. Recently, the same group have uncovered heterogeneity in hPSC-type 2 AECs and showed SFTPC-tdTomato+ cells reflect an immature, more proliferative type 2 AEC expressing less markers related to maturation. Instead, an ABCA3 fluorescence reporter was able to enrich for more mature type 2 AECs, which were shown to be isolated via NaPi2B expression (Sun et al., 2021). Profiling of our gene lists for type 2 AEC proliferating and transitional markers, as well as WNT signalling genes demonstrated increased expression of these transcripts in SFTPC+ compared to NaPi2B+ cells. Recently, a WNT-responsive population of type 2 AECs has been identified in human lung and has been posited to be the progenitor population of the alveoli (Travaglini et al., 2020). These findings suggest NaPi2B+ cells might reflect a more mature type 2 AEC phenotype, validating the use of this marker for enrichment.

Advances in technologies like RNA sequencing have enabled benchmarking of *in vitro* hiPSC-derived cells to primary tissue to determine similarities and differences in the transcriptome. We compared our RNA-seq datasets of NaPi2B+ cells to publicly available datasets of fetal and adult human lung. We observed separate clustering of NaPi2B populations from all other samples but found the expression of type 2 AEC genes was comparable between NaPi2B+ cells and early and late human fetal lung.

The expression of mature markers such as *SFTPA1*, *SFTPA2*, *LAMP3* and *NAPSA* were highly enriched in adult lung, as expected. These findings highlight an important consideration; hPSC-derived type 2 AECs still reflect a fetal-like, immature transcriptomic programme. Strategies that aim to promote the maturation of cells are required to better reflect adult cell physiology and function (Gu et al., 2021). To date, only one condition has been described to support the self-renewal of hPSC-derived type 2 AECs. We manipulated the signalling pathways to reflect conditions used to maintain adult type 2 AECs, including WNT and FGF signalling coupled with TGFβ and BMP signalling inhibition (Katsura et al., 2020; Salahudeen et al., 2020; Youk et al., 2020). We observed the induction of a NaPi2B+ population enriched for transcripts related to airway basal cell genes and type 2 AECs, suggesting a heterogeneous population. Dual-SMAD inhibition has been described to maintain a highly proliferative hiPSC-basal cell population *in vitro*, but expression of NaPi2B in these cells has not been reported to our knowledge (Hawkins et al., 2021).

These observations raise several questions that need further investigation to be answered. It is unclear what populations are induced in hiPSC-derived lung organoids in the presence of CHIR, FGF7, EGF and dual-SMAD inhibition, and whether hiPSC-type 2 AECs transition into metaplastic basal cells *in vitro*, or whether both populations co-exist in culture. Current protocols support the self-renewal of hPSC-type 2 AECs, but they retain an immature phenotype. Whether hiPSC-derived lung progenitor cells can be differentiated to more mature type 2 AECs with manipulation of BMP and TGFβ signalling remains to be demonstrated, although other reports indicate this might be achievable (Frum et al., 2023). It is possible lung progenitor cells need to be differentiated into an immature type 2 AEC before transitioning to a more mature phenotype. This might explain why the adult-like type 2 AEC conditions we tested led to lung progenitor cells becoming airway-fated. However, *in vivo*, type 2 AECs are usually quiescent and only rapidly proliferate in response to injury (Bhaskaran et al., 2007). This highlights an important consideration as to whether enhancing the maturation of stem cell-derived type 2 AECs would diminish their capacity to self-renew and propagate in culture.

Importantly, we demonstrated type 2 AECs differentiated from homozygous SFTPC-mutant hiPSCs displayed aberrant lamellar body biosynthesis and diminished secretion

of SFTPC compared to homozygous-corrected cells. Mutations in the BRICHOS domain of SFTPC are known to impair the folding and trafficking of pro-SFTPC that precedes activation of the unfolded protein response, ER stress and cell death (Mulugeta et al., 2007; Mulugeta et al., 2005). These findings show our differentiation platform can generate hiPSC-type 2 AECs with functionally mature characteristics such as lamellar bodies and surfactant protein synthesis and secretion, which can recapitulate aspects of human lung disease.

Utilising our optimised protocol, our final aim was to demonstrate the ability to differentiate hiPSCs to type 2 AECs in fully defined, xeno-free conditions. We showed we could perform all stages of differentiation in defined medium on laminin-521 substrate, and CPM(hi) progenitors could give rise to NaPi2B+/EPCAM+ cells, that we know to be enriched for type 2 AECs, in recombinant silk protein scaffolds. Further optimisation and characterisation are needed, but to our knowledge, this is the first report of a xeno-free differentiation protocol to generate type 2 AECs.

References

Adewumi, O., Aflatoonian, B., Ahrlund-Richter, L., Amit, M., Andrews, P. W., Beighton, G., Bello, P. A., Benvenisty, N., Berry, L. S., Bevan, S., Blum, B., Brooking, J., Chen, K. G., Choo, A. B. H., Churchill, G. A., Corbel, M., Damjanov, I., Draper, J. S., Dvorak, P., . . . Zhang, W. (2007). Characterization of human embryonic stem cell lines by the International Stem Cell Initiative [Article]. *Nature Biotechnology*, *25*(7), 803-816. https://doi.org/10.1038/nbt1318

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nature Communications*, *9*(1), 1911. https://doi.org/10.1038/s41467-018-04252-2

Aisenbrey, E. A., & Murphy, W. L. (2020). Synthetic alternatives to Matrigel. *Nat Rev Mater*, *5*(7), 539-551. https://doi.org/10.1038/s41578-020-0199-8

Alfahad, A. J., Alzaydi, M. M., Aldossary, A. M., Alshehri, A. A., Almughem, F. A., Zaidan, N. M., & Tawfik, E. A. (2021). Current views in chronic obstructive pulmonary disease pathogenesis and management. *Saudi Pharmaceutical Journal*, *29*(12), 1361-1373. https://doi.org/https://doi.org/10.1016/j.jsps.2021.10.008

Allen, R. J., Porte, J., Braybrooke, R., Flores, C., Fingerlin, T. E., Oldham, J. M., Guillen-Guio, B., Ma, S.-F., Okamoto, T., John, A. E., Obeidat, M. e., Yang, I. V., Henry, A., Hubbard, R. B., Navaratnam, V., Saini, G., Thompson, N., Booth, H. L., Hart, S. P., . . . Jenkins, R. G. (2017). Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. *The Lancet Respiratory Medicine*, *5*(11), 869-880. https://doi.org/10.1016/S2213-2600(17)30387-9

Alysandratos, K. D., Russo, S. J., Petcherski, A., Taddeo, E. P., Acín-Pérez, R., Villacorta-Martin, C., Jean, J. C., Mulugeta, S., Rodriguez, L. R., Blum, B. C., Hekman, R. M., Hix, O. T., Minakin, K., Vedaie, M., Kook, S., Tilston-Lunel, A. M., Varelas, X., Wambach, J. A., Cole, F. S., . . . Kotton, D. N. (2021). Patient-specific iPSCs carrying an SFTPC mutation reveal the intrinsic alveolar epithelial dysfunction at the inception of interstitial lung disease. *Cell Rep*, *36*(9), 109636. https://doi.org/10.1016/j.celrep.2021.109636

Ang, S.-L., & Rossant, J. (1994). HNF-3β is essential for node and notochord formation in mouse development. *Cell*, *78*(4), 561-574. https://doi.org/https://doi.org/10.1016/0092-8674(94)90522-3

Anzalone, A. V., Koblan, L. W., & Liu, D. R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology*, *38*(7), 824-844. https://doi.org/10.1038/s41587-020-0561-9

Arnold, S. J., Hofmann, U. K., Bikoff, E. K., & Robertson, E. J. (2008). Pivotal roles for eomesodermin during axis formation,epithelium-to-mesenchyme transition and endoderm specification in the mouse. *Development*, *135*(3), 501-511. https://doi.org/10.1242/dev.014357

Arnold, S. J., & Robertson, E. J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nature Reviews Molecular Cell Biology*, *10*(2), 91-103. https://doi.org/10.1038/nrm2618

Aros, C. J., Pantoja, C. J., & Gomperts, B. N. (2021). Wnt signaling in lung development, regeneration, and disease progression. *Communications Biology*, *4*(1), 601. https://doi.org/10.1038/s42003-021-02118-w

Barkauskas, C. E., Cronce, M. J., Rackley, C. R., Bowie, E. J., Keene, D. R., Stripp, B. R., Randell, S. H., Noble, P. W., & Hogan, B. L. (2013). Type 2 alveolar cells are

stem cells in adult lung. *J Clin Invest*, *123*(7), 3025-3036. https://doi.org/10.1172/jci68782

Barnes, P. J., Bonini, S., Seeger, W., Belvisi, M. G., Ward, B., & Holmes, A. (2015). Barriers to new drug development in respiratory disease. *European Respiratory Journal*, *45*(5), 1197-1207. https://doi.org/10.1183/09031936.00007915

Basil, M. C., Katzen, J., Engler, A. E., Guo, M., Herriges, M. J., Kathiriya, J. J., Windmueller, R., Ysasi, A. B., Zacharias, W. J., & Chapman, H. A. (2020). The cellular and physiological basis for lung repair and regeneration: past, present, and future. *Cell Stem Cell*, *26*(4), 482-502.

Beeh, K. M. (2016). The Role of Bronchodilators in Preventing Exacerbations of Chronic Obstructive Pulmonary Disease. *Tuberc Respir Dis (Seoul)*, *79*(4), 241-247. https://doi.org/10.4046/trd.2016.79.4.241

Beers, M. F., & Moodley, Y. (2017). When Is an Alveolar Type 2 Cell an Alveolar Type 2 Cell? A Conundrum for Lung Stem Cell Biology and Regenerative Medicine. *American Journal of Respiratory Cell and Molecular Biology*, *57*(1), 18-27. https://doi.org/10.1165/rcmb.2016-0426PS

Beers, M. F., & Mulugeta, S. (2005). SURFACTANT PROTEIN C BIOSYNTHESIS AND ITS EMERGING ROLE IN CONFORMATIONAL LUNG DISEASE. *Annual Review of Physiology*, *67*(1), 663-696. https://doi.org/10.1146/annurev.physiol.67.040403.101937

Beers, M. F., & Mulugeta, S. (2017). The biology of the ABCA3 lipid transporter in lung health and disease. *Cell Tissue Res*, *367*(3), 481-493. https://doi.org/10.1007/s00441-016-2554-z

Bennett, E. P., Petersen, B. L., Johansen, I. E., Niu, Y., Yang, Z., Chamberlain, C. A., Met, Ö., Wandall, H. H., & Frödin, M. (2020). INDEL detection, the 'Achilles heel' of precise genome editing: a survey of methods for accurate profiling of gene editing induced indels. *Nucleic Acids Research*, *48*(21), 11958-11981. https://doi.org/10.1093/nar/gkaa975

Besser, D. (2004). Expression of Nodal, Lefty-A, and Lefty-B in Undifferentiated Human Embryonic Stem Cells Requires Activation of Smad2/3 *. *Journal of Biological Chemistry*, *279*(43), 45076-45084. https://doi.org/10.1074/jbc.M404979200

Betensley, A., Sharif, R., & Karamichos, D. (2016). A Systematic Review of the Role of Dysfunctional Wound Healing in the Pathogenesis and Treatment of Idiopathic Pulmonary Fibrosis. *J Clin Med*, *6*(1). https://doi.org/10.3390/jcm6010002

Bhaskaran, M., Kolliputi, N., Wang, Y., Gou, D., Chintagari, N. R., & Liu, L. (2007). Trans-differentiation of alveolar epithelial type II cells to type I cells involves autocrine signaling by transforming growth factor beta 1 through the Smad pathway. *J Biol Chem*, *282*(6), 3968-3976. https://doi.org/10.1074/jbc.M609060200

Bikkavilli, R. K., & Malbon, C. C. (2009). Mitogen-activated protein kinases and Wnt/beta-catenin signaling: Molecular conversations among signaling pathways. *Commun Integr Biol*, *2*(1), 46-49. https://doi.org/10.4161/cib.2.1.7503

Boergermann, J. H., Kopf, J., Yu, P. B., & Knaus, P. (2010). Dorsomorphin and LDN-193189 inhibit BMP-mediated Smad, p38 and Akt signalling in C2C12 cells. *Int J Biochem Cell Biol*, *42*(11), 1802-1807. https://doi.org/10.1016/j.biocel.2010.07.018

Bolt, R. J., van Weissenbruch, M. M., Lafeber, H. N., & Delemarre-van de Waal, H. A. (2001). Glucocorticoids and lung development in the fetus and preterm infant. *Pediatric Pulmonology*, *32*(1), 76-91. https://doi.org/https://doi.org/10.1002/ppul.1092

Bone, H. K., Nelson, A. S., Goldring, C. E., Tosh, D., & Welham, M. J. (2011). A novel chemically directed route for the generation of definitive endoderm from human embryonic stem cells based on inhibition of GSK-3. *J Cell Sci*, *124*(Pt 12), 1992-2000. https://doi.org/10.1242/jcs.081679

Bourdin, A., Burgel, P.-R., Chanez, P., Garcia, G., Perez, T., & Roche, N. (2009). Recent advances in COPD: pathophysiology, respiratory physiology and clinical aspects, including comorbidities. *European Respiratory Review*, *18*(114), 198-212. https://doi.org/10.1183/09059180.00005509

Brafman, D. A., Moya, N., Allen-Soltero, S., Fellner, T., Robinson, M., McMillen, Z. L., Gaasterland, T., & Willert, K. (2013). Analysis of SOX2-expressing cell populations derived from human pluripotent stem cells. *Stem Cell Reports*, *1*(5), 464-478. https://doi.org/10.1016/j.stemcr.2013.09.005

Brewer, G. J., Torricelli, J. R., Evege, E. K., & Price, P. J. (1993). Optimized survival of hippocampal neurons in B27-supplemented neurobasal™, a new serum-free medium combination. *Journal of Neuroscience Research*, *35*(5), 567-576. https://doi.org/https://doi.org/10.1002/jnr.490350513

Bridges, J. P., Sudha, P., Lipps, D., Wagner, A., Guo, M., Du, Y., Brown, K., Filuta, A., Kitzmiller, J., Stockman, C., Chen, X., Weirauch, M. T., Jobe, A. H., Whitsett, J. A., & Xu, Y. (2020). Glucocorticoid regulates mesenchymal cell differentiation required for perinatal lung morphogenesis and function. *Am J Physiol Lung Cell Mol Physiol*, *319*(2), L239-l255. https://doi.org/10.1152/ajplung.00459.2019

Brown, S., Teo, A., Pauklin, S., Hannan, N., Cho, C. H.-H., Lim, B., Vardy, L., Dunn, N. R., Trotter, M., Pedersen, R., & Vallier, L. (2011). Activin/Nodal Signaling Controls Divergent Transcriptional Networks in Human Embryonic Stem Cells and in Endoderm Progenitors. *Stem Cells*, *29*(8), 1176-1185. https://doi.org/https://doi.org/10.1002/stem.666

Brownfield, D. G., de Arce, A. D., Ghelfi, E., Gillich, A., Desai, T. J., & Krasnow, M. A. (2022). Alveolar cell fate selection and lifelong maintenance of AT2 cells by FGF signaling. *Nature Communications*, *13*(1), 7137. https://doi.org/10.1038/s41467-022-34059-1

Burridge, P. W., Matsa, E., Shukla, P., Lin, Z. C., Churko, J. M., Ebert, A. D., Lan, F., Diecke, S., Huber, B., Mordwinkin, N. M., Plews, J. R., Abilez, O. J., Cui, B., Gold, J. D., & Wu, J. C. (2014). Chemically defined generation of human cardiomyocytes. *Nature Methods*, *11*(8), 855-860. https://doi.org/10.1038/nmeth.2999

Burtscher, I., & Lickert, H. (2009). Foxa2 regulates polarity and epithelialization in the endoderm germ layer of the mouse embryo. *Development*, *136*(6), 1029-1038. https://doi.org/10.1242/dev.028415

Cardoso, W. V., Itoh, A., Nogawa, H., Mason, I., & Brody, J. S. (1997). FGF-1 and FGF-7 induce distinct patterns of growth and differentiation in embryonic lung epithelium. *Dev Dyn*, *208*(3), 398-405. https://doi.org/10.1002/(sici)1097-0177(199703)208:3<398::Aid-aja10>3.0.Co;2-x

Cardoso, W. V., & Lü, J. (2006). Regulation of early lung morphogenesis: questions, facts and controversies. *Development*, *133*(9), 1611-1624. https://doi.org/10.1242/dev.02310

Chakraborty, M., & Kotecha, S. (2013). Pulmonary surfactant in newborn infants and children. *Breathe*, *9*(6), 476-488. https://doi.org/10.1183/20734735.006513

Chen, F., Cao, Y., Qian, J., Shao, F., Niederreither, K., & Cardoso, W. V. (2010). A retinoic acid-dependent network in the foregut controls formation of the mouse lung primordium. *J Clin Invest*, *120*(6), 2040-2048. https://doi.org/10.1172/jci40253

Chen, F., Desai, T. J., Qian, J., Niederreither, K., Lü, J., & Cardoso, W. V. (2007). Inhibition of Tgfβ signaling by endogenous retinoic acid is essential for primary lung bud induction. *Development*, *134*(16), 2969-2979. https://doi.org/10.1242/dev.006221

Chen, G., Hou, Z., Gulbranson, D. R., & Thomson, J. A. (2010). Actin-Myosin Contractility Is Responsible for the Reduced Viability of Dissociated Human Embryonic Stem Cells. *Cell Stem Cell*, *7*(2), 240-248. https://doi.org/10.1016/j.stem.2010.06.017

Chen, Q., Shen, Y., & Zheng, J. (2021). A review of cystic fibrosis: Basic and clinical aspects. *Animal Model Exp Med*, *4*(3), 220-232. https://doi.org/10.1002/ame2.12180

Chen, Y. W., Huang, S. X., de Carvalho, A., Ho, S. H., Islam, M. N., Volpi, S., Notarangelo, L. D., Ciancanelli, M., Casanova, J. L., Bhattacharya, J., Liang, A. F., Palermo, L. M., Porotto, M., Moscona, A., & Snoeck, H. W. (2017). A three-dimensional model of human lung development and disease from pluripotent stem cells. *Nat Cell Biol*, *19*(5), 542-549. https://doi.org/10.1038/ncb3510

Cheng, X., Ying, L., Lu, L., Galvão, A. M., Mills, J. A., Lin, H. C., Kotton, D. N., Shen, S. S., Nostro, M. C., Choi, J. K., Weiss, M. J., French, D. L., & Gadue, P. (2012). Self-renewing endodermal progenitor lines generated from human pluripotent stem cells. *Cell Stem Cell*, *10*(4), 371-384. https://doi.org/10.1016/j.stem.2012.02.024

Cheong, N., Zhang, H., Madesh, M., Zhao, M., Yu, K., Dodia, C., Fisher, A. B., Savani, R. C., & Shuman, H. (2007). ABCA3 is critical for lamellar body biogenesis in vivo. *J Biol Chem*, *282*(33), 23811-23817. https://doi.org/10.1074/jbc.M703927200

Chew, J. L., Loh, Y. H., Zhang, W., Chen, X., Tam, W. L., Yeap, L. S., Li, P., Ang, Y. S., Lim, B., Robson, P., & Ng, H. H. (2005). Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol Cell Biol*, *25*(14), 6031-6046. https://doi.org/10.1128/mcb.25.14.6031-6046.2005

Choi, J., Park, J.-E., Tsagkogeorga, G., Yanagita, M., Koo, B.-K., Han, N., & Lee, J.-H. (2020). Inflammatory signals induce AT2 cell-derived damage-associated transient progenitors that mediate alveolar regeneration. *Cell Stem Cell*, *27*(3), 366-382. e367.

Chung, M.-I., Bujnis, M., Barkauskas, C. E., Kobayashi, Y., & Hogan, B. L. M. (2018). Niche-mediated BMP/SMAD signaling regulates lung alveolar stem cell proliferation and differentiation. *Development*, *145*(9). https://doi.org/10.1242/dev.163014

Collard, H. R. (2010). The Age of Idiopathic Pulmonary Fibrosis. *American Journal of Respiratory and Critical Care Medicine*, *181*(8), 771-772. https://doi.org/10.1164/rccm.201001-0049ED

Colognato, H., & Yurchenco, P. D. (2000). Form and function: The laminin family of heterotrimers. *Developmental Dynamics*, *218*(2), 213-234. https://doi.org/https://doi.org/10.1002/(SICI)1097-0177(200006)218:2<213::AID-DVDY1>3.0.CO;2-R

Cuevas-Ocaña, S., Yang, J. Y., Aushev, M., Schlossmacher, G., Bear, C. E., Hannan, N. R. F., Perkins, N. D., Rossant, J., Wong, A. P., & Gray, M. A. (2023). A Cell-Based Optimised Approach for Rapid and Efficient Gene Editing of Human Pluripotent Stem Cells. *International Journal of Molecular Sciences*, *24*(12), 10266. https://www.mdpi.com/1422-0067/24/12/10266

Cuevas Ocana, S., Serna Valverde, A. L., Reed, L., & Hannan, N. (2022). *Next generation of gene-editing technologies for respiratory research and medicine*. https://doi.org/10.1183/13993003.congress-2022.2815

D'Amour, K. A., Agulnick, A. D., Eliazer, S., Kelly, O. G., Kroon, E., & Baetge, E. E. (2005). Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nature Biotechnology*, *23*(12), 1534-1541. https://doi.org/10.1038/nbt1163

D'Amour, K. A., Bang, A. G., Eliazer, S., Kelly, O. G., Agulnick, A. D., Smart, N. G., Moorman, M. A., Kroon, E., Carpenter, M. K., & Baetge, E. E. (2006). Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nat Biotechnol*, *24*(11), 1392-1401. https://doi.org/10.1038/nbt1259

Danopoulos, S., Thornton, M. E., Grubbs, B. H., Frey, M. R., Warburton, D., Bellusci, S., & Al Alam, D. (2019). Discordant roles for FGF ligands in lung branching morphogenesis between human and mouse. *The Journal of Pathology*, *247*(2), 254-265. https://doi.org/https://doi.org/10.1002/path.5188

Davis, J. D., & Wypych, T. P. (2021). Cellular and functional heterogeneity of the airway epithelium. *Mucosal Immunology*, *14*(5), 978-990. https://doi.org/10.1038/s41385-020-00370-7

de Carvalho, A., Strikoudis, A., Liu, H. Y., Chen, Y. W., Dantas, T. J., Vallee, R. B., Correia-Pinto, J., & Snoeck, H. W. (2019). Glycogen synthase kinase 3 induces multilineage maturation of human pluripotent stem cell-derived lung progenitors in 3D culture. *Development*, *146*(2). https://doi.org/10.1242/dev.171652

de Klerk, E., & Hebrok, M. (2021). Stem Cell-Based Clinical Trials for Diabetes Mellitus. *Front Endocrinol (Lausanne)*, *12*, 631463. https://doi.org/10.3389/fendo.2021.631463

Dean, C. H., & Snelgrove, R. J. (2018). New Rules for Club Development: New Insights into Human Small Airway Epithelial Club Cell Ontogeny and Function. *Am J Respir Crit Care Med*, *198*(11), 1355-1356. https://doi.org/10.1164/rccm.201805-0925ED

Dekkers, J. F., Alieva, M., Wellens, L. M., Ariese, H. C. R., Jamieson, P. R., Vonk, A. M., Amatngalim, G. D., Hu, H., Oost, K. C., Snippert, H. J. G., Beekman, J. M., Wehrens, E. J., Visvader, J. E., Clevers, H., & Rios, A. C. (2019). High-resolution 3D imaging of fixed and cleared organoids. *Nat Protoc*, *14*(6), 1756-1771. https://doi.org/10.1038/s41596-019-0160-8

DeLaForest, A., Di Furio, F., Jing, R., Ludwig-Kubinski, A., Twaroski, K., Urick, A., Pulakanti, K., Rao, S., & Duncan, S. A. (2018). HNF4A Regulates the Formation of Hepatic Progenitor Cells from Human iPSC-Derived Endoderm by Facilitating Efficient Recruitment of RNA Pol II. *Genes (Basel)*, *10*(1). https://doi.org/10.3390/genes10010021

Desai, T. J., Brownfield, D. G., & Krasnow, M. A. (2014). Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature*, *507*(7491), 190-194. https://doi.org/10.1038/nature12930

Desai, T. J., Malpel, S., Flentke, G. R., Smith, S. M., & Cardoso, W. V. (2004). Retinoic acid selectively regulates Fgf10 expression and maintains cell identity in the prospective lung field of the developing foregut. *Developmental Biology*, *273*(2), 402-415. https://doi.org/https://doi.org/10.1016/j.ydbio.2004.04.039

Devine, J. F. (2008). Chronic obstructive pulmonary disease: an overview. *Am Health Drug Benefits*, *1*(7), 34-42.

Diekmann, U., Lenzen, S., & Naujok, O. (2015). A reliable and efficient protocol for human pluripotent stem cell differentiation into the definitive endoderm based on dispersed single cells. *Stem Cells and Development*, *24*(2), 190-204.

Diekmann, U., Wolling, H., Dettmer, R., Niwolik, I., Naujok, O., & Buettner, F. F. R. (2019). Chemically defined and xenogeneic-free differentiation of human pluripotent stem cells into definitive endoderm in 3D culture. *Scientific Reports*, *9*(1), 996. https://doi.org/10.1038/s41598-018-37650-z

Domyan, E. T., Ferretti, E., Throckmorton, K., Mishina, Y., Nicolis, S. K., & Sun, X. (2011). Signaling through BMP receptors promotes respiratory identity in the foregut via repression of Sox2. *Development*, *138*(5), 971-981. https://doi.org/10.1242/dev.053694

Duan, R. R., Hao, K., & Yang, T. (2020). Air pollution and chronic obstructive pulmonary disease. *Chronic Dis Transl Med*, *6*(4), 260-269. https://doi.org/10.1016/j.cdtm.2020.05.004

Dye, B. R., Hill, D. R., Ferguson, M. A. H., Tsai, Y.-H., Nagy, M. S., Dyal, R., Wells, J. M., Mayhew, C. N., Nattiv, R., Klein, O. D., White, E. S., Deutsch, G. H., & Spence, J. R. (2015). In vitro generation of human pluripotent stem cell derived lung organoids. *Elife*, *4*, e05098. https://doi.org/10.7554/eLife.05098

Elborn, J. S. (2016). Cystic fibrosis. *The Lancet*, *388*(10059), 2519-2531. https://doi.org/https://doi.org/10.1016/S0140-6736(16)00576-6

Engert, S., Burtscher, I., Liao, W. P., Dulev, S., Schotta, G., & Lickert, H. (2013). Wnt/β-catenin signalling regulates Sox17 expression and is essential for organizer and endoderm formation in the mouse. *Development*, *140*(15), 3128-3138. https://doi.org/10.1242/dev.088765

Evans, K. V., & Lee, J. H. (2020). Alveolar wars: The rise of in vitro models to understand human lung alveolar maintenance, regeneration, and disease. *Stem Cells Transl Med*, *9*(8), 867-881. https://doi.org/10.1002/sctm.19-0433

Fathi, A., Eisa-Beygi, S., & Baharvand, H. (2017). Signaling Molecules Governing Pluripotency and Early Lineage Commitments in Human Pluripotent Stem Cells. *Cell J*, *19*(2), 194-203. https://doi.org/10.22074/cellj.2016.3915

Feng, R., Liebe, R., & Weng, H.-L. (2023). Transcription networks in liver development and acute liver failure. *Liver Research*, *7*(1), 47-55. https://doi.org/https://doi.org/10.1016/j.livres.2022.11.010

Firth, A. L., Dargitz, C. T., Qualls, S. J., Menon, T., Wright, R., Singer, O., Gage, F. H., Khanna, A., & Verma, I. M. (2014). Generation of multiciliated cells in functional airway epithelia from human induced pluripotent stem cells. *Proceedings of the National Academy of Sciences*, *111*(17), E1723-E1730. https://doi.org/doi:10.1073/pnas.1403470111

Fisher, J., Pulakanti, K., Rao, S., & Duncan, S. (2017). GATA6 is essential for endoderm formation from human pluripotent stem cells. *Biology Open*, *6*(7), 1084-1095.

Francis, G. L. (2010). Albumin and mammalian cell culture: implications for biotechnology applications. *Cytotechnology*, *62*(1), 1-16. https://doi.org/10.1007/s10616-010-9263-3

Frank, D. B., Peng, T., Zepp, J. A., Snitow, M., Vincent, T. L., Penkala, I. J., Cui, Z., Herriges, M. J., Morley, M. P., Zhou, S., Lu, M. M., & Morrisey, E. E. (2016). Emergence of a Wave of Wnt Signaling that Regulates Lung Alveologenesis by Controlling Epithelial Self-Renewal and Differentiation. *Cell Rep*, *17*(9), 2312-2325. https://doi.org/10.1016/j.celrep.2016.11.001

Frank, D. B., Penkala, I. J., Zepp, J. A., Sivakumar, A., Linares-Saldana, R., Zacharias, W. J., Stolz, K. G., Pankin, J., Lu, M., Wang, Q., Babu, A., Li, L., Zhou, S., Morley, M. P., Jain, R., & Morrisey, E. E. (2019). Early lineage specification defines alveolar epithelial ontogeny in the murine lung. *Proceedings of the National Academy of Sciences*, *116*(10), 4362-4371. https://doi.org/doi:10.1073/pnas.1813952116

Frum, T., Hsu, P. P., Hein, R. F. C., Conchola, A. S., Zhang, C. J., Utter, O. R., Anand, A., Zhang, Y., Clark, S. G., Glass, I., Sexton, J. Z., & Spence, J. R. (2023). Opposing roles for TGFβ- and BMP-signaling during nascent alveolar differentiation in the developing human lung. *npj Regenerative Medicine*, *8*(1), 48. https://doi.org/10.1038/s41536-023-00325-z

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, *551*(7681), 464-471. https://doi.org/10.1038/nature24644

Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. (2017). *Lancet Respir Med*, *5*(9), 691-706. https://doi.org/10.1016/s2213-2600(17)30293-x

Gokey, J. J., Snowball, J., Sridharan, A., Sudha, P., Kitzmiller, J. A., Xu, Y., & Whitsett, J. A. (2021). YAP regulates alveolar epithelial cell differentiation and AGER via NFIB/KLF5/NKX2-1. *iScience*, *24*(9). https://doi.org/10.1016/j.isci.2021.102967

Gonzales, L. W., Guttentag, S. H., Wade, K. C., Postle, A. D., & Ballard, P. L. (2002). Differentiation of human pulmonary type II cells in vitro by glucocorticoid plus cAMP. *Am J Physiol Lung Cell Mol Physiol*, *283*(5), L940-951. https://doi.org/10.1152/ajplung.00127.2002

Goodwin, J., Laslett, A. L., & Rugg-Gunn, P. J. (2020). The application of cell surface markers to demarcate distinct human pluripotent states. *Exp Cell Res*, *387*(1), 111749. https://doi.org/10.1016/j.yexcr.2019.111749

Goss, A. M., Tian, Y., Tsukiyama, T., Cohen, E. D., Zhou, D., Lu, M. M., Yamaguchi, T. P., & Morrisey, E. E. (2009). Wnt2/2b and β-Catenin Signaling Are Necessary and Sufficient to Specify Lung Progenitors in the Foregut. *Developmental Cell*, *17*(2), 290-298. https://doi.org/https://doi.org/10.1016/j.devcel.2009.06.005

Gotoh, S., Ito, I., Nagasaki, T., Yamamoto, Y., Konishi, S., Korogi, Y., Matsumoto, H., Muro, S., Hirai, T., Funato, M., Mae, S., Toyoda, T., Sato-Otsubo, A., Ogawa, S., Osafune, K., & Mishima, M. (2014). Generation of alveolar epithelial spheroids via isolated progenitor cells from human pluripotent stem cells. *Stem Cell Reports*, *3*(3), 394-403. https://doi.org/10.1016/j.stemcr.2014.07.005

Gouon-Evans, V., Boussemart, L., Gadue, P., Nierhoff, D., Koehler, C. I., Kubo, A., Shafritz, D. A., & Keller, G. (2006). BMP-4 is required for hepatic specification of mouse embryonic stem cell–derived definitive endoderm. *Nature Biotechnology*, *24*(11), 1402-1411. https://doi.org/10.1038/nbt1258

Graffmann, N., Scherer, B., & Adjaye, J. (2022). In vitro differentiation of pluripotent stem cells into hepatocyte like cells – Basic principles and current progress. *Stem Cell Research*, *61*, 102763. https://doi.org/https://doi.org/10.1016/j.scr.2022.102763

Green, M. D., Chen, A., Nostro, M. C., d'Souza, S. L., Schaniel, C., Lemischka, I. R., Gouon-Evans, V., Keller, G., & Snoeck, H. W. (2011). Generation of anterior foregut endoderm from human embryonic and induced pluripotent stem cells. *Nat Biotechnol*, *29*(3), 267-272. https://doi.org/10.1038/nbt.1788

Gu, X., Zhou, F., & Mu, J. (2021). Recent Advances in Maturation of Pluripotent Stem Cell-Derived Cardiomyocytes Promoted by Mechanical Stretch. *Med Sci Monit*, *27*, e931063. https://doi.org/10.12659/msm.931063

Günzburg, W. H., & Salmons, B. (2000). Xenotransplantation: is the risk of viral infection as great as we thought? *Molecular Medicine Today*, *6*(5), 199-208. https://doi.org/10.1016/S1357-4310(00)01708-1

Hamill, K. J., Kligys, K., Hopkinson, S. B., & Jones, J. C. R. (2009). Laminin deposition in the extracellular matrix: a complex picture emerges. *Journal of Cell Science*, *122*(24), 4409-4417. https://doi.org/10.1242/jcs.041095

Han, S., & Mallampalli, R. K. (2015). The Role of Surfactant in Lung Disease and Host Defense against Pulmonary Infections. *Ann Am Thorac Soc*, *12*(5), 765-774. https://doi.org/10.1513/AnnalsATS.201411-507FR

Hannan, N. R., Fordham, R. P., Syed, Y. A., Moignard, V., Berry, A., Bautista, R., Hanley, N. A., Jensen, K. B., & Vallier, L. (2013). Generation of multipotent foregut stem cells from human pluripotent stem cells. *Stem Cell Reports*, *1*(4), 293-306. https://doi.org/10.1016/j.stemcr.2013.09.003

Hannan, N. R., Sampaziotis, F., Segeritz, C. P., Hanley, N. A., & Vallier, L. (2015). Generation of Distal Airway Epithelium from Multipotent Human Foregut Stem Cells. *Stem Cells Dev*, *24*(14), 1680-1690. https://doi.org/10.1089/scd.2014.0512

Hannan, N. R., Segeritz, C. P., Touboul, T., & Vallier, L. (2013). Production of hepatocyte-like cells from human pluripotent stem cells. *Nat Protoc*, *8*(2), 430-437. https://doi.org/10.1038/nprot.2012.153

Hanssens, L. S., Duchateau, J., & Casimir, G. J. (2021). CFTR Protein: Not Just a Chloride Channel? *Cells*, *10*(11). https://doi.org/10.3390/cells10112844

Hawkins, F., Kramer, P., Jacob, A., Driver, I., Thomas, D. C., McCauley, K. B., Skvir, N., Crane, A. M., Kurmann, A. A., Hollenberg, A. N., Nguyen, S., Wong, B. G., Khalil, A. S., Huang, S. X., Guttentag, S., Rock, J. R., Shannon, J. M., Davis, B. R., & Kotton, D. N. (2017). Prospective isolation of NKX2-1-expressing human lung progenitors derived from pluripotent stem cells. *J Clin Invest*, *127*(6), 2277-2294. https://doi.org/10.1172/jci89950

Hawkins, F. J., Suzuki, S., Beermann, M. L., Barillà, C., Wang, R., Villacorta-Martin, C., Berical, A., Jean, J. C., Le Suer, J., Matte, T., Simone-Roach, C., Tang, Y., Schlaeger, T. M., Crane, A. M., Matthias, N., Huang, S. X. L., Randell, S. H., Wu, J., Spence, J. R., . . . Kotton, D. N. (2021). Derivation of Airway Basal Stem Cells from Human Pluripotent Stem Cells. *Cell Stem Cell*, *28*(1), 79-95.e78. https://doi.org/10.1016/j.stem.2020.09.017

He, P., Lim, K., Sun, D., Pett, J. P., Jeng, Q., Polanski, K., Dong, Z., Bolt, L., Richardson, L., & Mamanova, L. (2022). A human fetal lung cell atlas uncovers proximal-distal gradients of differentiation and key regulators of epithelial fates. *Cell*, *185*(25), 4841-4860. e4825.

Hein, R. F. C., Conchola, A. S., Fine, A. S., Xiao, Z., Frum, T., Brastrom, L. K., Akinwale, M. A., Childs, C. J., Tsai, Y. H., Holloway, E. M., Huang, S., Mahoney, J., Heemskerk, I., & Spence, J. R. (2022). Stable iPSC-derived NKX2-1+ lung bud tip progenitor organoids give rise to airway and alveolar cell types. *Development*, *149*(20). https://doi.org/10.1242/dev.200693

Heldin, C. H., & Moustakas, A. (2016). Signaling Receptors for TGF-β Family Members. *Cold Spring Harb Perspect Biol*, *8*(8). https://doi.org/10.1101/cshperspect.a022053

Hoffman, E. T., Uriarte, J. J., Uhl, F. E., Eckstrom, K., Tanneberger, A. E., Becker, C., Moulin, C., Asarian, L., Ikonomou, L., Kotton, D. N., & Weiss, D. J. (2023). Human alveolar hydrogels promote morphological and transcriptional differentiation in iPSC-derived alveolar type 2 epithelial cells. *Scientific Reports*, *13*(1), 12057. https://doi.org/10.1038/s41598-023-37685-x

Hogan, B. L. M. (2020, 2020//). The Alveolar Stem Cell Niche of the Mammalian Lung. Molecular Mechanism of Congenital Heart Disease and Pulmonary Hypertension, Singapore.

Holtzinger, A., Streeter, P. R., Sarangi, F., Hillborn, S., Niapour, M., Ogawa, S., & Keller, G. (2015). New markers for tracking endoderm induction and hepatocyte differentiation from human pluripotent stem cells. *Development*, *142*(24), 4253-4265. https://doi.org/10.1242/dev.121020

Horbelt, D., Boergermann, J. H., Chaikuad, A., Alfano, I., Williams, E., Lukonin, I., Timmel, T., Bullock, A. N., & Knaus, P. (2015). Small molecules dorsomorphin and LDN-193189 inhibit myostatin/GDF8 signaling and promote functional myoblast differentiation. *J Biol Chem*, *290*(6), 3390-3404. https://doi.org/10.1074/jbc.M114.604397

Hsiao, C., Lampe, M., Nillasithanukroh, S., Han, W., Lian, X., & Palecek, S. P. (2016). Human pluripotent stem cell culture density modulates YAP signaling. *Biotechnol J*, *11*(5), 662-675. https://doi.org/10.1002/biot.201500374

Hu, D., Sun, X., Magpusao, A., Fedorov, Y., Thompson, M., Wang, B., Lundberg, K., Adams, D. J., & Qi, X. (2021). Small-molecule suppression of calpastatin degradation reduces neuropathology in models of Huntington's disease. *Nature Communications*, *12*(1), 5305. https://doi.org/10.1038/s41467-021-25651-y

Huang, R., Zhang, C., Zheng, Y., Zhang, W., Huang, H., Qiu, M., Li, J., & Li, F. (2023). ISL1 regulates lung branching morphogenesis via Shh signaling pathway. *J Biol Chem*, *299*(8), 105034. https://doi.org/10.1016/j.jbc.2023.105034

Huang, S. K., & Horowitz, J. C. (2014). Outstaying their Welcome: The Persistent Myofibroblast in IPF. *Austin J Pulm Respir Med*, *1*(1), 3.

Huang, S. X., Islam, M. N., O'Neill, J., Hu, Z., Yang, Y. G., Chen, Y. W., Mumau, M., Green, M. D., Vunjak-Novakovic, G., Bhattacharya, J., & Snoeck, H. W. (2014). Efficient generation of lung and airway epithelial cells from human pluripotent stem cells. *Nat Biotechnol*, *32*(1), 84-91. https://doi.org/10.1038/nbt.2754

Huang, S. X. L., Green, M. D., de Carvalho, A. T., Mumau, M., Chen, Y.-W., D'Souza, S. L., & Snoeck, H.-W. (2015). The in vitro generation of lung and airway progenitor cells from human pluripotent stem cells. *Nature Protocols*, *10*(3), 413-425. https://doi.org/10.1038/nprot.2015.023

Hughes, C. S., Postovit, L. M., & Lajoie, G. A. (2010). Matrigel: A complex protein mixture required for optimal growth of cell culture. *PROTEOMICS*, *10*(9), 1886-1890. https://doi.org/https://doi.org/10.1002/pmic.200900758

Huizing, M., Anikster, Y., & Gahl, W. A. (2000). Hermansky-Pudlak syndrome and related disorders of organelle formation. *Traffic*, *1*(11), 823-835. https://doi.org/10.1034/j.1600-0854.2000.011103.x

Hurley, K., Ding, J., Villacorta-Martin, C., Herriges, M. J., Jacob, A., Vedaie, M., Alysandratos, K. D., Sun, Y. L., Lin, C., Werder, R. B., Huang, J., Wilson, A. A., Mithal, A., Mostoslavsky, G., Oglesby, I., Caballero, I. S., Guttentag, S. H., Ahangari, F., Kaminski, N., . . . Kotton, D. N. (2020). Reconstructed Single-Cell Fate Trajectories Define Lineage Plasticity Windows during Differentiation of Human PSC-Derived Distal Lung Progenitors. *Cell Stem Cell*, *26*(4), 593-608.e598. https://doi.org/10.1016/j.stem.2019.12.009

Hussain, M., Xu, C., Lu, M., Wu, X., Tang, L., & Wu, X. (2017). Wnt/β-catenin signaling links embryonic lung development and asthmatic airway remodeling. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1863*(12), 3226-3242. https://doi.org/https://doi.org/10.1016/j.bbadis.2017.08.031

Ikeo, S., Yamamoto, Y., Ikeda, K., Sone, N., Korogi, Y., Tomiyama, L., Matsumoto, H., Hirai, T., Hagiwara, M., & Gotoh, S. (2021). Core-shell hydrogel microfiber-expanded pluripotent stem cell-derived lung progenitors applicable to lung reconstruction in vivo. *Biomaterials*, *276*, 121031. https://doi.org/https://doi.org/10.1016/j.biomaterials.2021.121031

Irvin, C. G., & Bates, J. H. T. (2003). Measuring the lung function in the mouse: the challenge of size. *Respiratory Research*, *4*(1), 1. https://doi.org/10.1186/rr199

Jacob, A., Morley, M., Hawkins, F., McCauley, K. B., Jean, J. C., Heins, H., Na, C. L., Weaver, T. E., Vedaie, M., Hurley, K., Hinds, A., Russo, S. J., Kook, S., Zacharias, W., Ochs, M., Traber, K., Quinton, L. J., Crane, A., Davis, B. R., . . . Kotton, D. N. (2017). Differentiation of Human Pluripotent Stem Cells into Functional Lung Alveolar Epithelial Cells. *Cell Stem Cell*, *21*(4), 472-488.e410. https://doi.org/10.1016/j.stem.2017.08.014

Jacob, A., Vedaie, M., Roberts, D. A., Thomas, D. C., Villacorta-Martin, C., Alysandratos, K. D., Hawkins, F., & Kotton, D. N. (2019). Derivation of self-renewing lung alveolar epithelial type II cells from human pluripotent stem cells. *Nat Protoc*, *14*(12), 3303-3332. https://doi.org/10.1038/s41596-019-0220-0

Jang, K.-S., Kim, Y.-G., Adhya, M., Park, H.-M., & Kim, B.-G. (2013). The sweets standing at the borderline between allo- and xenotransplantation. *Xenotransplantation*, *20*(4), 199-208. https://doi.org/https://doi.org/10.1111/xen.12030

Jenkins, R. G., Moore, B. B., Chambers, R. C., Eickelberg, O., Königshoff, M., Kolb, M., Laurent, G. J., Nanthakumar, C. B., Olman, M. A., Pardo, A., Selman, M., Sheppard, D., Sime, P. J., Tager, A. M., Tatler, A. L., Thannickal, V. J., & White, E. S. (2017). An Official American Thoracic Society Workshop Report: Use of Animal Models for the Preclinical Assessment of Potential Therapies for Pulmonary Fibrosis. *American Journal of Respiratory Cell and Molecular Biology*, *56*(5), 667-679. https://doi.org/10.1165/rcmb.2017-0096ST

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, *337*(6096), 816-821. https://doi.org/doi:10.1126/science.1225829

Jones-Freeman, B., & Starkey, M. R. (2020). Bronchioalveolar stem cells in lung repair, regeneration and disease. *The Journal of Pathology*, *252*(3), 219-226. https://doi.org/https://doi.org/10.1002/path.5527

Joung, J. K., & Sander, J. D. (2013). TALENs: a widely applicable technology for targeted genome editing. *Nature Reviews Molecular Cell Biology*, *14*(1), 49-55. https://doi.org/10.1038/nrm3486

Kanai-Azuma, M., Kanai, Y., Gad, J. M., Tajima, Y., Taya, C., Kurohmaru, M., Sanai, Y., Yonekawa, H., Yazaki, K., Tam, P. P. L., & Hayashi, Y. (2002). Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development*, *129*(10), 2367-2379. https://doi.org/10.1242/dev.129.10.2367

Kaneko-Kawano, T., Takasu, F., Naoki, H., Sakumura, Y., Ishii, S., Ueba, T., Eiyama, A., Okada, A., Kawano, Y., & Suzuki, K. (2012). Dynamic regulation of myosin light chain phosphorylation by Rho-kinase. *PLoS One*, *7*(6), e39269. https://doi.org/10.1371/journal.pone.0039269

Kang, J., Nathan, E., Xu, S.-M., Tzahor, E., & Black, B. L. (2009). Isl1 is a direct transcriptional target of Forkhead transcription factors in second heart field-derived mesoderm. *Developmental Biology*, *334*(2), 513-522. https://doi.org/https://doi.org/10.1016/j.ydbio.2009.06.041

Karra, N., Swindle, E. J., & Morgan, H. (2019). Drug delivery for traditional and emerging airway models. *Organs-on-a-Chip*, *1*, 100002. https://doi.org/https://doi.org/10.1016/j.ooc.2020.100002

Kartikasari, A. E. R., Zhou, J. X., Kanji, M. S., Chan, D. N., Sinha, A., Grapin-Botton, A., Magnuson, M. A., Lowry, W. E., & Bhushan, A. (2013). The histone demethylase Jmjd3 sequentially associates with the transcription factors Tbx3 and Eomes to drive endoderm differentiation. *The EMBO Journal*, *32*(10), 1393-1408. https://doi.org/https://doi.org/10.1038/emboj.2013.78

Kathiriya, J. J., Wang, C., Zhou, M., Brumwell, A., Cassandras, M., Le Saux, C. J., Cohen, M., Alysandratos, K.-D., Wang, B., Wolters, P., Matthay, M., Kotton, D. N., Chapman, H. A., & Peng, T. (2022). Human alveolar type 2 epithelium transdifferentiates into metaplastic KRT5+ basal cells. *Nature Cell Biology*, *24*(1), 10-23. https://doi.org/10.1038/s41556-021-00809-4

Katsura, H., Sontake, V., Tata, A., Kobayashi, Y., Edwards, C. E., Heaton, B. E., Konkimalla, A., Asakura, T., Mikami, Y., Fritch, E. J., Lee, P. J., Heaton, N. S., Boucher, R. C., Randell, S. H., Baric, R. S., & Tata, P. R. (2020). Human Lung Stem Cell-Based Alveolospheres Provide Insights into SARS-CoV-2-Mediated Interferon Responses and Pneumocyte Dysfunction. *Cell Stem Cell*, *27*(6), 890-904.e898. https://doi.org/10.1016/j.stem.2020.10.005

Kearns, N. A., Genga, R. M. J., Ziller, M., Kapinas, K., Peters, H., Brehm, M. A., Meissner, A., & Maehr, R. (2013). Generation of organized anterior foregut epithelia from pluripotent stem cells using small molecules. *Stem Cell Research*, *11*(3), 1003-1012. https://doi.org/https://doi.org/10.1016/j.scr.2013.06.007

Kelly, G. M., & Drysdale, T. A. (2015). Retinoic Acid and the Development of the Endoderm. *Journal of Developmental Biology*, *3*(2), 25-56. https://www.mdpi.com/2221-3759/3/2/25

Kempf, H., Olmer, R., Haase, A., Franke, A., Bolesani, E., Schwanke, K., Robles-Diaz, D., Coffee, M., Göhring, G., Dräger, G., Pötz, O., Joos, T., Martinez-Hackert, E., Haverich, A., Buettner, F. F. R., Martin, U., & Zweigerdt, R. (2016). Bulk cell density and Wnt/TGFbeta signalling regulate mesendodermal patterning of human pluripotent stem cells. *Nature Communications*, *7*(1), 13602. https://doi.org/10.1038/ncomms13602

KHUBCHANDANI, K. R., & SNYDER, J. M. (2001). Surfactant protein A (SP-A): the alveolus and beyond. *The FASEB Journal*, *15*(1), 59-69. https://doi.org/https://doi.org/10.1096/fj.00-0318rev

Kido, T., Koui, Y., Suzuki, K., Kobayashi, A., Miura, Y., Chern, E. Y., Tanaka, M., & Miyajima, A. (2015). CPM Is a Useful Cell Surface Marker to Isolate Expandable Bi-Potential Liver Progenitor Cells Derived from Human iPS Cells. *Stem Cell Reports*, *5*(4), 508-515. https://doi.org/10.1016/j.stemcr.2015.08.008

Kilpinen, H., Goncalves, A., Leha, A., Afzal, V., Alasoo, K., Ashford, S., Bala, S., Bensaddek, D., Casale, F. P., Culley, O. J., Danecek, P., Faulconbridge, A., Harrison, P. W., Kathuria, A., McCarthy, D., McCarthy, S. A., Meleckyte, R., Memari, Y., Moens, N., . . . Gaffney, D. J. (2017). Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature*, *546*(7658), 370-375. https://doi.org/10.1038/nature22403

Kim, E., Jiang, M., Huang, H., Zhang, Y., Tjota, N., Gao, X., Robert, J., Gilmore, N., Gan, L., & Que, J. (2019). Isl1 Regulation of Nkx2.1 in the Early Foregut Epithelium Is Required for Trachea-Esophageal Separation and Lung Lobation. *Dev Cell*, *51*(6), 675-683.e674. https://doi.org/10.1016/j.devcel.2019.11.002

Klaffky, E., Williams, R., Yao, C.-C., Ziober, B., Kramer, R., & Sutherland, A. (2001). Trophoblast-specific expression and function of the integrin α7 subunit in the peri-implantation mouse embryo. *Developmental Biology*, *239*(1), 161-175.

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*(7603), 420-424. https://doi.org/10.1038/nature17946

Konagaya, S., & Iwata, H. (2019). Chemically defined conditions for long-term maintenance of pancreatic progenitors derived from human induced pluripotent stem cells. *Scientific Reports*, *9*(1), 640. https://doi.org/10.1038/s41598-018-36606-7

Konishi, S., Gotoh, S., Tateishi, K., Yamamoto, Y., Korogi, Y., Nagasaki, T., Matsumoto, H., Muro, S., Hirai, T., Ito, I., Tsukita, S., & Mishima, M. (2016). Directed Induction of Functional Multi-ciliated Cells in Proximal Airway Epithelial Spheroids from Human Pluripotent Stem Cells. *Stem Cell Reports*, *6*(1), 18-25. https://doi.org/10.1016/j.stemcr.2015.11.010

Korfei, M., Ruppert, C., Mahavadi, P., Henneke, I., Markart, P., Koch, M., Lang, G., Fink, L., Bohle, R. M., Seeger, W., Weaver, T. E., & Guenther, A. (2008). Epithelial endoplasmic reticulum stress and apoptosis in sporadic idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*, *178*(8), 838-846. https://doi.org/10.1164/rccm.200802-313OC

Korogi, Y., Gotoh, S., Ikeo, S., Yamamoto, Y., Sone, N., Tamai, K., Konishi, S., Nagasaki, T., Matsumoto, H., Ito, I., Chen-Yoshikawa, T. F., Date, H., Hagiwara, M., Asaka, I., Hotta, A., Mishima, M., & Hirai, T. (2019). In Vitro Disease Modeling of Hermansky-Pudlak Syndrome Type 2 Using Human Induced Pluripotent Stem Cell-Derived Alveolar Organoids. *Stem Cell Reports*, *12*(3), 431-440. https://doi.org/10.1016/j.stemcr.2019.01.014

Koutsourakis, M., Keijzer, R., Visser, P., Post, M., Tibboel, D., & Grosveld, F. (2001). Branching and differentiation defects in pulmonary epithelium with elevated Gata6 expression. *Mech Dev*, *105*(1-2), 105-114. https://doi.org/10.1016/s0925-4773(01)00386-0

Kozlowski, M. T., Crook, C. J., & Ku, H. T. (2021). Towards organoid culture without Matrigel. *Communications Biology*, *4*(1), 1387. https://doi.org/10.1038/s42003-021-02910-8

Kraus, M. R. C., & Grapin-Botton, A. (2012). Patterning and shaping the endoderm in vivo and in culture. *Current Opinion in Genetics & Development*, *22*(4), 347-353. https://doi.org/https://doi.org/10.1016/j.gde.2012.05.002

Krauss, E., Gehrken, G., Drakopanagiotakis, F., Tello, S., Dartsch, R. C., Maurer, O., Windhorst, A., von der Beck, D., Griese, M., Seeger, W., & Guenther, A. (2019). Clinical characteristics of patients with familial idiopathic pulmonary fibrosis (f-IPF). *BMC Pulm Med*, *19*(1), 130. https://doi.org/10.1186/s12890-019-0895-6

Kropski, J. A., Blackwell, T. S., & Loyd, J. E. (2015). The genetic basis of idiopathic pulmonary fibrosis. *Eur Respir J*, *45*(6), 1717-1727. https://doi.org/10.1183/09031936.00163814

Kugler, M. C., Joyner, A. L., Loomis, C. A., & Munger, J. S. (2015). Sonic hedgehog signaling in the lung. From development to disease. *Am J Respir Cell Mol Biol*, *52*(1), 1-13. https://doi.org/10.1165/rcmb.2014-0132TR

Kurath-Koller, S., Resch, B., Kraschl, R., Windpassinger, C., & Eber, E. (2015). Surfactant Protein B Deficiency Caused by Homozygous C248X Mutation-A Case Report and Review of the Literature. *AJP Rep*, *5*(1), e53-59. https://doi.org/10.1055/s-0035-1545668

Kwon, G. S., Viotti, M., & Hadjantonakis, A. K. (2008). The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages. *Dev Cell*, *15*(4), 509-520. https://doi.org/10.1016/j.devcel.2008.07.017

Laco, F., Woo, T. L., Zhong, Q., Szmyd, R., Ting, S., Khan, F. J., Chai, C. L. L., Reuveny, S., Chen, A., & Oh, S. (2018). Unraveling the Inconsistencies of Cardiac Differentiation Efficiency Induced by the GSK3β Inhibitor CHIR99021 in Human Pluripotent Stem Cells. *Stem Cell Reports*, *10*(6), 1851-1866. https://doi.org/https://doi.org/10.1016/j.stemcr.2018.03.023

Laresgoiti, U., Nikolić, M. Z., Rao, C., Brady, J. L., Richardson, R. V., Batchen, E. J., Chapman, K. E., & Rawlins, E. L. (2016). Lung epithelial tip progenitors integrate glucocorticoid- and STAT3-mediated signals to control progeny fate. *Development*, *143*(20), 3686-3699. https://doi.org/10.1242/dev.134023

Lau, H. H., Amirruddin, N. S., Loo, L. S. W., Chan, J. W., Iich, E., Krishnan, V. G., Hoon, S., & Teo, A. K. K. (2023). FGFR-mediated ERK1/2 signaling contributes to mesendoderm and definitive endoderm formation in vitro. *iScience*, *26*(8), 107265. https://doi.org/10.1016/j.isci.2023.107265

Lawson, K. A., Meneses, J. J., & Pedersen, R. A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development*, *113*(3), 891-911. https://doi.org/10.1242/dev.113.3.891

Lawson, W. E., Crossno, P. F., Polosukhin, V. V., Roldan, J., Cheng, D.-S., Lane, K. B., Blackwell, T. R., Xu, C., Markin, C., Ware, L. B., Miller, G. G., Loyd, J. E., & Blackwell, T. S. (2008). Endoplasmic reticulum stress in alveolar epithelial cells is prominent in IPF: association with altered surfactant protein processing and herpesvirus infection. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, *294*(6), L1119-L1126. https://doi.org/10.1152/ajplung.00382.2007

Lazzaro, D., Price, M., Felice, M. D., & Lauro, R. D. (1991). The transcription factor TTF-1 is expressed at the onset of thyroid and lung morphogenesis and in restricted regions of the foetal brain. *Development*, *113*(4), 1093-1104. https://doi.org/10.1242/dev.113.4.1093

Lee, C. C., Ding, X., Zhao, T., Wu, L., Perkins, S., Du, H., & Yan, C. (2019). Transthyretin Stimulates Tumor Growth through Regulation of Tumor, Immune, and Endothelial Cells. *J Immunol*, *202*(3), 991-1002. https://doi.org/10.4049/jimmunol.1800736

Lee, J.-H., & Rawlins, E. L. (2018). Developmental mechanisms and adult stem cells for therapeutic lung regeneration. *Developmental Biology*, *433*(2), 166-176. https://doi.org/https://doi.org/10.1016/j.ydbio.2017.09.016

Lei, L., Traore, S., Romano Ibarra, G. S., Karp, P. H., Rehman, T., Meyerholz, D. K., Zabner, J., Stoltz, D. A., Sinn, P. L., Welsh, M. J., McCray, P. B., Jr., & Thornell, I. M. (2023). CFTR-rich ionocytes mediate chloride absorption across airway epithelia. *J Clin Invest*, *133*(20). https://doi.org/10.1172/jci171268

Leibel, S. L., Winquist, A., Tseu, I., Wang, J., Luo, D., Shojaie, S., Nathan, N., Snyder, E., & Post, M. (2019). Reversal of Surfactant Protein B Deficiency in Patient Specific Human Induced Pluripotent Stem Cell Derived Lung Organoids by Gene Therapy. *Sci Rep*, *9*(1), 13450. https://doi.org/10.1038/s41598-019-49696-8

Lemaigre, F. P. (2009). Mechanisms of Liver Development: Concepts for Understanding Liver Disorders and Design of Novel Therapies. *Gastroenterology*, *137*(1), 62-79. https://doi.org/https://doi.org/10.1053/j.gastro.2009.03.035

Li, L., Ma, M., Duan, T., & Sui, X. (2022). The critical roles and therapeutic implications of tuft cells in cancer [Review]. *Frontiers in Pharmacology*, *13*. https://doi.org/10.3389/fphar.2022.1047188

Li, M., & Belmonte, J. C. I. (2017). Ground rules of the pluripotency gene regulatory network. *Nature Reviews Genetics*, *18*(3), 180-191. https://doi.org/10.1038/nrg.2016.156

Lian, X., Bao, X., Zilberter, M., Westman, M., Fisahn, A., Hsiao, C., Hazeltine, L. B., Dunn, K. K., Kamp, T. J., & Palecek, S. P. (2015). Chemically defined, albumin-free human cardiomyocyte generation. *Nature Methods*, *12*(7), 595-596. https://doi.org/10.1038/nmeth.3448

Liao, C.-M., Mukherjee, S., Tiyaboonchai, A., Maguire, J. A., Cardenas-Diaz, F. L., French, D. L., & Gadue, P. (2018). GATA6 suppression enhances lung specification from human pluripotent stem cells. *The Journal of Clinical Investigation*, *128*(7), 2944-2950. https://doi.org/10.1172/JCI96539

Lim, K., Donovan, A. P. A., Tang, W., Sun, D., He, P., Pett, J. P., Teichmann, S. A., Marioni, J. C., Meyer, K. B., Brand, A. H., & Rawlins, E. L. (2023). Organoid modeling of human fetal lung alveolar development reveals mechanisms of cell fate patterning and neonatal respiratory disease. *Cell Stem Cell*, *30*(1), 20-37.e29. https://doi.org/10.1016/j.stem.2022.11.013

Little, D. R., Gerner-Mauro, K. N., Flodby, P., Crandall, E. D., Borok, Z., Akiyama, H., Kimura, S., Ostrin, E. J., & Chen, J. (2019). Transcriptional control of lung alveolar type 1 cell development and maintenance by NK homeobox 2-1. *Proceedings of the National Academy of Sciences*, *116*(41), 20545-20555. https://doi.org/doi:10.1073/pnas.1906663116

Little, D. R., Lynch, A. M., Yan, Y., Akiyama, H., Kimura, S., & Chen, J. (2021). Differential chromatin binding of the lung lineage transcription factor NKX2-1 resolves

opposing murine alveolar cell fates in vivo. *Nature Communications*, *12*(1), 2509. https://doi.org/10.1038/s41467-021-22817-6

Liu, L. P., & Zheng, Y. W. (2019). Predicting differentiation potential of human pluripotent stem cells: Possibilities and challenges. *World J Stem Cells*, *11*(7), 375-382. https://doi.org/10.4252/wjsc.v11.i7.375

Liu, Q., Liu, K., Cui, G., Huang, X., Yao, S., Guo, W., Qin, Z., Li, Y., Yang, R., Pu, W., Zhang, L., He, L., Zhao, H., Yu, W., Tang, M., Tian, X., Cai, D., Nie, Y., Hu, S., . . . Zhou, B. (2019). Lung regeneration by multipotent stem cells residing at the bronchioalveolar-duct junction. *Nature Genetics*, *51*(4), 728-738. https://doi.org/10.1038/s41588-019-0346-6

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, *25*(4), 402-408. https://doi.org/10.1006/meth.2001.1262

Loebel, C., Weiner, A. I., Katzen, J. B., Morley, M. P., Bala, V., Cardenas-Diaz, F. L., Davidson, M. D., Shiraishi, K., Basil, M. C., Ochs, M., Beers, M. F., Morrisey, E. E., Vaughan, A. E., & Burdick, J. A. (2021). Microstructured hydrogels to guide self-assembly and function of lung alveolospheres. *bioRxiv*, 2021.2008.2030.457534. https://doi.org/10.1101/2021.08.30.457534

Loh, K. M., Ang, L. T., Zhang, J., Kumar, V., Ang, J., Auyeong, J. Q., Lee, K. L., Choo, S. H., Lim, C. Y., & Nichane, M. (2014). Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. *Cell Stem Cell*, *14*(2), 237-252.

Longmire, T. A., Ikonomou, L., Hawkins, F., Christodoulou, C., Cao, Y., Jean, J. C., Kwok, L. W., Mou, H., Rajagopal, J., Shen, S. S., Dowton, A. A., Serra, M., Weiss, D. J., Green, M. D., Snoeck, H. W., Ramirez, M. I., & Kotton, D. N. (2012). Efficient derivation of purified lung and thyroid progenitors from embryonic stem cells. *Cell Stem Cell*, *10*(4), 398-411. https://doi.org/10.1016/j.stem.2012.01.019

Ma, Y., Ma, M., Sun, J., Li, W., Li, Y., Guo, X., & Zhang, H. (2019). CHIR-99021 regulates mitochondrial remodelling via β-catenin signalling and miRNA expression during endodermal differentiation. *Journal of Cell Science*, *132*(15). https://doi.org/10.1242/jcs.229948

Mahavadi, P., Korfei, M., Henneke, I., Liebisch, G., Schmitz, G., Gochuico, B. R., Markart, P., Bellusci, S., Seeger, W., Ruppert, C., & Guenther, A. (2010). Epithelial stress and apoptosis underlie Hermansky-Pudlak syndrome-associated interstitial pneumonia. *Am J Respir Crit Care Med*, *182*(2), 207-219. https://doi.org/10.1164/rccm.200909-1414OC

Maher, T. M., Wells, A. U., & Laurent, G. J. (2007). Idiopathic pulmonary fibrosis: multiple causes and multiple mechanisms? *Eur Respir J*, *30*(5), 835-839. https://doi.org/10.1183/09031936.00069307

Mammoto, A., & Mammoto, T. (2019). Vascular Niche in Lung Alveolar Development, Homeostasis, and Regeneration [Review]. *Frontiers in Bioengineering and Biotechnology*, *7*. https://doi.org/10.3389/fbioe.2019.00318

Mandai, M., Watanabe, A., Kurimoto, Y., Hirami, Y., Morinaga, C., Daimon, T., Fujihara, M., Akimaru, H., Sakai, N., Shibata, Y., Terada, M., Nomiya, Y., Tanishima, S., Nakamura, M., Kamao, H., Sugita, S., Onishi, A., Ito, T., Fujita, K., . . . Takahashi, M. (2017). Autologous Induced Stem-Cell–Derived Retinal Cells for Macular Degeneration. *New England Journal of Medicine*, *376*(11), 1038-1046. https://doi.org/10.1056/NEJMoa1608368

McBennett, K. A., Davis, P. B., & Konstan, M. W. (2022). Increasing life expectancy in cystic fibrosis: Advances and challenges. *Pediatr Pulmonol*, *57 Suppl 1*(Suppl 1), S5-s12. https://doi.org/10.1002/ppul.25733

McCauley, K. B., Alysandratos, K. D., Jacob, A., Hawkins, F., Caballero, I. S., Vedaie, M., Yang, W., Slovik, K. J., Morley, M., Carraro, G., Kook, S., Guttentag, S. H., Stripp, B. R., Morrisey, E. E., & Kotton, D. N. (2018). Single-Cell Transcriptomic Profiling of Pluripotent Stem Cell-Derived SCGB3A2+ Airway Epithelium. *Stem Cell Reports*, *10*(5), 1579-1595. https://doi.org/10.1016/j.stemcr.2018.03.013

McLean, A. B., D'Amour, K. A., Jones, K. L., Krishnamoorthy, M., Kulik, M. J., Reynolds, D. M., Sheppard, A. M., Liu, H., Xu, Y., Baetge, E. E., & Dalton, S. (2007). Activin A Efficiently Specifies Definitive Endoderm from Human Embryonic Stem Cells Only When Phosphatidylinositol 3-Kinase Signaling Is Suppressed. *Stem Cells*, *25*(1), 29-38. https://doi.org/10.1634/stemcells.2006-0219

Mei, Q., Liu, Z., Zuo, H., Yang, Z., & Qu, J. (2021). Idiopathic Pulmonary Fibrosis: An Update on Pathogenesis. *Front Pharmacol*, *12*, 797292. https://doi.org/10.3389/fphar.2021.797292

Mendelson, C. R., & Boggaram, V. (1989). Regulation of pulmonary surfactant protein synthesis in fetal lung: A major role of glucocorticoids and cyclic AMP. *Trends in Endocrinology & Metabolism*, *1*(1), 20-25. https://doi.org/https://doi.org/10.1016/1043-2760(89)90025-8

Mendjan, S., Mascetti, V. L., Ortmann, D., Ortiz, M., Karjosukarso, D. W., Ng, Y., Moreau, T., & Pedersen, R. A. (2014). NANOG and CDX2 pattern distinct subtypes of human mesoderm during exit from pluripotency. *Cell Stem Cell*, *15*(3), 310-325.

Meng, X., Cui, G., & Peng, G. (2023). Lung development and regeneration: newly defined cell types and progenitor status. *Cell Regen*, *12*(1), 5. https://doi.org/10.1186/s13619-022-00149-0

Michalski, J. E., & Schwartz, D. A. (2020). Genetic Risk Factors for Idiopathic Pulmonary Fibrosis: Insights into Immunopathogenesis. *J Inflamm Res*, *13*, 1305-1318. https://doi.org/10.2147/jir.S280958

Miller, A. J., Dye, B. R., Ferrer-Torres, D., Hill, D. R., Overeem, A. W., Shea, L. D., & Spence, J. R. (2019). Generation of lung organoids from human pluripotent stem cells in vitro. *Nat Protoc*, *14*(2), 518-540. https://doi.org/10.1038/s41596-018-0104-8

Miller, A. J., Hill, D. R., Nagy, M. S., Aoki, Y., Dye, B. R., Chin, A. M., Huang, S., Zhu, F., White, E. S., Lama, V., & Spence, J. R. (2018). In Vitro Induction and In Vivo Engraftment of Lung Bud Tip Progenitor Cells Derived from Human Pluripotent Stem Cells. *Stem Cell Reports*, *10*(1), 101-119. https://doi.org/10.1016/j.stemcr.2017.11.012

Miller, A. J., Yu, Q., Czerwinski, M., Tsai, Y. H., Conway, R. F., Wu, A., Holloway, E. M., Walker, T., Glass, I. A., Treutlein, B., Camp, J. G., & Spence, J. R. (2020). In Vitro and In Vivo Development of the Human Airway at Single-Cell Resolution. *Dev Cell*, *53*(1), 117-128.e116. https://doi.org/10.1016/j.devcel.2020.01.033

Miner, J. H., Li, C., Mudd, J. L., Go, G., & Sutherland, A. E. (2004). Compositional and structural requirements for laminin and basement membranes during mouse embryo implantation and gastrulation. *Development*, *131*(10), 2247-2256. https://doi.org/10.1242/dev.01112

Minoo, P., Su, G., Drum, H., Bringas, P., & Kimura, S. (1999). Defects in tracheoesophageal and lung morphogenesis inNkx2. 1 (−/−) mouse embryos. *Developmental Biology*, *209*(1), 60-71.

Mithal, A., Capilla, A., Heinze, D., Berical, A., Villacorta-Martin, C., Vedaie, M., Jacob, A., Abo, K., Szymaniak, A., Peasley, M., Stuffer, A., Mahoney, J., Kotton, D. N., Hawkins, F., & Mostoslavsky, G. (2020). Generation of mesenchyme free intestinal organoids from human induced pluripotent stem cells. *Nature Communications*, *11*(1), 215. https://doi.org/10.1038/s41467-019-13916-6

Molkentin, J. D., Lin, Q., Duncan, S. A., & Olson, E. N. (1997). Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes & development*, *11*(8), 1061-1072.

Montoro, D. T., Haber, A. L., Biton, M., Vinarsky, V., Lin, B., Birket, S. E., Yuan, F., Chen, S., Leung, H. M., Villoria, J., Rogel, N., Burgin, G., Tsankov, A. M., Waghray, A., Slyper, M., Waldman, J., Nguyen, L., Dionne, D., Rozenblatt-Rosen, O., . . . Rajagopal, J. (2018). A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature*, *560*(7718), 319-324. https://doi.org/10.1038/s41586-018-0393-7

Moore, B., Lawson, W. E., Oury, T. D., Sisson, T. H., Raghavendran, K., & Hogaboam, C. M. (2013). Animal models of fibrotic lung disease. *Am J Respir Cell Mol Biol*, *49*(2), 167-179. https://doi.org/10.1165/rcmb.2013-0094TR

Morizane, A. (2023). Cell therapy for Parkinson's disease with induced pluripotent stem cells. *Inflamm Regen*, *43*(1), 16. https://doi.org/10.1186/s41232-023-00269-3

Morrisey, E. E., & Hogan, B. L. (2010). Preparing for the first breath: genetic and cellular mechanisms in lung development. *Developmental Cell*, *18*(1), 8-23.

Morrisey, E. E., Tang, Z., Sigrist, K., Lu, M. M., Jiang, F., Ip, H. S., & Parmacek, M. S. (1998). GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes & development*, *12*(22), 3579-3590.

Mosqueira, D., Mannhardt, I., Bhagwan, J. R., Lis-Slimak, K., Katili, P., Scott, E., Hassan, M., Prondzynski, M., Harmer, S. C., Tinker, A., Smith, J. G. W., Carrier, L., Williams, P. M., Gaffney, D., Eschenhagen, T., Hansen, A., & Denning, C. (2018). CRISPR/Cas9 editing in human pluripotent stem cell-cardiomyocytes highlights arrhythmias, hypocontractility, and energy depletion as potential therapeutic targets for hypertrophic cardiomyopathy. *Eur Heart J*, *39*(43), 3879-3892. https://doi.org/10.1093/eurheartj/ehy249

Mou, H., Vinarsky, V., Tata, P. R., Brazauskas, K., Choi, S. H., Crooke, A. K., Zhang, B., Solomon, G. M., Turner, B., Bihler, H., Harrington, J., Lapey, A., Channick, C., Keyes, C., Freund, A., Artandi, S., Mense, M., Rowe, S., Engelhardt, J. F., . . . Rajagopal, J. (2016). Dual SMAD Signaling Inhibition Enables Long-Term Expansion of Diverse Epithelial Basal Cells. *Cell Stem Cell*, *19*(2), 217-231. https://doi.org/10.1016/j.stem.2016.05.012

Mou, H., Zhao, R., Sherwood, R., Ahfeldt, T., Lapey, A., Wain, J., Sicilian, L., Izvolsky, K., Musunuru, K., Cowan, C., & Rajagopal, J. (2012). Generation of multipotent lung and airway progenitors from mouse ESCs and patient-specific cystic fibrosis iPSCs. *Cell Stem Cell*, *10*(4), 385-397. https://doi.org/10.1016/j.stem.2012.01.018

Mulugeta, S., Maguire, J. A., Newitt, J. L., Russo, S. J., Kotorashvili, A., & Beers, M. F. (2007). Misfolded BRICHOS SP-C mutant proteins induce apoptosis via caspase-4- and cytochrome c-related mechanisms. *American Journal of*

Physiology-Lung Cellular and Molecular Physiology, *293*(3), L720-L729. https://doi.org/10.1152/ajplung.00025.2007

Mulugeta, S., Nguyen, V., Russo, S. J., Muniswamy, M., & Beers, M. F. (2005). A surfactant protein C precursor protein BRICHOS domain mutation causes endoplasmic reticulum stress, proteasome dysfunction, and caspase 3 activation. *Am J Respir Cell Mol Biol*, *32*(6), 521-530. https://doi.org/10.1165/rcmb.2005-0009OC

Mulugeta, S., Nureki, S., & Beers, M. F. (2015). Lost after translation: insights from pulmonary surfactant for understanding the role of alveolar epithelial dysfunction and cellular quality control in fibrotic lung disease. *Am J Physiol Lung Cell Mol Physiol*, *309*(6), L507-525. https://doi.org/10.1152/ajplung.00139.2015

Murphy, W. L., McDevitt, T. C., & Engler, A. J. (2014). Materials as stem cell regulators. *Nat Mater*, *13*(6), 547-557. https://doi.org/10.1038/nmat3937

Nabhan, A. N., Brownfield, D. G., Harbury, P. B., Krasnow, M. A., & Desai, T. J. (2018). Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science*, *359*(6380), 1118-1123. https://doi.org/10.1126/science.aam6603

Nakamura, Y., & Suda, T. (2015). Idiopathic Pulmonary Fibrosis: Diagnosis and Clinical Manifestations. *Clin Med Insights Circ Respir Pulm Med*, *9*(Suppl 1), 163-171. https://doi.org/10.4137/CCRPM.S39897

Narayanan, M., Owers-Bradley, J., Beardsmore, C. S., Mada, M., Ball, I., Garipov, R., Panesar, K. S., Kuehni, C. E., Spycher, B. D., Williams, S. E., & Silverman, M. (2012). Alveolarization continues during childhood and adolescence: new evidence from helium-3 magnetic resonance. *Am J Respir Crit Care Med*, *185*(2), 186-191. https://doi.org/10.1164/rccm.201107-1348OC

Naujok, O., Diekmann, U., & Lenzen, S. (2014). The Generation of Definitive Endoderm from Human Embryonic Stem Cells is Initially Independent from Activin A but Requires Canonical Wnt-Signaling. *Stem Cell Reviews and Reports*, *10*(4), 480-493. https://doi.org/10.1007/s12015-014-9509-0

Nguyen, N. M., Kelley, D. G., Schlueter, J. A., Meyer, M. J., Senior, R. M., & Miner, J. H. (2005). Epithelial laminin α5 is necessary for distal epithelial cell maturation, VEGF production, and alveolization in the developing murine lung. *Developmental Biology*, *282*(1), 111-125. https://doi.org/https://doi.org/10.1016/j.ydbio.2005.02.031

Nguyen, N. M., & Senior, R. M. (2006). Laminin isoforms and lung development: All isoforms are not equal. *Developmental Biology*, *294*(2), 271-279. https://doi.org/https://doi.org/10.1016/j.ydbio.2006.03.032

Nikolić, M. Z., Caritg, O., Jeng, Q., Johnson, J.-A., Sun, D., Howell, K. J., Brady, J. L., Laresgoiti, U., Allen, G., Butler, R., Zilbauer, M., Giangreco, A., & Rawlins, E. L. (2017). Human embryonic lung epithelial tips are multipotent progenitors that can be expanded in vitro as long-term self-renewing organoids. *Elife*, *6*, e26575. https://doi.org/10.7554/eLife.26575

Nikolić, M. Z., Sun, D., & Rawlins, E. L. (2018). Human lung development: recent progress and new challenges. *Development*, *145*(16). https://doi.org/10.1242/dev.163485

Nishiuchi, R., Takagi, J., Hayashi, M., Ido, H., Yagi, Y., Sanzen, N., Tsuji, T., Yamada, M., & Sekiguchi, K. (2006). Ligand-binding specificities of laminin-binding integrins: a comprehensive survey of laminin-integrin interactions using recombinant

alpha3beta1, alpha6beta1, alpha7beta1 and alpha6beta4 integrins. *Matrix Biol*, *25*(3), 189-197. https://doi.org/10.1016/j.matbio.2005.12.001

Nogee, L. M., Dunbar, A. E., Wert, S. E., Askin, F., Hamvas, A., & Whitsett, J. A. (2001). A Mutation in the Surfactant Protein C Gene Associated with Familial Interstitial Lung Disease. *New England Journal of Medicine*, *344*(8), 573-579. https://doi.org/10.1056/nejm200102223440805

NOGEE, L. M., WERT, S. E., PROFFIT, S. A., HULL, W. M., & WHITSETT, J. A. (2000). Allelic Heterogeneity in Hereditary Surfactant Protein B (SP-B) Deficiency. *American Journal of Respiratory and Critical Care Medicine*, *161*(3), 973-981. https://doi.org/10.1164/ajrccm.161.3.9903153

Noguchi, M., Furukawa, K. T., & Morimoto, M. (2020). Pulmonary neuroendocrine cells: physiology, tissue homeostasis and disease. *Dis Model Mech*, *13*(12). https://doi.org/10.1242/dmm.046920

Nowotschin, S., Hadjantonakis, A.-K., & Campbell, K. (2019). The endoderm: a divergent cell lineage with many commonalities. *Development*, *146*(11). https://doi.org/10.1242/dev.150920

Olajuyin, A. M., Zhang, X., & Ji, H.-L. (2019). Alveolar type 2 progenitor cells for lung injury repair. *Cell Death Discovery*, *5*(1), 63. https://doi.org/10.1038/s41420-019-0147-9

Ono, S., Tanaka, T., Ishida, M., Kinoshita, A., Fukuoka, J., Takaki, M., Sakamoto, N., Ishimatsu, Y., Kohno, S., Hayashi, T., Senba, M., Yasunami, M., Kubo, Y., Yoshida, L. M., Kubo, H., Ariyoshi, K., Yoshiura, K., & Morimoto, K. (2011). Surfactant protein C G100S mutation causes familial pulmonary fibrosis in Japanese kindred. *European Respiratory Journal*, *38*(4), 861-869. https://doi.org/10.1183/09031936.00143610

Ori, C., Ansari, M., Angelidis, I., Theis, F. J., Schiller, H. B., & Drukker, M. (2021). Single cell trajectory analysis of human pluripotent stem cells differentiating towards lung and hepatocyte progenitors. *bioRxiv*, 2021.2002. 2023.432413.

Park, K.-S., Whitsett, J. A., Di Palma, T., Hong, J.-H., Yaffe, M. B., & Zannini, M. (2004). TAZ Interacts with TTF-1 and Regulates Expression of Surfactant Protein-C *. *Journal of Biological Chemistry*, *279*(17), 17384-17390. https://doi.org/10.1074/jbc.M312569200

Parker, D., & Prince, A. (2011). Innate immunity in the respiratory epithelium. *Am J Respir Cell Mol Biol*, *45*(2), 189-201. https://doi.org/10.1165/rcmb.2011-0011RT

Parviz, F., Matullo, C., Garrison, W. D., Savatski, L., Adamson, J. W., Ning, G., Kaestner, K. H., Rossi, J. M., Zaret, K. S., & Duncan, S. A. (2003). Hepatocyte nuclear factor 4alpha controls the development of a hepatic epithelium and liver morphogenesis. *Nat Genet*, *34*(3), 292-296. https://doi.org/10.1038/ng1175

Pastva, A. M., Wright, J. R., & Williams, K. L. (2007). Immunomodulatory roles of surfactant proteins A and D: implications in lung disease. *Proc Am Thorac Soc*, *4*(3), 252-257. https://doi.org/10.1513/pats.200701-018AW

Perrimon, N., Pitsouli, C., & Shilo, B. Z. (2012). Signaling mechanisms controlling cell fate and embryonic patterning. *Cold Spring Harb Perspect Biol*, *4*(8), a005975. https://doi.org/10.1101/cshperspect.a005975

Phillips, B. W., Hentze, H., Rust, W. L., Chen, Q. P., Chipperfield, H., Tan, E. K., Abraham, S., Sadasivam, A., Soong, P. L., Wang, S. T., Lim, R., Sun, W., Colman, A., & Dunn, N. R. (2007). Directed differentiation of human embryonic stem cells into the

pancreatic endocrine lineage. *Stem Cells Dev*, *16*(4), 561-578. https://doi.org/10.1089/scd.2007.0029

Pierce, R. A., Griffin, G. L., Miner, J. H., & Senior, R. M. (2000). Expression Patterns of Laminin α 1 and α 5 in Human Lung during Development. *American Journal of Respiratory Cell and Molecular Biology*, *23*(6), 742-747.

Piersma, B., Bank, R. A., & Boersema, M. (2015). Signaling in Fibrosis: TGF-β, WNT, and YAP/TAZ Converge. *Front Med (Lausanne)*, *2*, 59. https://doi.org/10.3389/fmed.2015.00059

Plasschaert, L. W., Žilionis, R., Choo-Wing, R., Savova, V., Knehr, J., Roma, G., Klein, A. M., & Jaffe, A. B. (2018). A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature*, *560*(7718), 377-381. https://doi.org/10.1038/s41586-018-0394-6

Rackley, C. R., & Stripp, B. R. (2012). Building and maintaining the epithelium of the lung. *The Journal of Clinical Investigation*, *122*(8), 2724-2730. https://doi.org/10.1172/JCI60519

Raghu, G., Remy-Jardin, M., Myers, J. L., Richeldi, L., Ryerson, C. J., Lederer, D. J., Behr, J., Cottin, V., Danoff, S. K., Morell, F., Flaherty, K. R., Wells, A., Martinez, F. J., Azuma, A., Bice, T. J., Bouros, D., Brown, K. K., Collard, H. R., Duggal, A., . . . Wilson, K. C. (2018). Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *American Journal of Respiratory and Critical Care Medicine*, *198*(5), e44-e68. https://doi.org/10.1164/rccm.201807-1255ST

Raghu, G., Rochwerg, B., Zhang, Y., Garcia, C. A., Azuma, A., Behr, J., Brozek, J. L., Collard, H. R., Cunningham, W., Homma, S., Johkoh, T., Martinez, F. J., Myers, J., Protzko, S. L., Richeldi, L., Rind, D., Selman, M., Theodore, A., Wells, A. U., . . . Latin American Thoracic, A. (2015). An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. *Am J Respir Crit Care Med*, *192*(2), e3-19. https://doi.org/10.1164/rccm.201506-1063ST

Rankin, S. A., Han, L., McCracken, K. W., Kenny, A. P., Anglin, C. T., Grigg, E. A., Crawford, C. M., Wells, J. M., Shannon, J. M., & Zorn, A. M. (2016). A Retinoic Acid-Hedgehog Cascade Coordinates Mesoderm-Inducing Signals and Endoderm Competence during Lung Specification. *Cell Rep*, *16*(1), 66-78. https://doi.org/10.1016/j.celrep.2016.05.060

Rankin, S. A., McCracken, K. W., Luedeke, D. M., Han, L., Wells, J. M., Shannon, J. M., & Zorn, A. M. (2018). Timing is everything: Reiterative Wnt, BMP and RA signaling regulate developmental competence during endoderm organogenesis. *Dev Biol*, *434*(1), 121-132. https://doi.org/10.1016/j.ydbio.2017.11.018

Rawlins, E. L., Clark, C. P., Xue, Y., & Hogan, B. L. M. (2009). The Id2+ distal tip lung epithelium contains individual multipotent embryonic progenitor cells. *Development*, *136*(22), 3741-3745. https://doi.org/10.1242/dev.037317

Refai, E., Dekki, N., Yang, S.-N., Imreh, G., Cabrera, O., Yu, L., Yang, G., Norgren, S., Rössner, S. M., Inverardi, L., Ricordi, C., Olivecrona, G., Andersson, M., Jörnvall, H., Berggren, P.-O., & Juntti-Berggren, L. (2005). Transthyretin constitutes a functional component in pancreatic β-cell stimulus-secretion coupling. *Proceedings of the National Academy of Sciences*, *102*(47), 17020-17025. https://doi.org/doi:10.1073/pnas.0503219102

Reid, L., Meyrick, B., Antony, V. B., Chang, L. Y., Crapo, J. D., & Reynolds, H. Y. (2005). The mysterious pulmonary brush cell: a cell in search of a function. *Am J Respir Crit Care Med*, *172*(1), 136-139. https://doi.org/10.1164/rccm.200502-203WS

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47-e47. https://doi.org/10.1093/nar/gkv007

Rodin, S., Antonsson, L., Niaudet, C., Simonson, O. E., Salmela, E., Hansson, E. M., Domogatskaya, A., Xiao, Z., Damdimopoulou, P., Sheikhi, M., Inzunza, J., Nilsson, A.-S., Baker, D., Kuiper, R., Sun, Y., Blennow, E., Nordenskjöld, M., Grinnemo, K.-H., Kere, J., . . . Tryggvason, K. (2014). Clonal culturing of human embryonic stem cells on laminin-521/E-cadherin matrix in defined and xeno-free environment. *Nature Communications*, *5*(1), 3195. https://doi.org/10.1038/ncomms4195

Rodin, S., Domogatskaya, A., Ström, S., Hansson, E. M., Chien, K. R., Inzunza, J., Hovatta, O., & Tryggvason, K. (2010). Long-term self-renewal of human pluripotent stem cells on human recombinant laminin-511. *Nature Biotechnology*, *28*(6), 611-615. https://doi.org/10.1038/nbt.1620

Rodrigues Toste de Carvalho, A. L., Liu, H.-Y., Chen, Y.-W., Porotto, M., Moscona, A., & Snoeck, H.-W. (2021). The in vitro multilineage differentiation and maturation of lung and airway cells from human pluripotent stem cell–derived lung progenitors in 3D. *Nature Protocols*, *16*(4), 1802-1829. https://doi.org/10.1038/s41596-020-00476-z

Romito, A., & Cobellis, G. (2016). Pluripotent Stem Cells: Current Understanding and Future Directions. *Stem Cells Int*, *2016*, 9451492. https://doi.org/10.1155/2016/9451492

Rovina, N., Koutsoukou, A., & Koulouris, N. G. (2013). Inflammation and immune response in COPD: where do we stand? *Mediators Inflamm*, *2013*, 413735. https://doi.org/10.1155/2013/413735

Sahabian, A., Sgodda, M., Naujok, O., Dettmer, R., Dahlmann, J., Manstein, F., Cantz, T., Zweigerdt, R., Martin, U., & Olmer, R. (2019). Chemically-Defined, Xeno-Free, Scalable Production of hPSC-Derived Definitive Endoderm Aggregates with Multi-Lineage Differentiation Potential. *Cells*, *8*(12). https://doi.org/10.3390/cells8121571

Saint-Criq, V., & Gray, M. A. (2017). Role of CFTR in epithelial physiology. *Cell Mol Life Sci*, *74*(1), 93-115. https://doi.org/10.1007/s00018-016-2391-y

Salahudeen, A. A., Choi, S. S., Rustagi, A., Zhu, J., van Unen, V., de la, O. S., Flynn, R. A., Margalef-Català, M., Santos, A. J. M., Ju, J., Batish, A., Usui, T., Zheng, G. X. Y., Edwards, C. E., Wagar, L. E., Luca, V., Anchang, B., Nagendran, M., Nguyen, K., . . . Kuo, C. J. (2020). Progenitor identification and SARS-CoV-2 infection in human distal lung organoids. *Nature*, *588*(7839), 670-675. https://doi.org/10.1038/s41586-020-3014-1

Salerno, T., Peca, D., Menchini, L., Schiavino, A., Boldrini, R., Esposito, F., Danhaive, O., & Cutrera, R. (2016). Surfactant Protein C-associated interstitial lung disease; three different phenotypes of the same SFTPC mutation. *Ital J Pediatr*, *42*, 23. https://doi.org/10.1186/s13052-016-0235-x

Sapey, E., & Stockley, R. A. (2006). COPD exacerbations . 2: aetiology. *Thorax*, *61*(3), 250-258. https://doi.org/10.1136/thx.2005.041822

Scheibner, K., Schirge, S., Burtscher, I., Büttner, M., Sterr, M., Yang, D., Böttcher, A., Ansarullah, Irmler, M., Beckers, J., Cernilogar, F. M., Schotta, G., Theis, F. J., & Lickert, H. (2021). Epithelial cell plasticity drives endoderm formation during gastrulation. *Nature Cell Biology*, *23*(7), 692-703. https://doi.org/10.1038/s41556-021-00694-x

Schittny, J. C. (2017). Development of the lung. *Cell Tissue Res*, *367*(3), 427-444. https://doi.org/10.1007/s00441-016-2545-0

Scotet, V., L'Hostis, C., & Férec, C. (2020). The Changing Epidemiology of Cystic Fibrosis: Incidence, Survival and Impact of the CFTR Gene Discovery. *Genes (Basel)*, *11*(6). https://doi.org/10.3390/genes11060589

Sekine, K., Takebe, T., Suzuki, Y., Kamiya, A., Nakauchi, H., & Taniguchi, H. (2012). Highly Efficient Generation of Definitive Endoderm Lineage from Human Induced Pluripotent Stem Cells. *Transplantation Proceedings*, *44*(4), 1127-1129. https://doi.org/https://doi.org/10.1016/j.transproceed.2012.03.001

Serls, A. E., Doherty, S., Parvatiyar, P., Wells, J. M., & Deutsch, G. H. (2005). Different thresholds of fibroblast growth factors pattern the ventral foregut into liver and lung. *Development*, *132*(1), 35-47. https://doi.org/10.1242/dev.01570

Serra, M., Alysandratos, K. D., Hawkins, F., McCauley, K. B., Jacob, A., Choi, J., Caballero, I. S., Vedaie, M., Kurmann, A. A., Ikonomou, L., Hollenberg, A. N., Shannon, J. M., & Kotton, D. N. (2017). Pluripotent stem cell differentiation reveals distinct developmental pathways regulating lung- versus thyroid-lineage specification. *Development*, *144*(21), 3879-3893. https://doi.org/10.1242/dev.150193

Setting standards for stem cells. (2023). *Nature Methods*, *20*(9), 1267-1267. https://doi.org/10.1038/s41592-023-02016-5

Sever, N., Miličić, G., Bodnar, N. O., Wu, X., & Rapoport, T. A. (2021). Mechanism of Lamellar Body Formation by Lung Surfactant Protein B. *Mol Cell*, *81*(1), 49-66.e48. https://doi.org/10.1016/j.molcel.2020.10.042

Shaw-White, J. R., Bruno, M. D., & Whitsett, J. A. (1999). GATA-6 Activates Transcription of Thyroid Transcription Factor-1 *. *Journal of Biological Chemistry*, *274*(5), 2658-2664. https://doi.org/10.1074/jbc.274.5.2658

Shiraishi, K., Nakajima, T., Shichino, S., Deshimaru, S., Matsushima, K., & Ueha, S. (2019). In vitro expansion of endogenous human alveolar epithelial type II cells in fibroblast-free spheroid culture. *Biochemical and Biophysical Research Communications*, *515*(4), 579-585. https://doi.org/https://doi.org/10.1016/j.bbrc.2019.05.187

Sikkema, L., Ramírez-Suástegui, C., Strobl, D. C., Gillett, T. E., Zappia, L., Madissoon, E., Markov, N. S., Zaragosi, L.-E., Ji, Y., Ansari, M., Arguel, M.-J., Apperloo, L., Banchero, M., Bécavin, C., Berg, M., Chichelnitskiy, E., Chung, M.-i., Collin, A., Gay, A. C. A., . . . Lung Biological Network, C. (2023). An integrated cell atlas of the lung in health and disease. *Nature Medicine*, *29*(6), 1563-1577. https://doi.org/10.1038/s41591-023-02327-2

Siller, R., Naumovska, E., Mathapati, S., Lycke, M., Greenhough, S., & Sullivan, G. J. (2016). Development of a rapid screen for the endodermal differentiation potential of human pluripotent stem cell lines. *Sci Rep*, *6*, 37178. https://doi.org/10.1038/srep37178

Singh, A. M., Reynolds, D., Cliff, T., Ohtsuka, S., Mattheyses, A. L., Sun, Y., Menendez, L., Kulik, M., & Dalton, S. (2012). Signaling network crosstalk in human pluripotent cells: a Smad2/3-regulated switch that controls the balance

between self-renewal and differentiation. *Cell Stem Cell*, *10*(3), 312-326. https://doi.org/10.1016/j.stem.2012.01.014

Singh, V. K., Kalsan, M., Kumar, N., Saini, A., & Chandra, R. (2015). Induced pluripotent stem cells: applications in regenerative medicine, disease modeling, and drug discovery. *Front Cell Dev Biol*, *3*, 2. https://doi.org/10.3389/fcell.2015.00002

Snell, N., Strachan, D., Hubbard, R., Gibson, J., Maher, T., & Jarrold, I. (2016). P272 Epidemiology of idiopathic pulmonary fibrosis in the uk: findings from the british lung foundation's 'respiratory health of the nation' project. *Thorax*, *71*(Suppl 3), A236-A236. https://doi.org/10.1136/thoraxjnl-2016-209333.415

Soldner, F., Laganière, J., Cheng, Albert W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, Lawrence I., Myers, Richard H., Lindquist, S., Zhang, L., Guschin, D., Fong, Lauren K., Vu, B. J., Meng, X., Urnov, Fyodor D., Rebar, Edward J., Gregory, Philip D., Zhang, H. S., & Jaenisch, R. (2011). Generation of Isogenic Pluripotent Stem Cells Differing Exclusively at Two Early Onset Parkinson Point Mutations. *Cell*, *146*(2), 318-331. https://doi.org/10.1016/j.cell.2011.06.019

Stahlman, M. T., Gray, M. P., Falconieri, M. W., Whitsett, J. A., & Weaver, T. E. (2000). Lamellar body formation in normal and surfactant protein B-deficient fetal mice. *Lab Invest*, *80*(3), 395-403. https://doi.org/10.1038/labinvest.3780044

Steele, M. P., Speer, M. C., Loyd, J. E., Brown, K. K., Herron, A., Slifer, S. H., Burch, L. H., Wahidi, M. M., Phillips, J. A., 3rd, Sporn, T. A., McAdams, H. P., Schwarz, M. I., & Schwartz, D. A. (2005). Clinical and pathologic features of familial interstitial pneumonia. *Am J Respir Crit Care Med*, *172*(9), 1146-1152. https://doi.org/10.1164/rccm.200408-1104OC

Strikoudis, A., Cieslak, A., Loffredo, L., Chen, Y. W., Patel, N., Saqi, A., Lederer, D. J., & Snoeck, H. W. (2019). Modeling of Fibrotic Lung Disease Using 3D Organoids Derived from Human Pluripotent Stem Cells. *Cell Rep*, *27*(12), 3709-3723 e3705. https://doi.org/10.1016/j.celrep.2019.05.077

Stuart, W. D., Fink-Baldauf, I. M., Tomoshige, K., Guo, M., & Maeda, Y. (2021). CRISPRi-mediated functional analysis of NKX2-1-binding sites in the lung. *Communications Biology*, *4*(1), 568. https://doi.org/10.1038/s42003-021-02083-4

Sui, L., Mfopou, J. K., Geens, M., Sermon, K., & Bouwens, L. (2012). FGF signaling via MAPK is required early and improves Activin A-induced definitive endoderm formation from human embryonic stem cells. *Biochemical and Biophysical Research Communications*, *426*(3), 380-385. https://doi.org/https://doi.org/10.1016/j.bbrc.2012.08.098

Sun, X., Perl, A. K., Li, R., Bell, S. M., Sajti, E., Kalinichenko, V. V., Kalin, T. V., Misra, R. S., Deshmukh, H., Clair, G., Kyle, J., Crotty Alexander, L. E., Masso-Silva, J. A., Kitzmiller, J. A., Wikenheiser-Brokamp, K. A., Deutsch, G., Guo, M., Du, Y., Morley, M. P., . . . Morrisey, E. E. (2022). A census of the lung: CellCards from LungMAP. *Dev Cell*, *57*(1), 112-145.e112. https://doi.org/10.1016/j.devcel.2021.11.007

Sun, Y. L., Hurley, K., Villacorta-Martin, C., Huang, J., Hinds, A., Gopalan, K., Caballero, I. S., Russo, S. J., Kitzmiller, J. A., Whitsett, J. A., Beers, M. F., & Kotton, D. N. (2021). Heterogeneity in Human Induced Pluripotent Stem Cell-derived Alveolar Epithelial Type II Cells Revealed with ABCA3/SFTPC Reporters. *Am J Respir Cell Mol Biol*, *65*(4), 442-460. https://doi.org/10.1165/rcmb.2020-0259OC

Sundarakrishnan, A., Chen, Y., Black, L. D., Aldridge, B. B., & Kaplan, D. L. (2018). Engineered cell and tissue models of pulmonary fibrosis. *Advanced Drug Delivery Reviews*, *129*, 78-94. https://doi.org/https://doi.org/10.1016/j.addr.2017.12.013

Swarr, D. T., & Morrisey, E. E. (2015). Lung endoderm morphogenesis: gasping for form and function. *Annu Rev Cell Dev Biol*, *31*, 553-573. https://doi.org/10.1146/annurev-cellbio-100814-125249

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, *131*(5), 861-872. https://doi.org/10.1016/j.cell.2007.11.019

Talbot, N. C., & Caperna, T. J. (2015). Proteome array identification of bioactive soluble proteins/peptides in Matrigel: relevance to stem cell responses. *Cytotechnology*, *67*(5), 873-883. https://doi.org/10.1007/s10616-014-9727-y

Tam, P. P. L., & Behringer, R. R. (1997). Mouse gastrulation: the formation of a mammalian body plan. *Mechanisms of Development*, *68*(1), 3-25. https://doi.org/https://doi.org/10.1016/S0925-4773(97)00123-8

Tam, P. P. L., Williams, E. A., & Chan, W. Y. (1993). Gastrulation in the mouse embryo: Ultrastructural and molecular aspects of germ layer morphogenesis. *Microscopy Research and Technique*, *26*(4), 301-328. https://doi.org/https://doi.org/10.1002/jemt.1070260405

Tashiro, J., Rubio, G. A., Limper, A. H., Williams, K., Elliot, S. J., Ninou, I., Aidinis, V., Tzouvelekis, A., & Glassberg, M. K. (2017). Exploring Animal Models That Resemble Idiopathic Pulmonary Fibrosis. *Front Med (Lausanne)*, *4*, 118. https://doi.org/10.3389/fmed.2017.00118

Teo, A. K., Arnold, S. J., Trotter, M. W., Brown, S., Ang, L. T., Chng, Z., Robertson, E. J., Dunn, N. R., & Vallier, L. (2011). Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev*, *25*(3), 238-250. https://doi.org/10.1101/gad.607311

Teo, A. K., Valdez, I. A., Dirice, E., & Kulkarni, R. N. (2014). Comparable generation of activin-induced definitive endoderm via additive Wnt or BMP signaling in absence of serum. *Stem Cell Reports*, *3*(1), 5-14. https://doi.org/10.1016/j.stemcr.2014.05.007

Teo, A. K. K., Ali, Y., Wong, K. Y., Chipperfield, H., Sadasivam, A., Poobalan, Y., Tan, E. K., Wang, S. T., Abraham, S., Tsuneyoshi, N., Stanton, L. W., & Dunn, N. R. (2012). Activin and BMP4 Synergistically Promote Formation of Definitive Endoderm in Human Embryonic Stem Cells. *Stem Cells*, *30*(4), 631-642. https://doi.org/10.1002/stem.1022

ter Steege, E. J., & Bakker, E. R. M. (2021). The role of R-spondin proteins in cancer biology. *Oncogene*, *40*(47), 6469-6478. https://doi.org/10.1038/s41388-021-02059-y

Thomas, A. Q., Lane, K., Phillips, J., 3rd, Prince, M., Markin, C., Speer, M., Schwartz, D. A., Gaddipati, R., Marney, A., Johnson, J., Roberts, R., Haines, J., Stahlman, M., & Loyd, J. E. (2002). Heterozygosity for a surfactant protein C gene mutation associated with usual interstitial pneumonitis and cellular nonspecific interstitial pneumonitis in one kindred. *Am J Respir Crit Care Med*, *165*(9), 1322-1328. https://doi.org/10.1164/rccm.200112-123OC

Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., & Jones, J. M. (1998). Embryonic Stem Cell Lines Derived from

Human Blastocysts. *Science*, *282*(5391), 1145-1147. https://doi.org/doi:10.1126/science.282.5391.1145

Toivonen, S., Lundin, K., Balboa, D., Ustinov, J., Tamminen, K., Palgi, J., Trokovic, R., Tuuri, T., & Otonkoski, T. (2013). Activin A and Wnt-dependent specification of human definitive endoderm cells. *Experimental Cell Research*, *319*(17), 2535-2544. https://doi.org/https://doi.org/10.1016/j.yexcr.2013.07.007

Tomic, R., Antonescu-Turcu, A., & Jacobs, E. R. (2013). Development, Anatomy, and Physiology of the Lungs. In R. Shaker, P. C. Belafsky, G. N. Postma, & C. Easterling (Eds.), *Principles of Deglutition: A Multidisciplinary Text for Swallowing and its Disorders* (pp. 199-214). Springer New York. https://doi.org/10.1007/978-1-4614-3794-9_14

Torres-Padilla, M.-E., & Chambers, I. (2014). Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. *Development*, *141*(11), 2173-2181. https://doi.org/10.1242/dev.102624

Totaro, A., Panciera, T., & Piccolo, S. (2018). YAP/TAZ upstream signals and downstream responses. *Nature Cell Biology*, *20*(8), 888-899. https://doi.org/10.1038/s41556-018-0142-z

Touboul, T., Hannan, N. R. F., Corbineau, S., Martinez, A., Martinet, C., Branchereau, S., Mainot, S., Strick-Marchand, H., Pedersen, R., Di Santo, J., Weber, A., & Vallier, L. (2010). Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. *Hepatology*, *51*(5), 1754-1765. https://doi.org/https://doi.org/10.1002/hep.23506

Travaglini, K. J., Nabhan, A. N., Penland, L., Sinha, R., Gillich, A., Sit, R. V., Chang, S., Conley, S. D., Mori, Y., Seita, J., Berry, G. J., Shrager, J. B., Metzger, R. J., Kuo, C. S., Neff, N., Weissman, I. L., Quake, S. R., & Krasnow, M. A. (2020). A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature*, *587*(7835), 619-625. https://doi.org/10.1038/s41586-020-2922-4

Treutlein, B., Brownfield, D. G., Wu, A. R., Neff, N. F., Mantalas, G. L., Espinoza, F. H., Desai, T. J., Krasnow, M. A., & Quake, S. R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature*, *509*(7500), 371-375. https://doi.org/10.1038/nature13173

Tuyl, M. v., Liu, J., Groenman, F., Ridsdale, R., Han, R. N. N., Venkatesh, V., Tibboel, D., & Post, M. (2006). Iroquois genes influence proximo-distal morphogenesis during rat lung development. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, *290*(4), L777-L789. https://doi.org/10.1152/ajplung.00293.2005

Vallier, L., Alexander, M., & Pedersen, R. A. (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *Journal of Cell Science*, *118*(19), 4495-4509. https://doi.org/10.1242/jcs.02553

Vallier, L., Touboul, T., Chng, Z., Brimpari, M., Hannan, N., Millan, E., Smithers, L. E., Trotter, M., Rugg-Gunn, P., Weber, A., & Pedersen, R. A. (2009). Early Cell Fate Decisions of Human Embryonic Stem Cells and Mouse Epiblast Stem Cells Are Controlled by the Same Signalling Pathways. *PLoS One*, *4*(6), e6082. https://doi.org/10.1371/journal.pone.0006082

Vicary, G. W., Vergne, Y., Santiago-Cornier, A., Young, L. R., & Roman, J. (2016). Pulmonary Fibrosis in Hermansky-Pudlak Syndrome. *Ann Am Thorac Soc*, *13*(10), 1839-1846. https://doi.org/10.1513/AnnalsATS.201603-186FR

Vincent, S. D., Dunn, N. R., Hayashi, S., Norris, D. P., & Robertson, E. J. (2003). Cell fate decisions within the mouse organizer are governed by graded Nodal signals. *Genes Dev*, *17*(13), 1646-1662. https://doi.org/10.1101/gad.1100503

Viotti, M., Nowotschin, S., & Hadjantonakis, A.-K. (2014). SOX17 links gut endoderm morphogenesis and germ layer segregation. *Nature Cell Biology*, *16*(12), 1146-1156. https://doi.org/10.1038/ncb3070

Viswanathan, P., Gaskell, T., Moens, N., Culley, O. J., Hansen, D., Gervasio, M. K., Yeap, Y. J., & Danovi, D. (2014). Human pluripotent stem cells on artificial microenvironments: a high content perspective. *Front Pharmacol*, *5*, 150. https://doi.org/10.3389/fphar.2014.00150

Volpato, V., & Webber, C. (2020). Addressing variability in iPSC-derived models of human disease: guidelines to promote reproducibility. *Dis Model Mech*, *13*(1). https://doi.org/10.1242/dmm.042317

Wambach, J. A., Yang, P., Wegner, D. J., An, P., Hackett, B. P., Cole, F. S., & Hamvas, A. (2010). Surfactant protein-C promoter variants associated with neonatal respiratory distress syndrome reduce transcription. *Pediatr Res*, *68*(3), 216-220. https://doi.org/10.1203/PDR.0b013e3181eb5d68

Wan, H., Dingle, S., Xu, Y., Besnard, V., Kaestner, K. H., Ang, S. L., Wert, S., Stahlman, M. T., & Whitsett, J. A. (2005). Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J Biol Chem*, *280*(14), 13809-13816. https://doi.org/10.1074/jbc.M414122200

Wan, H., Xu, Y., Ikegami, M., Stahlman, M. T., Kaestner, K. H., Ang, S.-L., & Whitsett, J. A. (2004). Foxa2 is required for transition to air breathing at birth. *Proceedings of the National Academy of Sciences*, *101*(40), 14449-14454. https://doi.org/doi:10.1073/pnas.0404424101

Wang, Y., Srinivasan, K., Siddiqui, M. R., George, S. P., Tomar, A., & Khurana, S. (2008). A novel role for villin in intestinal epithelial cell survival and homeostasis. *J Biol Chem*, *283*(14), 9454-9464. https://doi.org/10.1074/jbc.M707962200

Weaver, T. E., & Conkright, J. J. (2001). Function of surfactant proteins B and C. *Annual Review of Physiology*, *63*(1), 555-578.

Whitsett, J. A. (2018). Airway Epithelial Differentiation and Mucociliary Clearance. *Annals of the American Thoracic Society*, *15*(Supplement_3), S143-S148. https://doi.org/10.1513/AnnalsATS.201802-128AW

Whitsett, J. A., Wert, S. E., & Weaver, T. E. (2010). Alveolar surfactant homeostasis and the pathogenesis of pulmonary disease. *Annu Rev Med*, *61*, 105-119. https://doi.org/10.1146/annurev.med.60.041807.123500

Will, M., Qin, A. C. R., Toy, W., Yao, Z., Rodrik-Outmezguine, V., Schneider, C., Huang, X., Monian, P., Jiang, X., & De Stanchina, E. (2014). Rapid induction of apoptosis by PI3K inhibitors is dependent upon their transient inhibition of RAS–ERK signaling. *Cancer discovery*, *4*(3), 334-347.

Wong, A. P., Bear, C. E., Chin, S., Pasceri, P., Thompson, T. O., Huan, L. J., Ratjen, F., Ellis, J., & Rossant, J. (2012). Directed differentiation of human pluripotent stem cells into mature airway epithelia expressing functional CFTR protein. *Nat Biotechnol*, *30*(9), 876-882. https://doi.org/10.1038/nbt.2328

Wrighton, P. J., Klim, J. R., Hernandez, B. A., Koonce, C. H., Kamp, T. J., & Kiessling, L. L. (2014). Signals from the surface modulate differentiation of human pluripotent stem cells through glycosaminoglycans and integrins. *Proceedings of the National Academy of Sciences*, *111*(51), 18126-18131. https://doi.org/doi:10.1073/pnas.1409525111

Wu, D., & Pan, W. (2010). GSK3: a multifaceted kinase in Wnt signaling. *Trends in Biochemical Sciences*, *35*(3), 161-168. https://doi.org/10.1016/j.tibs.2009.10.002

Wu, M., Zhang, X., Lin, Y., & Zeng, Y. (2022). Roles of airway basal stem cells in lung homeostasis and regenerative medicine. *Respiratory Research*, *23*(1), 122. https://doi.org/10.1186/s12931-022-02042-5

Wuyts, W. A., Agostini, C., Antoniou, K. M., Bouros, D., Chambers, R. C., Cottin, V., Egan, J. J., Lambrecht, B. N., Lories, R., Parfrey, H., Prasse, A., Robalo-Cordeiro, C., Verbeken, E., Verschakelen, J. A., Wells, A. U., & Verleden, G. M. (2013). The pathogenesis of pulmonary fibrosis: a moving target. *Eur Respir J*, *41*(5), 1207-1218. https://doi.org/10.1183/09031936.00073012

Xu, X., Browning, V. L., & Odorico, J. S. (2011). Activin, BMP and FGF pathways cooperate to promote endoderm and pancreatic lineage cell differentiation from human embryonic stem cells. *Mechanisms of Development*, *128*(7), 412-427. https://doi.org/https://doi.org/10.1016/j.mod.2011.08.001

Yamamoto, M., Saijoh, Y., Perea-Gomez, A., Shawlot, W., Behringer, R. R., Ang, S.-L., Hamada, H., & Meno, C. (2004). Nodal antagonists regulate formation of the anteroposterior axis of the mouse embryo. *Nature*, *428*(6981), 387-392. https://doi.org/10.1038/nature02418

Yamamoto, Y., Gotoh, S., Korogi, Y., Seki, M., Konishi, S., Ikeo, S., Sone, N., Nagasaki, T., Matsumoto, H., Muro, S., Ito, I., Hirai, T., Kohno, T., Suzuki, Y., & Mishima, M. (2017). Long-term expansion of alveolar stem cells derived from human iPS cells in organoids. *Nature Methods*, *14*(11), 1097-1106. https://doi.org/10.1038/nmeth.4448

Yamanaka, S. (2020). Pluripotent stem cell-based cell therapy—promise and challenges. *Cell Stem Cell*, *27*(4), 523-531.

Yang, H., Lu, M. M., Zhang, L., Whitsett, J. A., & Morrisey, E. E. (2002). GATA6 regulates differentiation of distal lung epithelium. *Development*, *129*(9), 2233-2246. https://doi.org/10.1242/dev.129.9.2233

Yang, K., Wang, X., Zhang, H., Wang, Z., Nan, G., Li, Y., Zhang, F., Mohammed, M. K., Haydon, R. C., Luu, H. H., Bi, Y., & He, T.-C. (2016). The evolving roles of canonical WNT signaling in stem cells and tumorigenesis: implications in targeted cancer therapies. *Laboratory Investigation*, *96*(2), 116-136. https://doi.org/10.1038/labinvest.2015.144

Yap, L., Tay, H. G., Nguyen, M. T. X., Tjin, M. S., & Tryggvason, K. (2019). Laminins in Cellular Differentiation. *Trends in Cell Biology*, *29*(12), 987-1000. https://doi.org/10.1016/j.tcb.2019.10.001

Ye, L., Swingen, C., & Zhang, J. (2013). Induced pluripotent stem cells and their potential for basic and clinical sciences. *Curr Cardiol Rev*, *9*(1), 63-72. https://doi.org/10.2174/157340313805076278

Yiangou, L., Ross, A. D., Goh, K. J., & Vallier, L. (2018). Human pluripotent stem cell-derived endoderm for modeling development and clinical applications. *Cell Stem Cell*, *22*(4), 485-499.

Yin, B. W. T., Kiyamova, R., Chua, R., Caballero, O. L., Gout, I., Gryshkova, V., Bhaskaran, N., Souchelnytskyi, S., Hellman, U., Filonenko, V., Jungbluth, A. A., Odunsi, K., Lloyd, K. O., Old, L. J., & Ritter, G. (2008). Monoclonal antibody MX35 detects the membrane transporter NaPi2b (SLC34A2) in human carcinomas. *Cancer Immunity*, *8*(1). https://doi.org/10.1158/1424-9634.Dcl-3.8.1

Youk, J., Kim, T., Evans, K. V., Jeong, Y. I., Hur, Y., Hong, S. P., Kim, J. H., Yi, K., Kim, S. Y., Na, K. J., Bleazard, T., Kim, H. M., Fellows, M., Mahbubani, K. T., Saeb-Parsy, K., Kim, S. Y., Kim, Y. T., Koh, G. Y., Choi, B. S., . . . Lee, J. H. (2020). Three-Dimensional Human Alveolar Stem Cell Culture Models Reveal Infection Response to SARS-CoV-2. *Cell Stem Cell*, *27*(6), 905-919.e910. https://doi.org/10.1016/j.stem.2020.10.004

Yu, J. S., Ramasamy, T. S., Murphy, N., Holt, M. K., Czapiewski, R., Wei, S. K., & Cui, W. (2015). PI3K/mTORC2 regulates TGF-β/Activin signalling by modulating Smad2/3 activity via linker phosphorylation. *Nat Commun*, *6*, 7212. https://doi.org/10.1038/ncomms8212

Yu, P. B., Deng, D. Y., Lai, C. S., Hong, C. C., Cuny, G. D., Bouxsein, M. L., Hong, D. W., McManus, P. M., Katagiri, T., Sachidanandan, C., Kamiya, N., Fukuda, T., Mishina, Y., Peterson, R. T., & Bloch, K. D. (2008). BMP type I receptor inhibition reduces heterotopic ossification. *Nature Medicine*, *14*(12), 1363-1369. https://doi.org/10.1038/nm.1888

Zacharias, W. J., Frank, D. B., Zepp, J. A., Morley, M. P., Alkhaleel, F. A., Kong, J., Zhou, S., Cantu, E., & Morrisey, E. E. (2018). Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. *Nature*, *555*(7695), 251-255. https://doi.org/10.1038/nature25786

Zakrzewski, W., Dobrzyński, M., Szymonowicz, M., & Rybak, Z. (2019). Stem cells: past, present, and future. *Stem Cell Research & Therapy*, *10*(1), 68. https://doi.org/10.1186/s13287-019-1165-5

Zepp, J. A., Zacharias, W. J., Frank, D. B., Cavanaugh, C. A., Zhou, S., Morley, M. P., & Morrisey, E. E. (2017). Distinct mesenchymal lineages and niches promote epithelial self-renewal and myofibrogenesis in the lung. *Cell*, *170*(6), 1134-1148. e1110.

Zhang, H., Menzies, K. J., & Auwerx, J. (2018). The role of mitochondria in stem cell fate and aging. *Development*, *145*(8). https://doi.org/10.1242/dev.143420

Zolin, A., Bossi, A., Cirilli, N., Kashirskaya, N., & Padoan, R. (2018). Cystic Fibrosis Mortality in Childhood. Data from European Cystic Fibrosis Society Patient Registry. *Int J Environ Res Public Health*, *15*(9). https://doi.org/10.3390/ijerph15092020

Zorn, A. M., & Wells, J. M. (2009). Vertebrate endoderm development and organ formation. *Annu Rev Cell Dev Biol*, *25*, 221-251. https://doi.org/10.1146/annurev.cellbio.042308.113344