



Caldera-Quevedo, Irene (2024) *Investigating the impact of lung cancer cell-of-origin on tumour metabolic phenotype and heterogeneity*. PhD thesis.

<https://theses.gla.ac.uk/84078/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>  
[research-enlighten@glasgow.ac.uk](mailto:research-enlighten@glasgow.ac.uk)



University  
of Glasgow

# Investigating the Impact of Lung Cancer Cell-of-origin on Tumour Metabolic Phenotype and Heterogeneity

Irene Caldera-Quevedo

Submitted in fulfilment of the requirements for the Degree  
of Degree of Doctor of Philosophy

School of Cancer Sciences, College of Medical, Veterinary  
& Life Sciences University of Glasgow

September 2023



CANCER  
RESEARCH  
UK

BEATSON  
INSTITUTE

## Abstract

Non-small-cell lung cancer has been described as highly heterogenous which results in different metabolic phenotypes. There are multiple factors which contribute to this heterogeneity, one of which is the tumour cell-of-origin. In the lung, there are five cell types reported to be cells-of-origin: alveolar epithelial type 2, club, basal, neuroendocrine and bronchioalveolar stem cells. This project focuses on the interaction between the cell-of-origin and the metabolic phenotype of lung cancer, and we aim to assess the contribution of the cell-of-origin to lung cancer metabolic resultant phenotype and heterogeneity.

To accomplish this, we have established two complementary model systems, one in vitro and one in vivo. In our in vitro model, we isolated specific lung cell types, including AT2 cells, basal cells, and club cells, utilising their unique cell surface markers. By introducing oncogenic KRAS mutations and deleting the P53 gene, we are creating lineage-restricted organoids. These organoids will serve as valuable tools for characterizing the metabolic aspects of tumours arising from different cell-of-origin backgrounds within an in vitro setting.

In our in vivo model, we induced NSCLC tumours in mice with genetic modifications using viral vectors, namely Ad5-mSPC-Cre, Ad5-CC10-Cre, and Ad5-bk5-Cre. These vectors are selectively expressed in AT2, club, and basal cells, respectively. To ensure the validity of our comparisons, we have carefully monitored tumour growth dynamics and burden in these mouse models. Our comprehensive analysis has revealed three distinct transcriptomic subtypes (S1, S2, and Acetate) within these NSCLC tumours. Notably, S1 and Acetate subtypes are enriched in tumours originating from specific cell types. Positron emission tomography (PET) imaging has unveiled metabolic variations, with S1 tumours displaying heightened [ $^{18}\text{F}$ ]FDG uptake and the Acetate subtype exhibiting increased [ $^{11}\text{C}$ ]acetate uptake. Furthermore, our multi-omics approach, encompassing transcriptomics, proteomics, and metabolomics, has exposed disparities in critical metabolic pathways, such as glycolysis, hypoxia response, and apoptosis.

In summary, our research provides a comprehensive examination of the metabolic heterogeneity of NSCLC based on the cell-of-origin independently of genomic alterations.

## **Author's Declaration**

I hereby declare that I am the only author of this thesis and all work presented in this document is my own, unless stated otherwise.

I certify that no part of this thesis has been submitted for the award of any other degree at the University of Glasgow or any other institution.

Irene Caldera-Quevedo

September 2023

## Acknowledgments

This research project and this thesis would not have been possible without the support provided by all the amazing people at the CRUK Beatson Institute. I would like to express my deepest gratitude to my supervisor Dr David Lewis, for giving me the opportunity to embark on this PhD journey and for his continuous guidance and support. I would also like to thank to my second supervisor Dr. Daniel Murphy for his mentorship. I am also thankful to my two reviewers Dr. Laura Machesky and Dr Oliver Maddocks for their advice and helpful discussions on the project.

I would also like to thank to all the people who is part or have been part of the R07/Y65 group during these four years. I'd like to thank Agata Mackintosh for her unvaluable support, for listening to me when I needed to vent, and for caring about me during these four years. I would also like to thank to Dr Gaurav Malviya and Emma Johnson for their patience when they were teaching me new lab techniques, for always supporting my learning process, answering my numerous questions and for all their help, kindness and support. Special thanks to all three of them for being patient with me when I was first learning how to work with animals and for kindly correcting my mistakes. I would also like to thank Emer for her support and friendship these years and all the rest of the members of our research group. Their support, training, friendship and advice during this project.

I would also like to thank all the people at the Beatson who has helped to bring this project including the BSU/BRU staff, people from the molecular technology group and the flow cytometry team and Dr Maria Terradas for their training in flow cytometry and analysis. I also would like to thank people from the metabolomic, proteomic and bioinformatic department for helping me with such complex analysis. Many thanks to the people working in the immunohistochemistry and the molecular pathology department for their guidance in data analysis. Additionally, many thanks to Cancer Research UK for funding my studentship.

My deepest appreciation goes to my Glaswegian family. All the people who made this city feel like home during the most difficult times, and who show me their love and support every day. Specially to Irati, Ilse, Ane and Rachel. These last four years have been the happiest thanks to you and our shared moments. You make me want to stay in Glasgow and you make me want to keep growing and challenge myself. Thanks to Andres for his Spanish food when I most miss home. Thanks to Seoras for our endless philosophical conversations. Thanks to Dalia for turning a flat into a home. This thesis has been submitted because of you all.

Thanks to Lavinia and Judith. I started this journey with you and even though I finish it 2,000 kilometres away, you never let me go. I feel I had you both with me every step of the way.

I also want to thank my friends here at the Beatson, especially Rachel, Adiba and Anastasia. Working with such amazing scientists and wonderful people has made my time here unforgettable, it has made me grow as a scientist and it has made me challenge myself in and outside the lab.

Last but not least, I would like to deeply thank my family: specially my parents whose unconditional love and support have allowed me to move to this cold and rainy country in pursuit of my passion for science and who always welcome me back home as if I have never left. They have always inspired me to do my best and stay motivated even when things don't work out as we initially planned. Finally, I would like to dedicate this thesis to my sister Nuria - I would not be here without you teasing me all the time - and to my grandmother Carmen, who was never allowed to indulge her curiosity and eagerness to learn because of the time she was born in, but who never stopped trying.

# Table of Contents

Abstract .....	i
Author's Declaration .....	iii
Acknowledgments .....	iv
Table of Contents .....	vi
List of Tables.....	x
List of Figures.....	xi
List of Abbreviations and Definitions .....	xiv
Chapter 1 Introduction.....	1
1.1 Pulmonary anatomy and cellular composition .....	1
1.1.1 Tissue stem cells proliferation and differentiation.....	5
1.2 Lung embryonic development .....	7
1.3 Lung cancer: current state .....	11
1.3.1 Epidemiology and risk factors .....	11
1.3.2 Lung cancer staging and histological patterns .....	12
1.3.3 Lung cancer symptoms and diagnostics .....	13
1.3.4 Lung cancer current treatments.....	14
1.4 Lung cancer development.....	15
1.5 Lung cancer heterogeneity.....	17
1.5.1 Lung cancer heterogeneity: genomic alterations .....	18
1.5.2 Lung cancer heterogeneity: cell-of-origin .....	20
1.6 Cancer metabolism.....	26
1.6.1 The effect of genomic alterations on lung cancer metabolism .....	35
1.6.2 The effect of cell-of-origin on cancer metabolism .....	36
1.7 Hypothesis, aims and objectives .....	37
Chapter 2 Materials and Methods.....	39
2.1 Animal work .....	39
2.1.1 Breeding and Maintenance of animals.....	39
2.1.2 SPC, CC10, CGRP and CMV Mouse Induction .....	39
2.1.3 Naphthalene optimisation .....	40
2.1.4 KT5 Mouse Induction .....	40
2.1.5 Mouse Model Characterisation .....	40
2.1.6 Histology .....	42
2.1.7 H&E staining.....	42
2.1.8 Immunohistochemistry.....	43

2.1.9	Histological Analysis.....	44
2.1.10	DNA, RNA and Protein Isolation.....	45
2.1.11	Metabolite Isolation .....	45
2.1.12	Metabolomic analysis of Tumours and Normal lung Tissue .....	46
2.1.13	Bioinformatic Analysis of Metabolomic data.....	47
2.1.14	RNA Sequencing of tumours and normal lung tissue.....	47
2.1.15	Bioinformatic analysis of tumours and normal lung tissue .....	48
2.1.16	Proteomics.....	50
2.1.17	Proteomic Data Analysis .....	51
2.1.18	PET Imaging .....	52
2.2	Tissue culture .....	53
2.2.1	Lung Tissue Dissociation .....	53
2.2.2	Trachea Tissue Dissociation .....	54
2.2.3	Extracellular Cell Staining.....	54
2.2.4	Compensation and flow cytometry analysis .....	55
2.2.5	Lung epithelial and mesenchymal cell sorting .....	56
2.2.6	Mesenchymal cell isolation.....	56
2.2.7	Endothelial Cells Isolation.....	57
2.2.8	Organoids Culture .....	58
2.3	Statistics .....	59
Chapter 3 Optimising lineage-specific lung cancer organoid cell culture protocol.		
60		
3.1	Introduction .....	60
3.1.1	Hypothesis and Aims .....	65
3.2	Results .....	66
3.2.1	Epithelial lung cell isolation .....	66
3.2.2	Detection of lung cancer AT2, BASCs, Club and neuroendocrine cells with flow cytometry .....	67
3.2.3	Detection of lung cancer basal cell-of-origin with Flow Cytometry.	71
3.2.4	Generation of lung organoids from lung epithelial cells .....	73
3.2.5	Isolation of support cells .....	77
3.2.6	AT2 cells organoid culture .....	78
3.2.7	Protocol checkpoints.....	80
3.3	Discussion .....	83
3.3.1	Epithelial lung cell isolation .....	83
3.3.2	Virus vectors specificity .....	83
3.3.3	Organoids protocol development.....	85

3.3.4	Organoids protocol checkpoints .....	88
3.4	Conclusion .....	88
Chapter 4	Mouse model characterisation .....	89
4.1	Introduction .....	89
4.1.1	Hypothesis and Aims .....	96
4.2	Results .....	96
4.2.1	Generation of KT5 mouse model .....	96
4.2.2	Effect of the virus vector titre on target cells .....	98
4.2.3	Mouse models characterisation.....	102
4.3	Discussion.....	109
4.3.1	KT5 mouse model.....	109
4.3.2	Targeted cells by cell activation .....	110
4.3.3	Mouse model characterisation and immunohistochemistry subtype 111	
4.3.4	Clinical subtypes classification .....	112
4.4	Conclusion .....	114
Chapter 5	Metabolic, Transcriptomic and Proteomic Tumour Characterisation.	115
5.1	Introduction .....	115
5.2	Hypothesis and Aims .....	120
5.3	Results .....	121
5.3.1	Transcriptomic results tumour vs normal tissue .....	121
5.3.2	Differences between CC10 vs SPC tumours .....	123
5.3.3	Description of S1, S2 and ACE subgroups.....	134
5.3.4	PET imaging .....	137
5.3.5	Differences between S1 and ACE tumours .....	138
5.4	Discussion.....	157
5.4.1	Transcriptomics results tumour vs normal tissue .....	157
5.4.2	Differences between CC10 and SPC tumours .....	157
5.4.3	Description of S1, S2 and ACE subgroups.....	162
5.4.4	PET imaging .....	164
5.4.5	Differences between S1 and ACE tumours .....	165
5.5	Conclusion .....	170
Chapter 6	Discussion .....	172
6.1	Effect of the cell-of-origin in metabolism of the resultant tumour .....	172
6.2	Limitations of the study .....	173
6.3	Future work .....	174
6.3.1	Lung Cancer Organoids .....	174

6.3.2 Metabolic, Transcriptomic and Proteomic characterisation .....175  
Supplementary information..... 1788  
References..... 180

## List of Tables

Table 2.1: List of antibodies, autostrainers and antigen retrieval methods.....	44
Table 2.2: List of antibodies used in flow cytometry analysis of lung cells .....	55
Table 3.1: Summary of percentage of each cell type .....	71
Table 3.2: Summary of percentage of each cell type - basal cells.....	73
Table 3.3: Summary of percentage of each cell type - EGFP cells.....	82
Table 4.1: Summary of coefficient of variation.....	104
Table 5.1: Summary of significant genes in SPC and CC10 tumours .....	124

## List of Figures

Figure 1.1. Schematic representation of lung main anatomical elements.....	1
Figure 1.2. Schematic representation of the cellular component and their distribution along the lung airways. ....	3
Figure 1.3. Stages of human and mouse lung embryony development and the differential expression of SOX2, SOX9, and FGF10 during human and mouse lung in the pseudoglandular stage. ....	10
Figure 1.4. Evolutionary development of lung cancer. ....	17
Figure 1.5. Schematic representation of the cells-of-origin of lung cancer in the murine lung epithelium and their specifically expressed genes. ....	21
Figure 1.6. Schematic representation of glycolysis and TCA pathways, and its metabolic branches. ....	28
Figure 1.7. Schematic representation of purine and pyrimidine biosynthesis....	30
Figure 1.8. Effect on MYC expression in glycolytic genes.....	32
Figure 1.9. Effect on MYC expression in nucleotide biosynthesis genes.....	33
Figure 1.10. Schematic representation of changes in glucose metabolism due to the activation of hypoxia pathway.....	34
Figure 3.1. Schematic representation of mTmG reporter gene set in mice. ....	64
Figure 3.2. Schematic representation of both KP and mTmG systems combination in the MTMGKP mouse model .....	65
Figure 3.3. Lung dissociation protocol B shows a significant increase in the percentage of living cells .....	66
Figure 3.4. Gating strategy for lung epithelial cells panel used to identify AT2 cells and BASCs .....	68
Figure 3.5. Gating strategy for lung epithelial cells panel used to identify club cells and BASCs .....	69
Figure 3.6. Gating strategy for lung epithelial cells panel used to identify neuroendocrine cells.....	70
Figure 3.7. Gating strategy for lung epithelial cells panel used to identify basal cells. ....	72
Figure 3.8. AT2 sorted cells 3D seeding strategies are not able to sustain the organoids growth and proliferation in vitro. ....	75
Figure 3.9. Seeded whole lung single cell suspension show a successful organoid culture with different Matrigel stiffness. ....	76
Figure 3.10. Endothelial cells proliferate and grow faster in culture than mesenchymal cells and therefore they are used as the most optimal cell support system for 3D lung culture. ....	78
Figure 3.11. Schematic representation of the protocol followed to isolate, activate and seed AT2 cells in 3D cultures. ....	79

Figure 3.12. Organoids are growing in culture as a result of complementing the media with FGF, cholera toxin and insulin and adding support cells. ....	79
Figure 3.13. Organoids are growing in culture as a result of complementing the media with FGF, cholera toxin and insulin and adding support cells and they expand themselves in vitro. ....	80
Figure 3.14. Gating strategy for lung epithelial cells panel used to identify AT2, BASCs and club cells, respectively. ....	82
Figure 4.1. Schematic representation of KP reporter gene set in mice. ....	92
Figure 4.2. Summary of the KRAS pathway and its effects in cell culture. ....	94
Figure 4.3. Schematic representation of P53 pathway and its activation effects in cellular processes. ....	95
Figure 4.4. Naphthalene treatment of 135 mg/Kg reduces CC10 staining in the trachea epithelium but maintains CC10 staining in the lung epithelium which will allow the regeneration of CC10 and the mouse survival. ....	97
Figure 4.5. Gating strategy for lung epithelial cells panel used to identify SPC, CC10 and KT5 EGFP positive cells. ....	99
Figure 4.6. Gating strategy for lung epithelial cells panel used to identify SPC and CC10 EGFP positive cells. ....	101
Figure 4.7. Growth dynamic and location data of CC10, SPC and KT5 tumours. ....	103
Figure 4.8. Histopathological analysis of TTF1 and P63 in SPC, CC10 and KT5 tumours. ....	106
Figure 4.9. H&E patterns and EVG staining for SPC, CC10, KT5 tumours. ....	108
Figure 5.1. KRas and P53 expression in SPC and CC10 tumours and PCA showing the separation between healthy lung tissue and tumour samples. ....	122
Figure 5.2. CC10 and SPC groups are separated in an unsupervised PCA analysis ....	123
Figure 5.3. Volcano plot showing significant differential genes between CC10 and SPC tumours. ....	124
Figure 5.4. Hallmark GSEA transcriptomic analysis highlighting the impact of the cell-of-origin on glycolysis and proliferation-related pathways. ....	125
Figure 5.5. KEGG GSEA transcriptomic analysis highlighting the impact of the cell-of-origin on glycolysis and immune pathways. ....	126
Figure 5.6. Hallmark GSEA transcriptomic analysis highlighting the impact of the cell-of-origin on glycolysis and proliferation-related pathways. ....	127
Figure 5.7. KEGG GSEA transcriptomic analysis highlighting the impact of the cell-of-origin on glycolysis and neuroactive interactions. ....	128
Figure 5.8. Proteomic data indicates an increase of glycolytic and nucleotide synthesis related genes in CC10 tumours. ....	129
Figure 5.9. Schematic representation of the glycolysis and the proteins found upregulated in the analysis. ....	131
Figure 5.10. Schematic representation of the nucleotide biosynthesis and the genes found upregulated in this analysis. ....	132

Figure 5.11. CC10 and SPC samples are do not differ a lot from a metabolic point of view. ....	133
Figure 5.12. Gene expression patterns of S1 and S2 gene signatures, revealed at least two transcriptomic subtypes .....	134
Figure 5.13. Gene expression patterns of S1, S2, FDG and Ace gene signatures, revealed ACE as novel transcriptomic subtype.....	135
Figure 5.14. SPC tumours are significantly enriched in ACE samples while CC10 tumours are enriched in S1 group.....	136
Figure 5.15. SPC and Cc10 tumours differ in [18F]FDG and [11C]acetate uptake. ....	138
Figure 5.16. Gene expression patterns of S1, and ACE samples of significant genes in glycolysis gene sets.....	140
Figure 5.17. Glycolysis map shows the genes upregulated in both S1 and ACE tumours. ....	141
Figure 5.18. Nucleotide synthesis map with the upregulated genes in S1 tumours. ....	143
Figure 5.19. Gene expression patterns of S1, and ACE samples of significant genes in hypoxia gene set. ....	144
Figure 5.20. Schematic representation of glycolysis and the upregulated genes in S1 and ACE tumours based on the hypoxia gene sets. ....	146
Figure 5.21. Gene expression patterns of S1, and ACE samples of significant genes in apoptosis (a) and P53 (b) gene set .....	147
Figure 5.22. Sftpc expression is significantly higher in ACE than in S1 tumours. ....	149
Figure 5.23. Proteomic data suggest a higher expression of glycolytic (a), nucleotide (b) and arginine (c) biosynthesis, one carbon metabolism (c) and translation (d) in S1 tumours. ....	150
Figure 5. 24. Schematic representation of glycolysis and TCA pathway. ....	152
Figure 5.25. Schematic representation of nucleotides synthesis pathway.....	153
Figure 5.26. S1 tumours are metabolic different from ACE tumours. ....	155
Figure 5.27. Joint pathway analysis of transcriptomics and metabolomics revealed differences in the glycolytic metabolism and nucleotide synthesis ...	156
Figure 5.28. Mechanism of action of E2F transcription factor in cell proliferation. ....	160
Figure 5.29. Schematic representation of the overall picture of S1 and ACE tumours. ....	171

## List of Abbreviations and Definitions

1,3BPG	1,3 biphosphoglyceric acid
2PG	2-phosphoglyceric acid
3PG	3-phosphoglyceric acid
Acetyl-CoA	Acetyl coenzyme A
ADC	Adenocarcinoma
ADP	Adenosine diphosphate
AK	Adenylate kinase
ALK	Anaplastic Lymphoma Kinase
AMP	Adenosine monophosphate
APOBEC	Apolipoprotein B mRNA editing enzyme
AT1	Alveolar cells type 1
AT2	Alveolar cells type 2
ASCL1	Achaete-scute homolog 1
ATP	Adenosine triphosphate
BADJ	Bronchioalveolar duct junction
BASCs	Bronchioalveolar stem cells
BSA	Bovine serum albumin
CC10	Scgb1a1
CDP	Cytosine diphosphate
CMP	Cytosine monophosphate
CTP	Cytosine triphosphate
CDKN2A	Cyclin dependent kinase inhibitor 2A
CGRP	Calcitonin related polypeptide
CTCs	Circulating tumour cells
ctDNA	Circulating tumour DNA
CPS1	Carbamoyl phosphate synthetase-1
CS	Citrate synthase
CT	Computerised tomography
DHAP	Dihydroxyacetone phosphate
DNA	Deoxyribonucleic acid

ECGS	Endothelial cell growth supplement
EGFP	Enhanced green fluorescent protein
EpCam	Epithelial cell adhesion molecule
EGFR	Epidermal growth factor receptor
ES	Enrichment score
ESI	Electrospray ionisation
EVG	Elastin Van Gieson
F1,6BP	Fructose 1,6-biphosphate
F2,6BP	Fructose 2,6-biphosphate
F6P	Fructose 6-phosphate
FBS	Fetal bovine serum
FDR	False discovery rate
FGF	Fibroblast growth factor
FGFR	Fibroblast growth factor receptor
FH	Fumarate hydratase
G12D	Glycine - 12 - Aspartic acid
G6P	Glucose 6-phosphate
GAPD	Glyceraldehyde 3-phosphate
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GDP	Guanidine diphosphate
GEMM	Genetically engineered mouse model
GLS	Glutaminase
GLUT	Glucose transporter
GMP	Guanidine monophosphate
GO	Gene Ontology
GPI	Glucose phosphate isomerase
GOT2	Glutamic-Oxaloacetic Transaminase 2
GS	Glutamine synthetase
GSH	Glutathione
GTP	Guanidine triphosphate
H&E	Haematoxylin eosin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HIF-1	Hypoxia inducible factor -1
HK	Hexokinase kinase
IDH1	Isocitrate dehydrogenase
IHC	Immunohistochemistry
IMP	Inosine monophosphate
ITS	Insulin / Transferrin / Selenium
KEGG	Kyoto Encyclopedia of Genes and Genomes
KT5	Keratin-5
KT7	Keratin-7
KT14	Keratin-14
KP	LSL-KRas <sup>G12D</sup> /+ P53 <sup>fl/fl</sup>
KGDH	Ketoglutarate dehydrogenase
LKB1	Liver kinase B1
LTOs	Lung tumour organoids
MAPK/ERK	Mitogen activated protein kinase
MCT	Monocarboxylate transporter
MDH	Malate dehydrogenase
MHCII	Major histocompatibility complex II
MRI	Magnetic resonance imaging
MsigDB	Molecular Signature Database
mT	tdTomato
mTORC1	Mammalian target of rapamycin complex 1
NADH	Nicotinamide adenine dinucleotide hydrogen
NADPH	Nicotinamide adenine dinucleotide phosphate hydrogen
NE	Neuroendocrine
NES	Normalised enrichment score
NGFR	Nerve growth factor receptor
NSCLC	Non-small cell lung cancer
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PDH	Pyruvate dehydrogenase
PDK1	Pyruvate dehydrogenase kinase 1

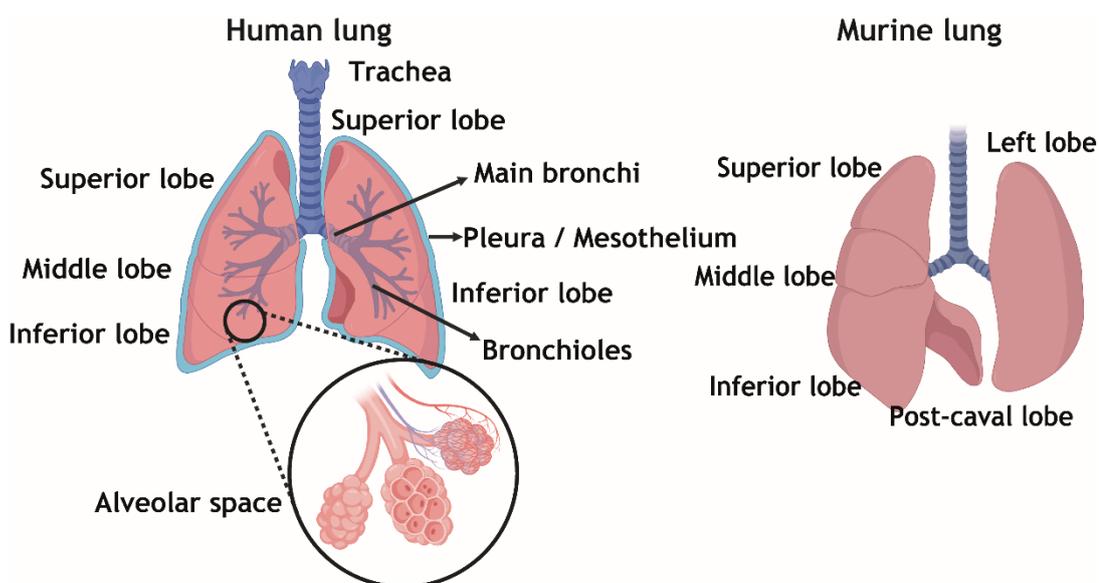
PET	Positron emission tomography
PEP	Phosphoenolpyruvate
PFK1	Phosphofructokinase
PGAM	Phosphoglycerate mutase
PGK	Kyoto Encyclopedia of Genes and Genomes
PI3K	Phosphoinositide 3-kinases
PKA	Pyruvate kinase A
PKM	Pyruvate kinase M
PPP	Pentose phosphate pathway
PRPP	Phosphoribosyl pyrophosphate
R5P	Ribose 5-phosphate
Ri5P	Ribulose 5-phosphate
RNA	Ribonucleic acid
RB	Retinoblastoma
RNAseq	RNA sequencing
ROI	Region of interest
ROS	Reactive oxygen species
SCC	Squamous cell carcinoma
SCLC	Small cell lung cancer
SDH	Succinate dehydrogenase
Shh	Sonic hedgehog
SLC	Solute carrier family
SOX	SRY-box transcription factor
SS	Succinate synthetase
SPC	Sfptc
TCA	Tricarboxylic acid
TDP	Thymidine diphosphate
TKI	Tyrosine kinase inhibitor
TMP	Thymidine monophosphate
TMT	Tandem mass tag
TGFB	Transforming growth factor beta
Trp63	Tumour protein 63

Tp53	Tumour protein 53
TPI	Triosephosphate isomerase
TTF1	Thyroid transcription factor 1
TTP	Thymidine triphosphate
UDP	Uracil diphosphate
UMP	Uracil monophosphate
UTP	Uracil triphosphate

## Chapter 1 Introduction

### 1.1 Pulmonary anatomy and cellular composition

Human lungs are located in the thoracic cavity on either side of the heart consisting of three right and two left lobes. The lungs are complex structures made of branched airways and blood vessels. The lungs are surrounded by a membrane known as the pleura, or mesothelium in mice. The proximal lung airways consist of the trachea, which has a pseudostratified epithelium and diverges into the two main bronchi (left and right). Each main bronchi divides further into secondary bronchi and subsequently into progressively narrower airways called bronchioles until the smallest one connects to the alveolar space or distal airways, which is surrounded by capillary networks. Air is inhaled and transported from the proximal airways into the alveolar space where gas exchange occurs. Oxygen then passes through the epithelial cells of the alveoli into the capillary blood vessels (Nikolic et al., 2018). Lungs are organised in lobules, but these are different between mouse and humans. In humans the right lung is divided in three lobules superior, middle and inferior. The left lung is divided in two lobules, superior and inferior. In mice, the right lung is divided in four lobules, superior, middle, inferior and post-caval lobule. The left lung is a unique lobule (Figure 1.1).



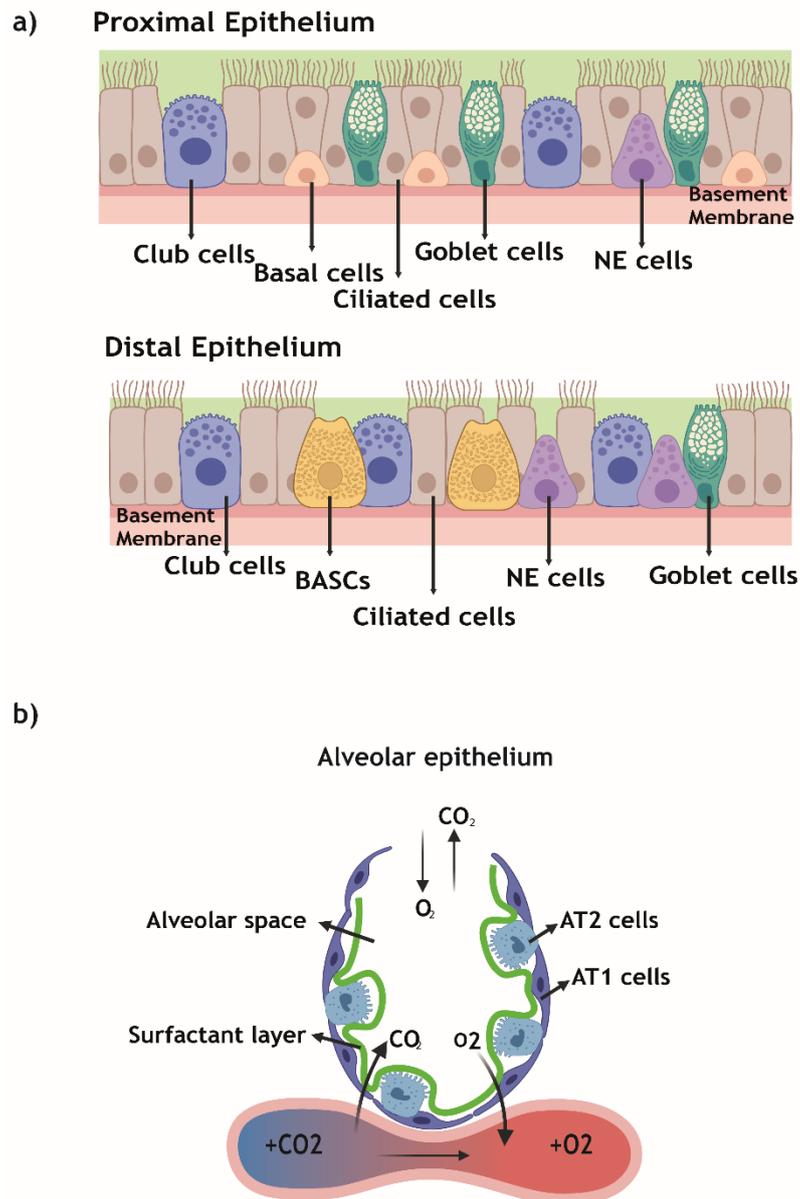
**Figure 1.1. Schematic representation of lung main anatomical elements.** Murine and human lungs differ in the organisation of lung lobules. Human lungs have three lobules in the right lung and two lobules in the left one. Murine lungs have four lobules in the right lung and one lobule in the left lung. The rest of the anatomical elements are consistent between human and murine

lungs. The proximal airways consist of the trachea which grows and branch into bronchi and bronchioles and the alveolar space at the end of these branches. Lungs are surrounded by the pleura, also called mesothelioma in mice. Created with BioRender.com.

In addition, there are several functional differences between murine and human lungs. The volume of one alveolus is  $4.2 \times 10^6 \mu\text{m}^3$  in humans compared to  $2.2 \times 10^4 \mu\text{m}^3$  in mice, and the estimated average number of alveoli is lower in mice than human,  $2.3 \times 10^6$  and  $4.8 \times 10^8$  respectively (Wansleeben et al., 2013). There are also some differences in the structure between mouse and human lungs. Human cartilage rings are present all the way from the trachea to the bronchioles, whereas in mouse lungs they are only present in the trachea and the main bronchus (Danopoulos et al., 2019). Moreover, submucosal glands are located below the epithelium, and they secrete mucus and other substances that help protect the lungs from infectious agents (Xiaoming Liu, 2004). In mouse lungs submucosal glands are limited to the trachea while in humans they are present from the trachea to bronchioles.

Finally, there are some differences in the cellular composition. Both mouse and human lung comprises various cell types, such as epithelial, endothelial, pericytes, pleura/mesothelial, smooth muscle cells surrounding bronchioles, fibroblasts, nerves, and immune cells. These cell types are distributed along the airways sometimes in specific locations (Nikolic et al., 2018). In humans, proximal airways, trachea, bronchus and bronchioles consist of a pseudostratified epithelium, in which a single layer of epithelial cells is in direct contact with the basement membrane. These cells feature apico-basal polarity, but they are more elongated than in other epithelium types and their nuclei are more densely packed (Norden, 2017). The cell composition of this epithelium varies along the proximal-distal axis of the airways tree. Below this structure blood and lymphatic vessels, smooth muscle, cartilage fibroblasts and nerves are found. Mouse tracheal and human proximal epithelium are similar, although distal branches of murine airways switch from pseudostratified to a simple columnar epithelium (Hogan et al., 2014). This consists of a single layer of tall/columnar cells attached to the basal membrane whose nuclei is towards the basal pole of the cell (Wright, 2008). Proximal airway epithelium consists of mainly secretory goblet, secretory club, ciliated, neuroendocrine, and basal cells (Figure 1.2a). The human lung epithelium consists of more goblet cells in the proximal

epithelium, with club cells found in less frequency. Mouse lung contains more club cells through the trachea and bronchi, with less goblet cells throughout (Danopoulos et al., 2019). Both in human and mice, basal and club cells have self-renew capacity which allows them to differentiate into other cell types of these epithelium after injury. In humans they are located throughout the proximal and distal airway tree although in murine airways they are only present in the trachea (Thai et al., 2021).



**Figure 1.2. Schematic representation of the cellular component and their distribution along the lung airways.** (a) Proximal epithelia consist of the trachea, bronchi and part of the bronchioles. Pseudostratified epithelium is found in both human and murine trachea and the rest of the human proximal airways. The cell composition of this epithelium varies between mice and humans and along the airways but, overall, it consists mainly secretory goblet, secretory club,

ciliated, neuroendocrine, and basal cells. Distal epithelium is found in the distal part of the bronchioles and the bronchioalveolar duct junction. Basal cells are not present in these distal parts of the epithelium, club cells frequency is reduced and bronchioalveolar stem cells (BASCs) appear. (b) Alveolar epithelium is composed of AT1 cells which carry out the oxygen exchange and AT2 cells which produce surfactant proteins. Created with BioRender.com.

Murine trachea epithelium and human airways contain a partially undifferentiated cell types called basal cells, which specifically expressed keratin-5 (*Kt5*), keratin-14 (*Kt14*), tumour protein 63 (*Trp63*) and nerve growth factor receptor (*Ngfr*). They play a structural role in the respiratory epithelium; after an injury or an epithelium insult, they are capable of proliferating and differentiating into secretory, ciliated cells or club cells after naphthalene administration (Rock and Hogan, 2011, Jason R Rock, 2009). It is unclear how these cells differentiate but some studies based on intestinal stem cells of the *Drosophila* larval midgut indicate that basal cells turn into an earlier progenitor *Trp63* negative and *KT5* positive that proliferate and differentiate to either secretory or ciliated cells in a Notch dependent manner (Rock et al., 2011b).

Club cells are located in the murine bronchioles, bronchi and trachea and they are the source for *Scgb1a1* secretoglobin (*CC10*). This is a heterogenous population in terms of the location, morphology, gene expression, and secretory profile, but few markers are known to distinguish between these subtypes. Club cells are known to act as tissue stem cells with self-renewal capacity, and they can generate ciliated cells after an insult or an injury (Rawlins et al., 2009). Systematic treatment of mice with naphthalene kills certain subpopulations of club cells, which express cytochrome *Cyp2f2*. The reconstitution of the bronchus and bronchioles epithelia are based on naphthalene-resistant club cells normally located close to neuroendocrine cell clusters and the bronchioalveolar duct junction (BADJ) (Hogan et al., 2014).

Neuroendocrine cells (NE) are a rare type of cell type in the bronchioles as they account only for 1% of the epithelium. Their role is unclear, but it is believed they play an important role in sensing stimuli in the airways such as hypoxia or nicotine, their secretory product include calcitonin (*CGRP*) and it is thought that they have a role in immune function, flow of air and blood in the airways. It has also been described that this cell type is innervated by sensory nerves fibres, and it is believed that they potentially transmit stimuli to the central nervous

system (Lommel, 2001, Rock and Hogan, 2011). Amongst other insults, NE cells serve as a sensor for hypoxia and bacterial infections. In mouse and human trachea, NE cells are clustered into groups of cells up to 30 or 8 in mouse and human airways, respectively (Nikolic et al., 2018).

Bronchoalveolar stem cells (BASCs) are found in the intersection between the airway and the alveolar space or BADJ and they are the source of secreted surfactant-associated protein C (*Sftpc* or *SPC*) and *CC10*. BASCs can proliferate and differentiate into cells from both bronchiolar and alveolar lineages after an injury (Kim et al., 2005). Their frequency is low, with less than one BASC per BADJ in average (Hogan et al., 2014). BASCs have not been identified in human airways (Kim et al., 2005).

The distal lung epithelium is called alveolar epithelium and it consists of type I and type II alveolar cells (AT1 and AT2 cells) which are surrounded by capillaries and fibroblasts (Figure 1.2b) (Herzog et al., 2008). AT1 cells perform the gas exchange, and they cover more than 95% of the gas exchange surface area. AT2 cells are cuboidal, higher in number, and they contain many secretory vesicles where they store and secrete SPC. This protein is expressed in AT2 cells. Surfactant is a mix of proteins and lipids that decreases alveolar surface tension, and it is involved in immune defence (Whitsett et al., 2010). It is known that AT2 cells are the progenitors for other cell types in the lung, including AT1 cells which perform the gas exchange and they have been described as non-proliferative, although cell turnover is low in a steady state. Changes in AT2 cell behaviour, including proliferation, differentiation and phenotypic plasticity occur in response to nitric oxide, high levels of oxygen, the chemotherapy drug bleomycin, cigarette smoke, irradiation, and viral infection (Hogan et al., 2014). Both AT1 and AT2 cells account for 95% of the epithelium surface (Rock and Hogan, 2011). AT2 cells are the main alveolar epithelial stem cell as they self-renew and differentiate into AT1 cells after an injury (Rock et al., 2011a).

### **1.1.1 Tissue stem cells proliferation and differentiation**

The differentiation process for basal, club, NE, AT2 and BASCs after an epithelium injury is not well understood. It may be that this plasticity involves dedifferentiation to a multipotent cell type followed by redifferentiation. This

switch from one cell type to another may be direct, not requiring an undifferentiated intermediate.

The interaction between epithelial cells and the stroma cells seems to be essential for lung repair and regeneration. The mechanisms behind this differentiation are not very well understood either, although it has been proved that the cell composition and cell-to-cell interactions between AT2 and BASCs cells and platelet derived growth factor receptor alpha fibroblasts and endothelial cells in the alveolar niche are key in AT2 and BASCs cell proliferation and differentiation (Hogan et al., 2014). The expression of thrombospondin-1 and bone morphogenetic protein 4 (BMP4) signalling in endothelial lung cells regulates the differentiation of BASCs (Lee et al., 2014).

*Wnt*, which plays a role in lung development, is also implicated in lung regeneration. *Wnt* signalling is activated after club cell depletion due to naphthalene treatment. As a result, the expression of the GATA6 transcription factor in BASCs leads to the proliferation of BASCs after naphthalene injury (Zhang et al., 2008). *Notch* signalling promotes the differentiation of basal cells to secretory cells at the expense of the ciliated cell differentiation (Tsao et al., 2009). *Notch* signalling has also been described to be active under the presence of reactive oxygen species (ROS) through the expression of Nuclear Factor Erythroid 2. This pair is important for basal cell self-renewal as it may be critical to maintain the correct number of basal cells in the proximal airways (Paul et al., 2014).

Finally, epigenetic regulation has also been described to play a role in lung repair and regeneration. Histone deacetylase 1 and 2 are required for the regeneration of secretory epithelium after naphthalene-induced depletion in the lung (Kazuhiro Ito et al., 2005). Histone deacetylase seems to be important for AT1 and AT2 cells to balance HOPX, which is expressed by AT1 cells and by alveolar progenitors, which recruits Histone deacetylase 2 to negatively regulate AT2-specific gene expression (Yin et al., 2006). This histone deacetylase can bind to other proteins such as FOXP transcriptions factors. This combination inhibits FOXP and results in spontaneous differentiation of club cells to goblet cells, resulting in an epithelium unable to repair itself after naphthalene injury (Li et al., 2012).

## 1.2 Lung embryonic development

To understand the different potential cells-of-origin of lung cancer and for a deeper understanding of pulmonary anatomy and tissue stem cell proliferation and differentiation, it is necessary to understand the cellular embryonic development, signalling and cell markers.

Human and mouse lung development is divided into different morphological stages: embryonic, pseudoglandular, canalicular, saccular and alveolar stages. In both mouse and human, lung epithelium is derived from the endoderm whereas the mesenchyme comes from the mesoderm (Nikolic et al., 2018).

The embryonic phase of human lung development occurs approximately 4-7 weeks post-conception and around embryonic day 9.5 in mice (Figure 1.3a). At this stage, the primary left and right lung buds appear from the foregut endoderm and are surrounded by a mesoderm, a vascular plexus. The endoderm undergoes branching to build the overall airway structure of the lung by the end of week 5. Wnt signalling is an essential regulator of early lung endoderm specification and development (Hogan et al., 2014). The pseudoglandular phase occurs from approximately 5 to 17 weeks post-conception in humans and around embryonic days 12 to 15 in mice. During this period, the lung continues to grow and branch creating epithelial tips (Figure 1.3a). The mesoderm layer begins to differentiate to cartilage, smooth muscle, and mucous glands. At the same time as epithelial branching, there is blood vessel development from the vascular plexus. At the end of this stage, the complete airway tree in the human and mouse has been built (Hogan et al., 2014, Nikolic et al., 2018).

In these early stages of embryonic development, thyroid nuclear factor NK2 homeobox (*Nkx2-1*) expression is ubiquitous among all lung endodermal cells. *Nkx2-1* represents the initial factor to be activated within the embryonic foregut endoderm, which serves as a precursor to the emerging lung. NKX2-1 expression is maintained in adult epithelial lung cells. Furthermore, it is crucial for lung development, as demonstrated by the absence of lung formation in knockout mice lacking NKX2-1 (Parviz Minoo and Pablo Bringas, 1999).

As the pseudoglandular phase progresses, differential gene expression patterns appear in the endoderm. Firstly, there is a proximal-distal gene expression pattern by SRY-box transcription factor 2 (SOX2) and SRY-box transcription factor 9 (SOX9) transcription factors. SOX2 is expressed in proximal epithelial cells in both mouse and human whereas SOX9 is expressed in the mouse distal epithelial tips cells. Co-expression of SOX2 and SOX9 transcription factors defined human distal epithelial cells (Figure 1.3b) (Chang et al., 2013, Danopoulos et al., 2019). Notch dependent differentiation of proximal cells generates cell precursors of all cell types in the intrapulmonary bronchi and bronchioles, including NE, club and ciliated cells (Morimoto et al., 2012). Distal cells eventually differentiate into alveolar cell types, and it has been suggested that this differentiation is sonic hedgehog (SHH) dependent (Hogan et al., 2014).

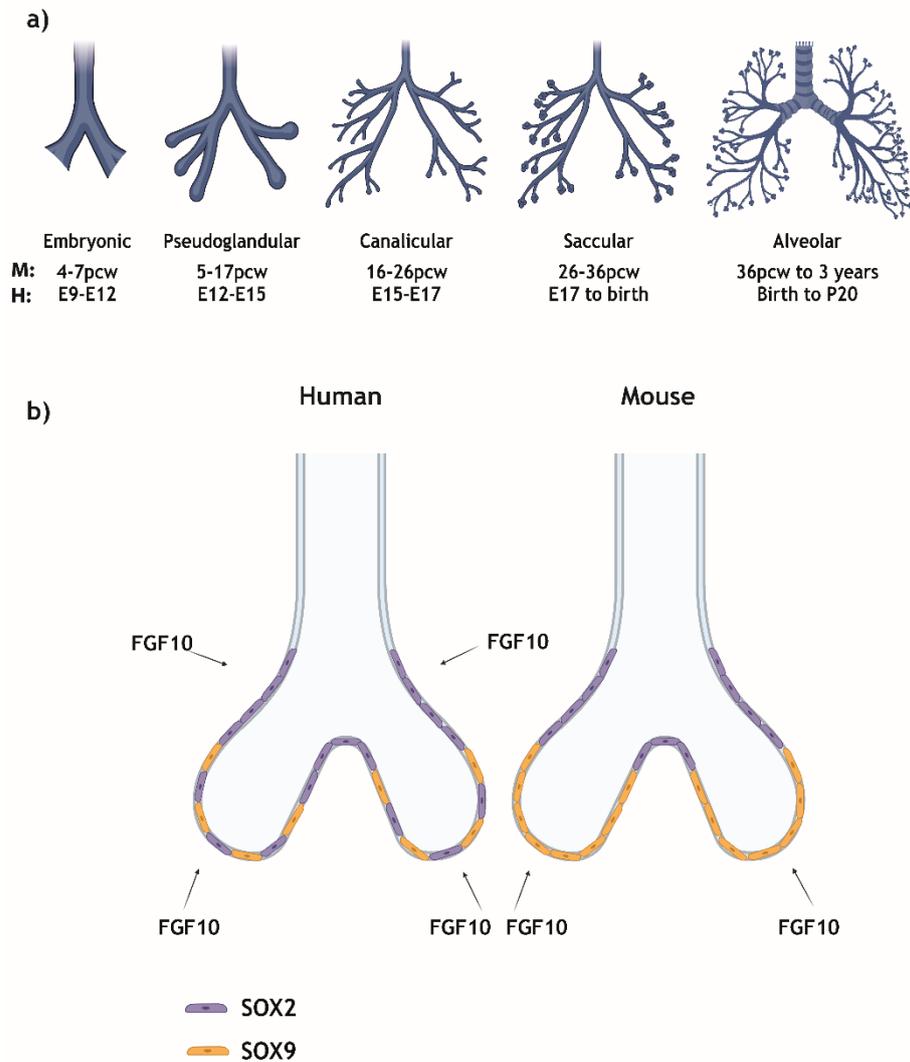
The canalicular phase of lung development occurs for 16-26 weeks post-conception in human and embryonic day 15-17 in mice (Figure 1.3a). During this stage, continuous branching occurs, existing airways continue to increase in size and the most distal epithelial tips widen into the airspaces in order to form the future alveolar regions, which are surrounded by mesenchyme thins, and the capillary networks grow closer to the distal epithelial airspaces. This is the first step into morphological alveolar epithelial cell differentiation (Hogan et al., 2014, Nikolic et al., 2018).

The saccular stage occurs between 24- and 38-weeks post-conception in humans and following embryonic day 15 in mice (Figure 1.3a). At this point, the branching stops and the distal airspaces keep increasing in size until they are surrounded by a capillary system. At this point alveolar epithelial differentiation increases and cells in the distal part of the airspaces begin to express specific genes for the two epithelial cell types present in the alveoli: AT1 cells, which express *Hopx*, and AT2 cells, which express surfactant proteins such as *SPC*. Moreover, alveoli development continuous after birth, as their number and surface area increase dramatically postnatally for weeks in mice and months in humans (Nikolic et al., 2018).

Finally, the alveolar stage is when the lung airspaces are subdivided into alveoli. At this stage, the area for gas exchange increases. At the same time, the

capillary networks fuse into a single system making every gas exchange surface surrounded by capillaries (Hogan et al., 2014).

In these last stages, fibroblast growth factor (*FGF*) genes have been identified as regulators for lung development. *FGF10* is described to be the driver of lung branching morphogenesis in mouse. The cells surrounding the distal epithelial tips produce FGF10 and function as a critical signalling population that drive outgrowth of the distal buds and branching morphogenesis. *Fgf10* is expressed in human lungs during the pseudoglandular and the canicular phases whereas in mouse FGF10 expression is detectable during the pseudoglandular stage onwards. The spatial distribution of FGF10 differs between mouse and human. In mouse lung FGF10 is highly expressed in the distal mesenchyme adjacent to the branching tip epithelial cells and the mesenchyme cells surrounding the airways that differentiate to smooth muscle cells and alveolar lipofibroblasts, whereas FGF10 in human is barely expressed in smooth muscle cells and it is dispersed throughout in the lung mesenchyme (Hogan et al., 2014, Danopoulos et al., 2019).



**Figure 1.3. Stages of human and mouse lung embryony development and the differential expression of SOX2, SOX9, and FGF10 during human and mouse lung in the pseudoglandular stage.** (a) Figure adapted from (Nikolic et al., 2018). Schematics representation of general lung morphology across the five different stages of human and mouse lung development: embryonic, pseudoglandular, canalicular, saccular and alveolar. For each stage, the developmental period is indicated, for human in post-conception weeks (pcw) and for mouse in embryonic days (E) and postnatal days (P). (b) Figure adapted from (Danopoulos et al., 2019). SOX2/SOX9 double positive progenitor cells are present in the distal epithelial tips of the human lung only, and it is predicted that they drive its elongation, whereas only SOX9 positive cells are present in the distal tips of the mouse. FGF10 is highly expressed in the mesenchyme adjacent to the epithelial surrounding the airways in mouse lungs, while in humans FGF10 positive cells are found dispersed throughout the mesenchyme with no accumulation in the distal tips. Created with BioRender.com.

## 1.3 Lung cancer: current state

### 1.3.1 Epidemiology and risk factors

Globally in 2020, lung cancer had an estimated of 2.2 million new cases and over 1.7 million related deaths, making it the second most frequently diagnosed cancer type and the largest contributor to cancer-related deaths worldwide. In the United Kingdom is the fourth most prevalent cancer type, with nearly 30,000 new cases diagnosed in 2020 and the leading cause of cancer-related deaths with nearly 18,000 related deaths in 2020 for both men and women (WHO).

To explain these high rates of incidence and mortality, it is important to understand that lung cancer has multiple risk factors. Most importantly, lung cancer mortality is highly related to smoking patterns. When smoking rates increase, lung cancer incidence and mortality rise in the following decades (Barta et al., 2019, Ahmad, 2016). This causality was first described in 1964 (Brawley et al., 2014) either with cigarettes, e-cigarettes or second-hand smoking. There are at least 55 described substances in a cigarette which are carcinogenic. Their presence leads to the formation of DNA adducts, gene methylation, gene punctual mutations, gene amplification or deletion, or whole chromosome gains or losses (Hecht, 1999). Smoking is the biggest cause of lung cancer in the UK, and 7 out of 10 diagnosed cases are caused by direct or passive smoking (Brown et al., 2018).

A second risk factor is biomass fuels, including wood, crop residues, dung and coal, which are used around the world for home cooking and heating. Indoor emissions in these households contain high concentrations of polycyclic aromatic hydrocarbons, benzene, and other carcinogenic compounds (Barta et al., 2019). Pre-existing pulmonary health conditions are related to lung cancer incidence and mortality. For instance, the development of chronic obstructive pulmonary disease, idiopathic pulmonary fibrosis, lower body mass index or emphysema (de Torres et al., 2011). Exposure to certain chemicals in the workplace also represent an important risk factor for lung cancer. For example, exposure to asbestos is one of the most well-recognised occupational causes of mesothelioma as it results in an oxidative damage and subsequent DNA deletions, somatic gene alterations, and enhanced delivery of tobacco carcinogens to the airway

epithelium (Kelsey, 2002, Klebe et al., 2019). Another example is silica which is used in construction and diesel engine exhaust fume. Regular environmental exposure to air pollution causes around 10% of lung cancer cases in the UK (Brown et al., 2018). Certain genetic factors like polymorphisms in enzymes such as cytochrome p450 enzymes and DNA repair genes are also lung cancer risk factors (Barta et al., 2019).

### **1.3.2 Lung cancer staging and histological patterns**

Lung cancer tumour can be classified in different stages depending on the level of invasive and dissemination. This classification aids patient prognostication and therapeutic decision making. The tumour-node-metastasis (TNM) system is the most commonly used staging system. Tumour (T) describes the size of the tumour (ranging from T1 to T4), node (N) describes whether the tumour has spread to the lymph nodes, ranging from N0 (lymph nodes free), N1, N2 and N3 (range of cancer cell lymph node invasion). Finally, metastasis (M) which describes if the tumour has spread to other parts of the body. M0 is the absence of metastasis and M1 is the presence of a secondary tumour (Lababede and Meziane, 2018).

In lung cancer, four stages summarise the TNM system. Stage I describes a tumour up to 3 cm of diameter that has not spread to the pleura, other airways, nearby lymph nodes or other distant parts in the body. Stage II describes one or more tumours up to 7 cm which may have invaded the pleura, but which have not invaded the lymph nodes or other organs. These two stages are referred to as early-stage tumours. Stage III is defined by tumour of 7 cm or bigger which has spread to the diaphragm, the heart, the blood vessels, the trachea, or other structures including the lymph nodes (in the lung, neck and bronchus) but has not spread to other parts of the body. This stage is referred as locally advanced cancer. Stage IV tumours are lesions which have grown into the nearby structures, lymph nodes or organs from outside the chest. This stage is referred as metastatic advanced lung cancer (Rami-Porta et al., 2018).

Lung adenocarcinomas have five different histological patterns, and they are classified into three prognostic groups: low grade (lepidic predominant tumours), intermediate grade (acinar or papillary predominant tumours) and high grade

(solid or micropapillary predominant tumours). Given that these patterns can be found together in the same tumour, each tumour grade can be calculated depending on the combination of histologic patterns. For instance, grade 1 tumours are well differentiated, and they have predominantly lepidic patterns with less than a 20% of a high-grade pattern. Grade 2 tumours are moderately differentiated, and they have predominantly acinar or papillary pattern with less than a 20% of a high-grade pattern. Finally, grade 3 tumours are poorly differentiated, and they are characterised by the presence of any pattern with more than 20% of a high-grade pattern (Nicholson et al., 2022).

### **1.3.3 Lung cancer symptoms and diagnostics**

Lung cancer patients show few symptoms in early stages and in later stages, patient's symptoms can be confused with other medical conditions. The main symptoms are persistent cough or hemoptysis, breathlessness, chest pain, recurrent chest infection, loss of appetite, fatigue, and weight loss (2013).

Lung cancer is commonly diagnosed in advanced stages. The lack of sensitive and specific biomarkers for early diagnosis (Nooreldeen and Bach, 2021) makes it difficult to detect early stages tumours. In Scotland more than 50% of the cases are diagnosed in advanced stages, 46% of cases are diagnosed in stage IV and 21.6% of patients are diagnosed in stage III. One-year survival drops from nearly 90% in lung tumours diagnosed in stage I to around 50% when the detected tumour is in stage III and to 20% in stage IV. Moreover, five-year survival drops from 60% in patients diagnosed in stage I to 10% on stage III and 5% in stage IV (Scotland, 2023).

There are different techniques to diagnose lung cancer. Traditionally, chest radiography (x-ray) has been used. However, radiography can only detect tumours when they are 1 cm in diameter. Tumours of that size tend to be in advanced stages. In contrast, computerised tomography (CT) scans can acquire data in a shorter scanning time than radiography, a lower radiation exposure, and are effective in detecting small, 1-5 mm lung nodules. Nevertheless, cost and accessibility prevent its wide use around the UK (Nooreldeen and Bach, 2021). Combined positron emission tomography (PET)-CT scans can provide additional information. For instance, PET-CT can better determine tumour

location, tumour stage, aggressiveness, treatment strategy and response monitoring (Manafi-Farid et al., 2021).

The most effective way to confirm a lung tumour is through a tissue biopsy. Samples can be obtained through a percutaneous lung biopsy, surgical biopsy or through an ultrasound guided neck lymph biopsy. These biopsies serve a dual purpose, not only for cancer confirmation but also for tumour characterisation through histological classification, which aids in prognostic and defines therapeutic strategies (Nooreldeen and Bach, 2021).

Currently, there is an effort to obtain this information with less invasive methods. Circulating tumour DNA (ctDNA) is released by tumour cells during apoptosis, necrosis, or secretion. ctDNA can be found in the plasma which allows analysis via a routine blood sample. Free circulating DNA is present in healthy patients due to processes like inflammation or apoptosis, but the amount of free circulating DNA is significantly higher in cancer patients. This allows for ctDNA amplification with a polymerase chain reaction and the downstream analysis of the genetic alterations include point mutations, methylation patterns, chromosomal rearrangements, structural rearrangements, and copy number variations. Circulating tumour cells (CTCs) are a novel predictor method for cancer development and progression. CTCs can be identified and isolated based on the expression of epithelial cell adhesion molecule (EpCAM) or the expression of markers of epithelial-mesenchymal transition (Nooreldeen and Bach, 2021). The presence of CTCs is associated with a lower survival in lung cancer patients and can be an indicator of a propensity for cell migration and metastasis.

#### **1.3.4 Lung cancer current treatments**

Lung cancer has a very limited range of treatments available. As mentioned above, less than 50% of the cases are diagnosed in early stages. However, tumours identified in early stages can often have complete surgical resection. More advanced tumours can still be surgically resected if small and at N0 stage. In more advanced or complex cases, lung cancer can be treated with systemic chemotherapy or adjuvant chemotherapy (after surgery). However, most lung tumours are either intrinsically resistant to chemotherapy or they develop

acquired resistance over time. Therefore, novel therapeutic strategies for inoperable cases are being developed (Ahmad, 2016).

Other treatment options include radiotherapy, which applies radioactive beams directly to the tumour either after surgery, or following incomplete surgical resection, or chemotherapy (Minniti et al., 2012).

Finally, targeted therapy in non-small cell lung cancer (NSCLC) is mainly focused on epidermal growth factor receptor (EGFR), which is mutated in 15% of NSCLC cases. Inhibition of EGFR leads to an increase of pro-apoptotic pathway and results in cell death through the activation of the intrinsic mitochondrial apoptotic pathway. There are several studies in phase III of EGFR tyrosine kinase inhibitors (TKI) that show significant improvement in response rates (Costa et al., 2007). Moreover, 5% of NSCLC patients contain translocations in anaplastic lymphoma kinase (*ALK*) gene and 1% of them have *ROS1* translocations. Several TKI developed have later found to have activity in patients with *ALK* and *ROS1* translocations. Other targetable mutations include *BRAF*, *KRAS* or mitogen activated protein kinase (*MAPK/ERK*) are still under investigation (Duma et al., 2019, Ahmad, 2016)

## 1.4 Lung cancer development

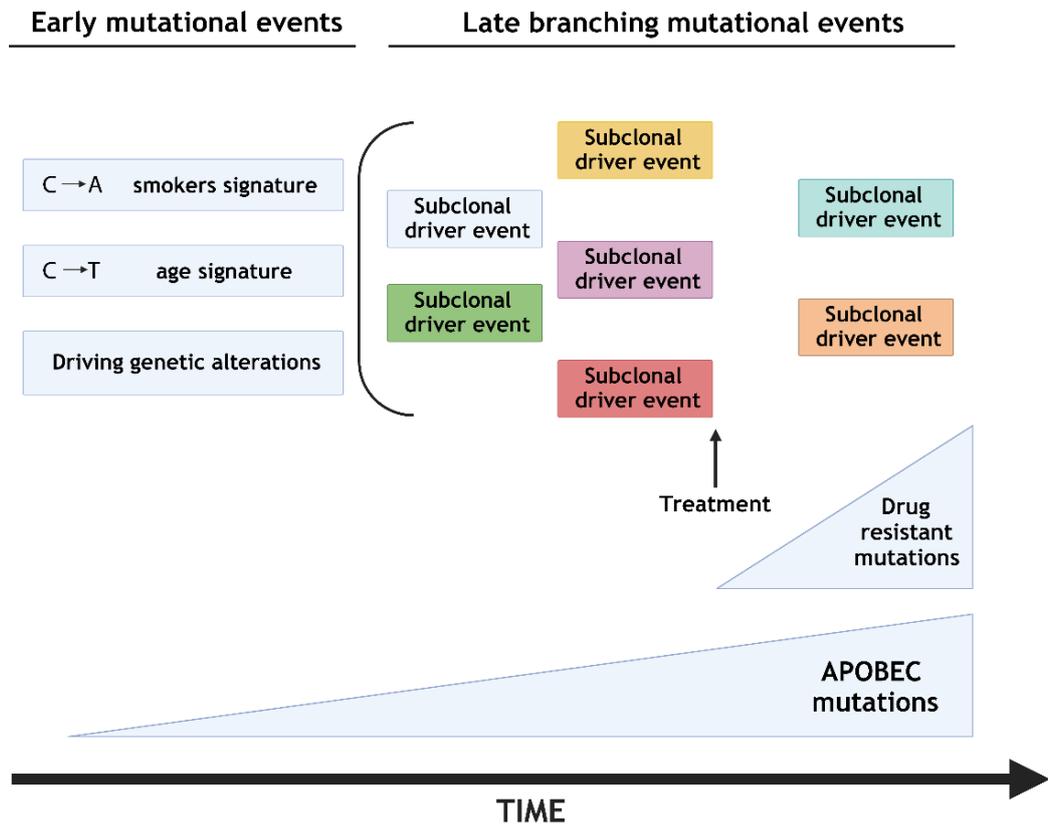
As explained above, tobacco is the main cause of lung cancer, and these cases present a high somatic mutations rate, regardless of the histological subtype, being 8 to 10 mutations per 1 million base pairs. Tumours from non-smokers patients have a much lower mutation rate being 0.8 to 1 mutation per million base pairs (Cancer Genome Atlas Research, 2014, Govindan et al., 2012).

Genomic analysis has revealed that lung tumour has clonal evolution where a lung tumour consists of different subclones that share a common ancestor but have differing genomic alterations occurring later in tumour development. These subclones might be mixed in the tumour or regionally separated and form the basis for intratumour heterogeneity. The evolutionary trajectory of these subclones is characterised by different genetic driver events (Yates and Campbell, 2012). The most common signature in lung tumours from tobacco smokers is the transversions of cytosines to adenines. Non-smokers may present

transversions from cytosine to thymine in CpG sites (DNA regions where a cytosine is followed by a guanine), a mutagenic signature related to advanced patient age. Simultaneously in early tumour evolution, there is evidence of somatic mutational and copy-number heterogeneity and of genome-doubling events which are present in every subclone population. These genetic alterations affect the driver genetic alterations in *EGFR*, *KRAS*, *BRAF* and tumour protein (*TP53*) genes (Figure 1.4) (Swanton and Govindan, 2016).

Subclonal driver events may occur after these initial genomic mutations which will contribute to an increase in the tumour burden. For instance, transversions of cytosines to adenines are less present in subclonal cells suggesting that an alternative mutational process appears later in the tumour evolution of smoking exposure. It has been identified that subclonal mutations are enriched with transversions from cytosines to thymine and guanine instead. This mutational process is linked to apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC), suggesting that this enzyme contributes to the clonal evolution of the tumour (Figure 1.4). This genome instability combined with age related mutations, or the mutations introduced due to cytotoxic therapies will result in alterations in pathways like phosphatidylinositol-4,5-bisphosphate 3-kinase (*PI3K*) and transforming growth factor beta (*TGFβ*). As a result of all these alterations and the genetic drifts, multiple subclones coexist within the tumour which may compete or cooperate for space and nutrients (Swanton and Govindan, 2016).

Consequently, in these subclonal populations, polyclonal drug resistance may develop. For instance, patients with somatic activation of *EGFR* are treated with *EGFR* TKI. However, a secondary mutation in *EGFR* gene (T790M) can contribute to drug resistance within 6 and 12 months after the first treatment (Figure 1.4) (Ma et al., 2011).



**Figure 1.4. Evolutionary development of lung cancer.** Adapted from (Swanton and Govindan, 2016). Early genetic alterations occur in individuals mainly characterised by transversions from cytosine to adenosine and thymine and driving alterations in multiple oncogenes. Subclonal driver events may occur later in time creating in different cells within the tumours creating the branches of the evolutionary tree. Later in tumour evolution, APOBEC mutational signature is enriched in branching subclonal cell populations, suggesting a branched evolution of lung cancer. Some subclonal populations are removed from the tumour after treatments but other acquired more subclonal mutations which will allow the tumour to become resistant to drug treatment. Created with BioRender.com.

## 1.5 Lung cancer heterogeneity

Lung cancer is a heterogeneous disease at multiple levels. This variability occurs within individual lung tumours (intratumoral heterogeneity) or between different lung tumours (intertumoral heterogeneity) (Visvader, 2011). In the case of intertumoral heterogeneity, lung cancer has a wide number of histological subtypes, genetic mutations or epigenetic alterations that are able to drive lung cancer through a range of cells-of-origin creating a heterogenic tumour microenvironment (Chen et al., 2014). There are two main mechanisms to explain intertumoral cancer heterogeneity. The first is different genetic or epigenetic alterations which result in different tumour phenotypes in different

cancer patients, and the second is due to tumours arising from different cell types in the lung. These two mechanisms are not exclusive, and they can both contribute to determine the resultant tumour phenotype (Visvader, 2011).

### 1.5.1 Lung cancer heterogeneity: genomic alterations

Heterogeneous genetic alterations are found in different cancer subtypes and can determine not only the histological subtype of lung cancer but also the tumour microenvironment composition. Most types of lung cancer share alterations in specific molecular pathways, for example alterations in the *RAS-RAF-MEK* pathway exist in 76% of lung cancer cases. Lung tumours also share deregulated RNA splicing pathways, activation of pathways related with the lung embryonic development, such as *NOTCH* alterations, altered oxidative stress response, cell cycle dysregulation, and epigenetic alterations (Swanton and Govindan, 2016). From a histological point of view, there is a wide range of phenotypes in lung cancer (Nicholson et al., 2022) and each histological subtype is driven by specific genetic alterations.

Lung cancer is mainly classified in NSCLC, which accounts for 85% of cases and SCLC, which represents the 15% of the total lung cancer cases. SCLC is mostly found in smokers (Ahmad, 2016). Genomic analysis of NSCLC tumours shows branching evolution in which cells within a tumour shares a common ancestor but have several different genetically altered subclones, acquired later during tumour evolution. Upregulation of oncogenic signalling through *RAS and BRAF and EGFR* mutations is shared among NSCLC tumours (Swanton and Govindan, 2016). NSCLC is further classified as adenocarcinomas (ADC), the most common subtype representing the 78% of cases, predominant type in non-smokers and arising in the distal airways (Ahmad, 2016, Swanton and Govindan, 2016), and the squamous cell carcinoma (SCC), arising in proximal airways (Thai et al., 2021) and being the predominant NSCLC type found in smokers. ADC is characterised by mutations and amplifications of other oncogenic pathways and receptors such as *ERBB2, MET*, fibroblast growth factor receptor 1 and 2 (*FGFR1* and *FGFR2*), as well as fusions oncogenes such as *ROS1* and *ALK*. ADC also have frequent mutations or deletion in tumour suppressor genes, the most common one being *TP53* (Chen et al., 2014). SCC tumours are characterised by the activation of cell differentiation pathway genes, including amplifications in *SOX2*

and *TP63*, as well as mutations in the NOTCH pathway. Additionally, these tumours exhibit disrupted cell cycle regulation due to mutations, overexpression, and deletions in genes like *P53*, *MYC*, and cyclin dependent kinase inhibitor (*CDKN2A*). Furthermore, SCC is characterised by alterations in genes associated with the PI3K pathway, such as mutations in *PTEN*, and irregularities in epigenetic regulators (Lau et al., 2022).

NSCLC histological subtypes can be classified with distinct biomarkers and in distinguished patterns. ADCs present with glandular histology and biomarkers consistent with a distal lung location such as thyroid transcription factor 1 (*TTF1*) and keratine-7 (*KT7*). ADCs can be divided into different subtypes depending on their morphologic growth patterns. These are acinar, papillary, lepidic, micropapillary and solid (Nicholson et al., 2022). Moreover, each pattern determines the tumour grade and patient prognosis. For instance, micropapillary and solid patterns are high risk, acinar and papillary are intermediate risk and lepidic are a low-risk (Nicholson et al., 2022). Whereas SCC is histologically more like the pseudostratified columnar epithelium of the trachea and the proximal airways, they can be identified by the expression of transcription factor *SOX2* and *TP63* (Chen et al., 2014). SCCs can also be divided into morphological patterns, including keratinizing, non-keratinizing and basaloid squamous cell carcinoma (Nicholson et al., 2022). Finally, SCLC lesions are characterised by deletions or mutations in retinoblastoma (*RB*), *p107*, *p130*, *PTEN* and *TP53*, as well as mutations in NOTCH receptors and the chromatin regulator *CREBBP*. In SCLC, there are some common mutations in oncogenes like amplifications in *FGFR1* and members of the *MYC* family (Rudin et al., 2021, Swanton and Govindan, 2016)

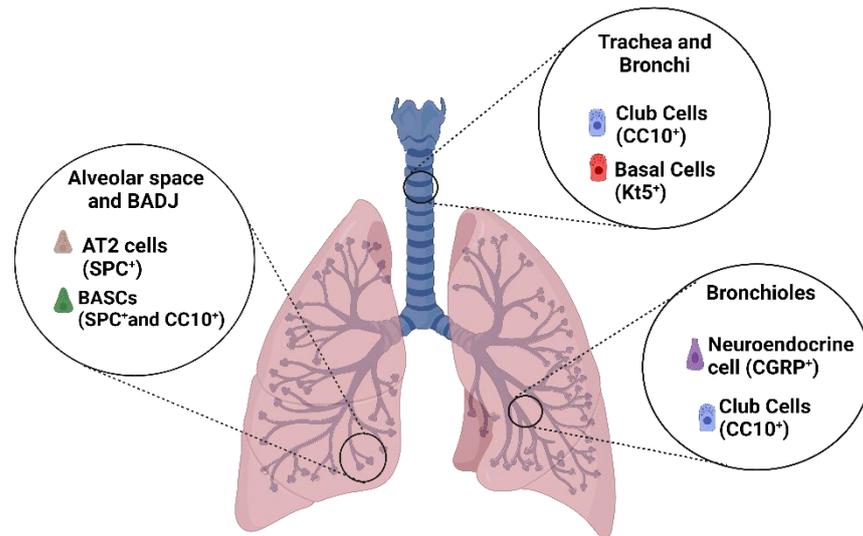
Genetic alterations can also drive differences in the tumour microenvironment. The tumour microenvironment describes all the cancer cell interactions within the extracellular matrix, mesenchymal cells like immune cells or fibroblasts and the vasculature surrounding the tumours (Chen et al., 2014). It is thought that lung tumours with different genetic backgrounds differ in the composition of their tumour microenvironment. For example, interleukin-6 regulates several immune and inflammatory responses, but it also plays a role in lung tumour progression and metastasis (Kishimoto, 2005). Lung tumours with *KRAS* (Ancrile et al., 2007) and *EGFR* (Gao et al., 2007) mutations have high levels of

interleukin-6 which drives angiogenesis and tumour growth and is a potential therapeutic target. Moreover, different genetic alterations in lung tumours drive a different immune cell composition. For instance, SCC with loss of serine threonine kinase 11 (*LKB1*) and *PTEN* genes shows an increase in the number of tumour-associated neutrophils (Xu et al., 2014) while adenocarcinomas driven by mutations or amplifications in *EGFR*, show a decrease in cytotoxic T cells due to the upregulation of PD-1, PD-L1 and CTL antigen-4 proteins. Thus, EGFR inhibitors downregulate the expression of PD-1 and PD-L1 in lung tumours (Akabay et al., 2013).

### **1.5.2 Lung cancer heterogeneity: cell-of-origin**

Recent studies suggest that many solid tumours are hierarchically organised as they are sustained by a self-renewing cell population. These tissue stem cells have been identified as the cells-of-origin for different organs, such as, lung, colon or brain, due to their proliferative capacity and their longevity, which allows enough time for the accumulation of genetic and epigenetic mutation for tumour initiation. However, tissue cells with limited self-renewal capacity and stem- cell-like properties can also start proliferating and generate tumours in response to genetic insults (Visvader, 2011). In the lung epithelium, different regions contain different populations of adult stem cells with the ability to self-renew after an injury. These cells not only function as tissue stem cells but they also express genes related with specialised functions such as surfactant protein synthesis or glycoprotein production and transport (Hogan et al., 2014).

In lung cancer, there five purported cell types that can act as cell-of-origin for a tumour, and these are distributed along the airways (Blanpain and Fuchs, 2014, Fuhrmann et al., 2014). In the alveolar space there are two potential cells-of-origin. Basal and club cells in the trachea and proximal airways, they express *KT5* and *CC10* specifically. NE cells in the bronchioles, which expressed *CGRP* and BASCs, expressing *SPC* and *CC10*, and AT2 cells, which express *SPC*, in the BADJ and alveolar space, respectively.



**Figure 1.5. Schematic representation of the cells-of-origin of lung cancer in the murine lung epithelium and their specifically expressed genes.** Murine trachea epithelium contains basal cells, which specifically express keratin-5; and club cells which express CC10. Bronchioles and proximal airways contain club cells again, and clusters of neuroendocrine cells, which express CGRP. Finally, in alveolar space and its intersection with the distal airways AT2 cells and BASCs can be found respectively. AT2 cells express SPC and BASCs express both CC10 and SPC genes. Created with BioRender.com.

### 1.5.2.1 The effect of cell-of-origin on lung cancer histology

Similarly to the relationship between histological subtypes and genetic alterations, it is possible that the biology of different cells-of-origin is what drives the phenotypic differences in NSCLCs. Differences may relate to specific gene expression patterns in each one of these cells, differences in routes of differentiation or differences in the local cellular microenvironment (Chen et al., 2014). To support that theory, it has recently been published that fundamental aspects of tumour biology are regulated by gene programs which are shared between tumours and non-malignant epithelial cells of the same tissue, suggesting that these gene expression programs are present before the tumorigenic process takes place, reflecting the biology of the tumour cell-of-origin (Gavish et al., 2023).

Nonetheless, there is uncertainty regarding whether multiple cell types possess equal potential to give rise to a particular type of lung tumour, or if it is exclusively a specific cell type within the appropriate microenvironment that must undergo oncogenic alterations to initiate the development of a specific

lung tumour (Chen et al., 2014). For instance, squamous cell carcinomas were first described to arise from basal cells (Davidson et al., 2013) but more recent studies showed that BASCs and club cells are likely cells-of-origin of SCC following the deletion of *LKB1* (Zhang et al., 2017). ADC lesions were first described to arise from club and AT2 cells in *KRAS* mutated and combined *KRAS* mutated and *TP53* deleted tumours (Sutherland et al., 2014). However, BASCs have more recently been proposed as cells-of-origin of lung ADC (Kim et al., 2005).

Finally, SCLC have been described to arise mainly from NE cells although *TP53* and RB inactivation in mouse models in SPC-expressing cells, including AT2 cells and BASCs can also lead to SCLC lesions although with lesser efficiency than in NE cells (Sutherland et al., 2011).

#### **1.5.2.2 Methods to study the effect of cell-of-origin on lung heterogeneity**

These complex scenarios can be partially explained by the fact that cell-of-origin research is technically difficult. In only a few cases, the tumour-initiating cells can be detected in grown tumour. For instance, basal cells have been suggested to be cells-of-origin of SCC as they normally arise in the proximal airways and these tumours express *TP63* and *SOX2*, which is a common marker of basal cells (Chen et al., 2014). However, in most cases the cell surface marker staining patterns of the resultant tumour might not match the initiating cell type markers (Chen et al., 2014). Therefore, different strategies are needed for the study of the cell-of-origin and its effects in lung tumour heterogeneity. The most-used method so far is the use of in vivo models. There are two main approaches when it comes to the use of genetically engineered mouse models (GEMMs). The first approach involves an ex vivo cell type genetic manipulation and subsequently transplantation into mice to assess their behaviour to tumour initiation in both human and mice cell types and tissues (Visvader, 2011). Secondly, the use of transgenic or conditionally targeted gene technologies, requiring cell-specific promoters, which drives the expression of a gene or of a *Cre* recombinase in a cell type specific manner. By targeting one cell type only, it is expected to obtain tumours that recapitulate human tumour phenotypes more accurately (Visvader, 2011). The main limitation for this approach is that there is limited knowledge of cell-lineage specific promoters. However, several

AdenoCre vectors have been developed where the Cre recombinase is under the control of cell-type specific promoters (Ferone et al., 2020).

Lung cancer cell-of-origin research is mainly based on the second approach using transgenic mice explained above. Several cell-type specific adenovirus vectors have been developed by Dr Anton Bern and they are produced and distributed by the Viral Vector Core Facility from the University of Iowa. In these vectors either the Cre recombinase or a reporter gene is under the control of *SPC*, *CC10*, *KT5* and *CGRP* promoters. These promoters are expressed in AT2, club, basal and NE cells respectively. *SPC* and *CC10* vectors are also expressed in BASCs (Chen et al., 2014). These viruses can be used to induce a variety of GEMMs, both by intranasal inhalation and intratracheal intubation (DuPage et al., 2009), which recapitulate several genetic alterations found in human lung cancer. For NSCLC models, the first GEMM was LSL-KRAS<sup>G12D</sup> described in 2001, using a non-cell specific Adeno-LoxP-CRE system in mouse models resulting in adenocarcinoma tumours (Jackson et al., 2001). RAS is known to activate several pathways related to cell proliferation such as MAPK and PI3K pathways, and pathways related to cell shape and cell migration (Kwon and Berns, 2013). The LSL-KRAS<sup>G12D</sup> GEMM has also been combined with deletions in suppressor genes. For instance, expression of mutant KRAS and the deletion of *Lkb1*, a combination often found in human tumours, in mice intranasally induced by Ad5-CMV-Cre virus shows an accelerated tumour development and with more malignant phenotype characteristics than LSL-KRAS<sup>G12D</sup> only tumours. This combination normally results in SCC lesions (Ji et al., 2007). Moreover, KRAS<sup>G12D</sup> mutations can be combined with P53 inactivation, in mice induced with lentivirus-Cre CMV and together with LSL-KRAS<sup>G12D</sup> GEMMs are the most common models for studying lung ADC (Winslow et al., 2011). In this case, tumours were demonstrated to be highly metastatic, expressing similar markers to NSCLC tumours (Kwon and Berns, 2013). Moreover, this model shows high intratumoral and intertumoral heterogeneity due to mutations, amplifications and deletions in oncogenes and suppressor genes beyond those engineered during murine tumorigenesis. Those spontaneously altered genes are also among the most prevalent in human lesions (Chung et al., 2017). GEMMs with other driving oncogenic mutations have also been developed. For instance, *Egfr* mutations mouse models have been widely used for drug resistance and drug combination studies, as different *EGFR*

mutations drive different genomic alterations seen in human tumours (Kwon and Berns, 2013). Finally, *Alk1* fusion gene mouse model expresses a constitutive active tyrosine kinase receptor which is not normally expressed in healthy lung tissue and is present in 5% of NSCLC (Manabu Sodaa and Hidenori Harutaa, 2008).

SCLC GEMMs have also been developed, mainly with retinoblastoma and *Tp53* suppressor mutations using *Rb<sup>fl/fl</sup>*, *Tp53<sup>fl/fl</sup>* mouse model induced with non-cell specific lentiviral or adenoviral vectors (Ralph Meuwissen, 2003). This tumour highly resembles human SCLC phenotype and their metastasis sites, although the tumorigenic process is slow, and it can take up to 9 months to develop tumours. These tumours acquire human SCLC mutations such as amplification of *Myc* (Dooley et al., 2011). The combined loss of *Rb* and *Tp53* with *p130* significantly accelerates the tumour development (Schaffer et al., 2010). As a result of these genetic alterations further mutations are found in these tumours related to PI3K resembling human tumours (Kwon and Berns, 2013).

Recently, in vitro 3D models, such as lung organoids have been developed to study the effect of the cell-of-origin on lung cancer heterogeneity. An organoid culture is defined as a 3D structure grown from organ-specific cell types that develops from pluripotent stem cells or tissue specific stem cells and self-organizes in a manner like in vivo (Lancaster and Knoblich, 2014). James Rheinwald and Howard Green described the first long-term culture of normal human cells in 1975 when they combined isolated keratinocytes with irradiated mouse fibroblasts (Green, 1975), and they are still being developed for both cancer and non-cancer research (Clevers, 2016).

Organoids can be grown from two main types of stem cells: pluripotent embryonic stem cells and induced pluripotent stem cells or organ-restricted adult stem cells. Since 1975, several organoid types have been established. From pluripotential stem cells seeded in 3D, they would differentiate into endoderm or ectoderm/neuroectoderm when cultured with or without activin A respectively. At this point different combination of growth factors will result into different organoid types: such as stomach, intestinal, lung, thyroid, liver, kidney and several structures from the central nervous system (Lancaster and Knoblich, 2014, Clevers, 2016). Finally, several regions in the body can be cultured as adult stem cell organoids like lungs, pancreas, large and small

intestine, stomach/oesophagus, fallopian tube, mammary gland, prostate liver and spleen (Clevers, 2016).

Organoids have a wide range of applications in cancer research both basic and translational, as they remain genetically stable, but they can be genetically modified, expanded long term and cryopreserved. Tumour organoids translational applications include drug development and personalised cancer treatment research and immunotherapy development (Drost and Clevers, 2018). 2D cell culture has been essential in cancer research for genetic predictions in cancer response so far but their lack of capacity to mirror tumour tissue may contribute to the high failure rate of new drugs in clinical trials. Patient-derived tumour organoids better recapitulate native tumour compared to 2D cancer cell line culture and might become the preferred model to test new cancer drugs and to this aim organoids biobanks are being created and shared with the research community to allow relevant drug screening (Clevers, 2016, Drost and Clevers, 2018).

Organoids have also been used in basic research to explore the link between infectious agents and cancer development. Several cancer types have been linked to infectious agents although the pathogen malignant transformation is not well understood. There are several studies that have succeeded in establishing organoid co-culture systems with different types of pathogens, this technology could potentially be used to study these processes. Most importantly, they have been used to interrogate mutational process underlying malignant transformation as organoid high genetic stability over time allows the researchers to study this process, especially in tumour early stages (Drost and Clevers, 2018).

As for lung tumour organoids, lineage restricted lung tumour organoids have recently been developed for a better understanding of the cell-of-origin in the resultant tumour phenotype. The main approach consists in the isolation of these stem cell like lung murine primary cells and culturing them with or without a support cell line (mesenchymal cells and endothelial cells have been widely used (Lee et al., 2014)) in an air-liquid interface method with either Matrigel or collagen, after inducing them in vitro. This system allows the researchers to test the ability of certain cell types to develop tumour but also to

understand the effects of the tumour initiating cells into the resultant tumour phenotype (Kim, 2017).

AT2 cells have been isolated and induced by KRAS activation to study the transcriptional and proteomic differences between normal epithelial progenitors and early-stages tumours to understand the effect of KRAS activation on phenotype (Dost et al., 2020). BASC lung organoids in lung cancer research have also been developed and have been used to demonstrate that lung epithelial progenitor cells (both AT2 cells and BASCs) from organoid cultures restore their native transcriptional state and retain progenitor cell function after lung reimplantation (Louie et al., 2022). Club cell organoids have also been developed and they have been used to confirm club cells as a proliferative cell population in the lung after *KRas* activation and therefore a likely cell of origin for NSCLC (Zhang et al., 2017). Club cell organoids have also been used to identify the accumulation and trajectory of club cell mutations during tumour progression (Chen et al., 2022). Finally, basal cell organoids have been built mainly to study of the trachea/proximal airways and to understand the distinct responses of the airway epithelium during injury-repair and in different disease types (Zhou et al., 2022, Cunniff et al., 2021).

## 1.6 Cancer metabolism

Metabolic alterations are considered to be a hallmark of cancer, contributing to tumour transformation and progression (Hanahan and Weinberg, 2011).

Therefore, understanding tumour metabolism is important as it can provide diagnostic, prognostic and therapeutic information.

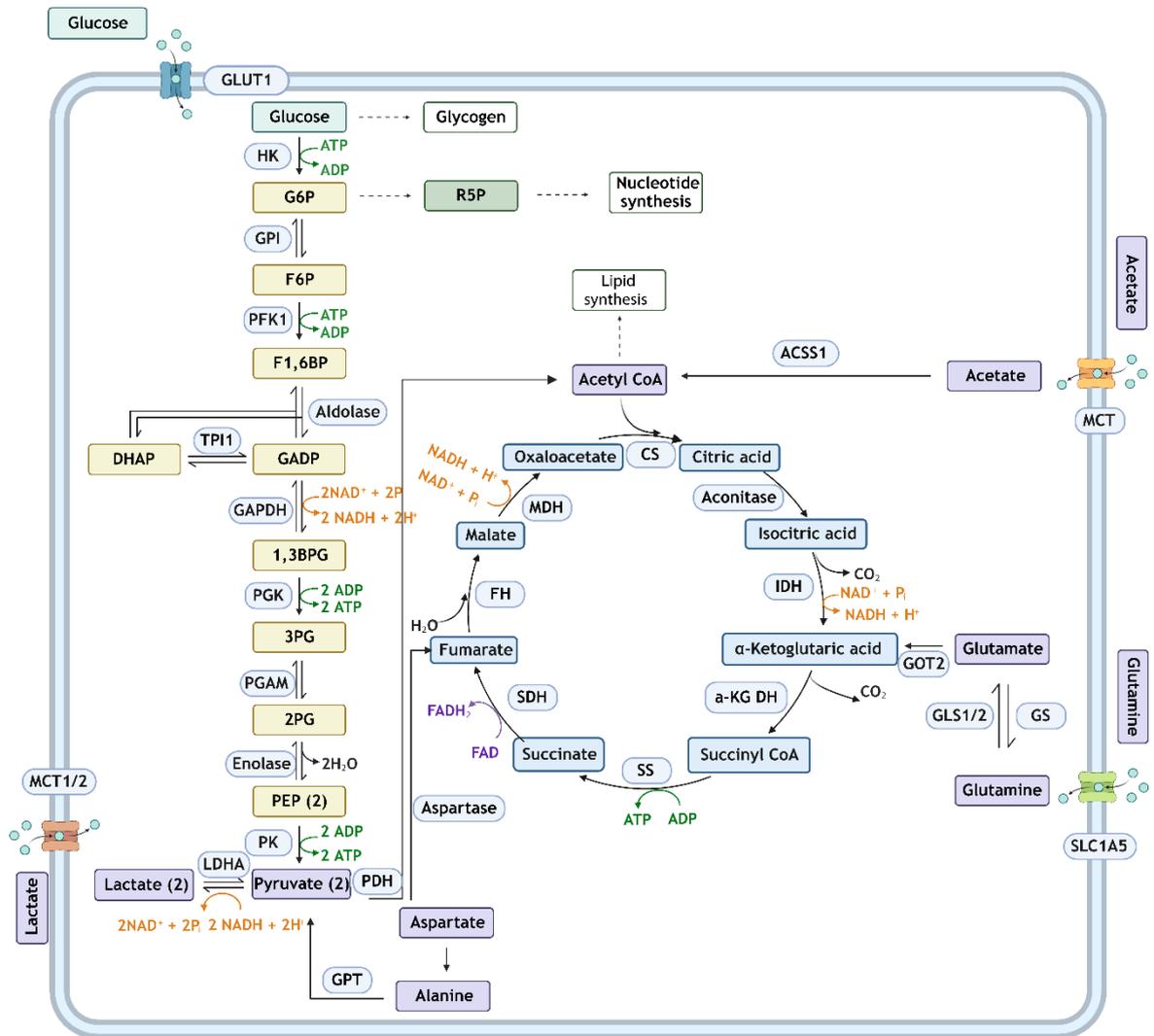
Reprogrammed metabolic pathways have a distinct functional classification. Metabolic changes can be classified as transforming, enabling or neutral activities. Transforming alterations have a direct contribution to cell transformation and blocking them will stop cancer progression. These alterations are normally caused by mutations in metabolic genes. The most common ones are: mutations in the genes encoding isocitrate dehydrogenases-1 and -2 (IDH1, IDH2); mutations in components of the succinate dehydrogenase (SDH) complex; and mutations in fumarate hydratase (FH) (Vander Heiden and DeBerardinis, 2017). Mutations in *IDH1/2* an enzyme that converts  $\alpha$ -ketoglutarate ( $\alpha$ KG) to

(D)-2-hydroxyglutarate. As a consequence, (D)-2-hydroxyglutarate accumulates and it interferes with the function of prolyl hydroxylases that target hypoxia inducible factor (HIF-1) subunits for degradation (Dang et al., 2009). SDH and FH catalyse sequential reactions in the tricarboxylic acid (TCA) cycle. Both enzymes act as tumour suppressors, and after their deletion tumours accumulate high levels of succinate and/or fumarate that interferes with several enzymes leading to cell proliferation (Vander Heiden and DeBerardinis, 2017).

Enabling alterations are altered metabolic pathways that do not contribute directly to cell transformation. Enabling alterations carry out typical metabolic functions such as energetics, maintenance redox state and generation of macromolecules. These pathways are under the regulation of oncogene mutations and tumour suppressor gene deletions such as *c-MYC*, which activates glucose metabolism; *KRAS*, which increases nutrient uptake and synthesis of macromolecules and redox homeostasis; and mammalian target of rapamycin (*mTOR*) which promotes lipid, protein and nucleotides synthesis (Vander Heiden and DeBerardinis, 2017).

The first enabling alteration in cancer metabolism described was the Warburg effect. This theory describes that cancer cell metabolise glucose in a distinct manner than normal cells. Cancer cells convert glucose into lactate even when there is sufficient oxygen to support oxidative phosphorylation in the mitochondria (Otto Warburg, 1926, Warburg, 1956). Glycolysis is a metabolic pathway that converts glucose into pyruvate in the cell cytoplasm, and as a result, 2 adenosine triphosphate (ATP) molecules are produced. The pathway is characterised by 10 chemical reactions with each is catalysed by a specific enzyme, being hexokinase (*HK*), phosphoglucose isomerase (*GPI*), phosphofructokinase (*PFK1*), aldolase, triosephosphate isomerase (*TPI*), glyceraldehyde 3 phosphate dehydrogenase (*GAPDH*), phosphoglycerate kinase (*PGK*), phosphoglycerate mutase (*PGAM*), enolase, and pyruvate kinase (*PKM*) (Figure 1.6). In normal tissues, pyruvate is converted into lactate, under reduced oxygen conditions otherwise pyruvate enters into the TCA cycle in the mitochondria, resulting in 36 ATP molecules (38 after the 2 molecules produced in glycolysis) per glucose molecule and producing CO<sub>2</sub> and H<sub>2</sub>O (Li et al., 2015), showing that TCA cycle is more energetically efficient than glycolysis (Figure 1.6). This seems contradictory for cancer cells as energy requirements for cancer

cell proliferation are better met with oxidative phosphorylation (Matthew G. Vander Heiden, 2009). But glycolysis can be an inefficient ATP resource only when there is a limitation to glucose access, which might not be the case for some tumours which are well perfused and obtain unlimited glucose from the circulation (Li et al., 2015).

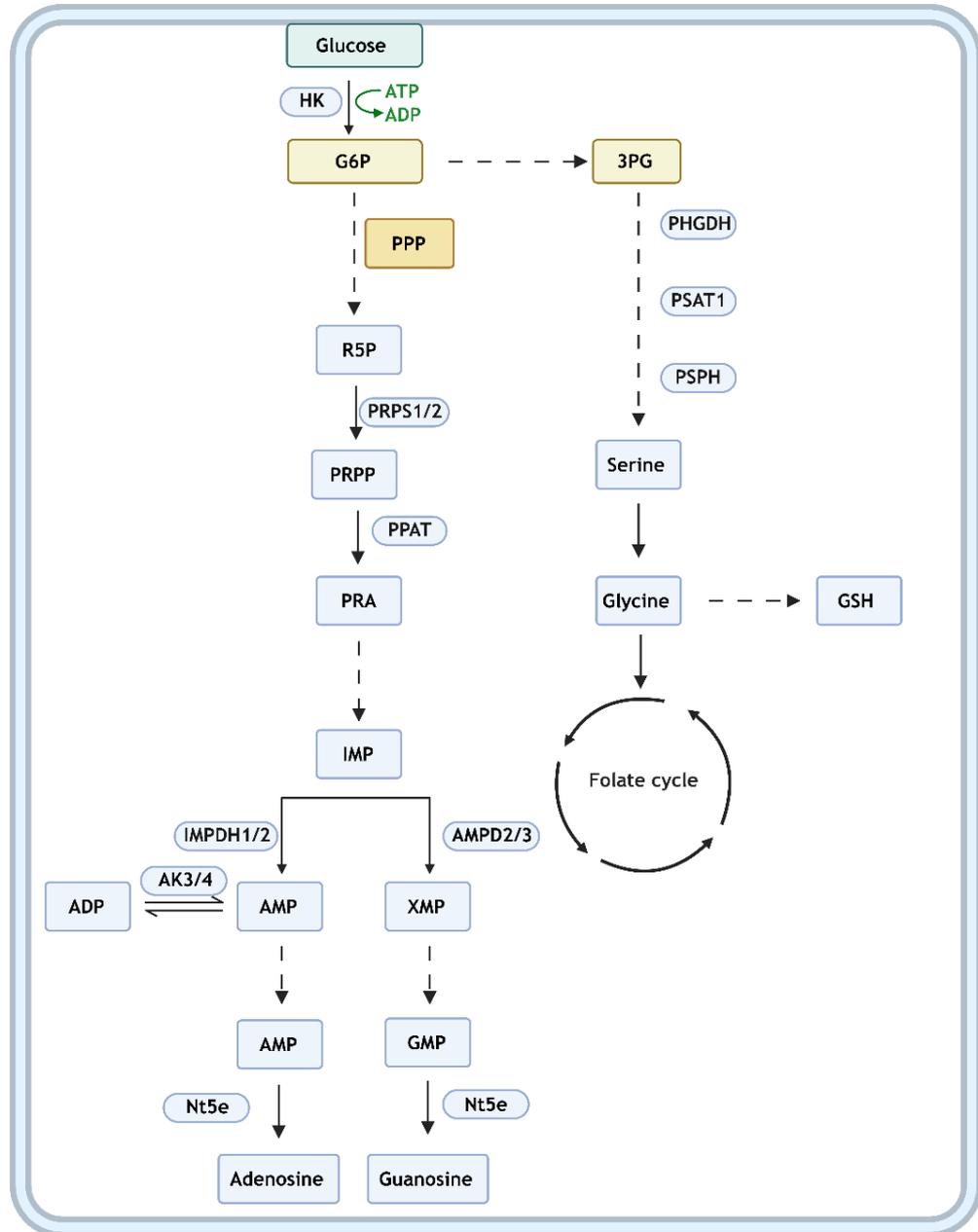


**Figure 1.6. Schematic representation of glycolysis and TCA pathways, and its metabolic branches.** Glucose is transported from the extracellular space or produced by glycogen degradation, and it can be transformed into pyruvate by producing ATP and NADH. Pyruvate can then be transformed into lactate or be introduced into the TCA cycle through Acetyl-CoA formation. This cycle can be filled by aspartate, acetate and glutamate. Glucose can also be derived to glycogen, nucleotide biosynthesis or lipid metabolism. 1,3BPG 1,3-bisphosphoglycerate, F1,6-BP fructose 1,6- bisphosphate, F6P fructose 6-phosphate, GAPD glyceraldehyde-3-phosphate, GAPDH glyceraldehyde-3-phosphate dehydrogenase, GLS glutaminase, GOT2 glutamate oxaloacetate transaminase, G6P glucose-6-phosphate, GPI phosphoglucose isomerase, GPT glutamine pyruvate transaminase, GS glutamine synthetase, HK

hexokinase, LDHA lactate dehydrogenase A, MCT monocarboxylate transporter, PDH pyruvate dehydrogenase, PEP phosphoenolpyruvate, PFK phosphofructokinase, PG phosphoglycerate, PGK phosphoglycerate kinase, PGAM phosphoglucomutase, PK pyruvate kinase, SLC solute carrier family, TCA tricarboxylic acid, TPI Triosephosphate isomerase, R5P ribose 5 phosphate, CS citrate synthase, IDH isocitrate dehydrogenase,  $\alpha$ -KG DH alpha-ketoglutarate dehydrogenase, SS succinyl-CoA synthetase, SDH succinic dehydrogenase, FH fumarate dehydrogenase, MDH malate dehydrogenase. Created with BioRender.com

Moreover, Warburg's theory fails to explain the fact that respiration and other mitochondrial activities are required for tumour growth in several tumour types (Chandel, 2016). This is important as, proliferation requires not only energy but nucleotides, lipids and amino acids. However, the biomass synthesis consumes energy, thus, if glycolysis generates only 2 ATPs per glucose molecule or if these carbons are used as a carbon source for new molecules, ATP and nicotinamide adenine dinucleotide hydrogen (NADH) / nicotinamide adenine dinucleotide phosphate hydrogen (NADPH) cannot be produced in sufficient quantities (Matthew G. Vander Heiden, 2009). As a result, the cell can use other metabolites as an energetic source such as glutamine which can be incorporated into the TCA cycle and be converted to pyruvate and lactate (Figure 1.6). The conversion of glutamine to lactate generates NADPH (Figure 1.6) (Ralph J. DeBerardinis and Thompson, 2007). Lactate can also be used as an energy source as it can be taken up by monocarboxylate transporter (*MCT*) and converted to pyruvate by NADH generation (Figure 1.6) (Rabinowitz and Enerback, 2020). Acetate, together with glutamine, have been proved to provide alternative acetyl-CoA sources when access to glucose-derived acetyl-CoA is limited, as well as branch chained amino acids which can be oxidated to acetyl coenzyme A (acetyl-CoA) (Figure 1.6) (Chandel, 2016). These alternative sources of fuel and reprogrammed metabolic activities in cancer cells allows the use of glycolysis as a source of metabolites for other pathways. For instance, glucose-6-phosphate (G6P) serves as a starting point for the pentose phosphate pathway, which converts G6P into ribulose-5-phosphate (Rib5P). At this point, several glycolysis intermediates can be synthesized (Werner et al., 2016) or Rib5P can be transformed to ribose-5-phosphate (R5P) which is the precursor of de novo-synthesised nucleotides (Figure 1.7) (Khedkar et al., 2016). Moreover, glucose can be stored as glycogen, which is involved in gluconeogenesis and acetyl-CoA can be used to synthesize fatty acids (Figure 1.6) (Li et al., 2015). It can also be diverted to glycine and serine biosynthesis which will provide antioxidants, or it

will sustain the folate cycle and support nucleotide synthesis (Figure 1.7). This partly explains why some cancer cells switch to a less efficient energetic pathway to sustain their proliferation, as by potentiating glycolysis cells have access to other metabolites pathway and other metabolites.

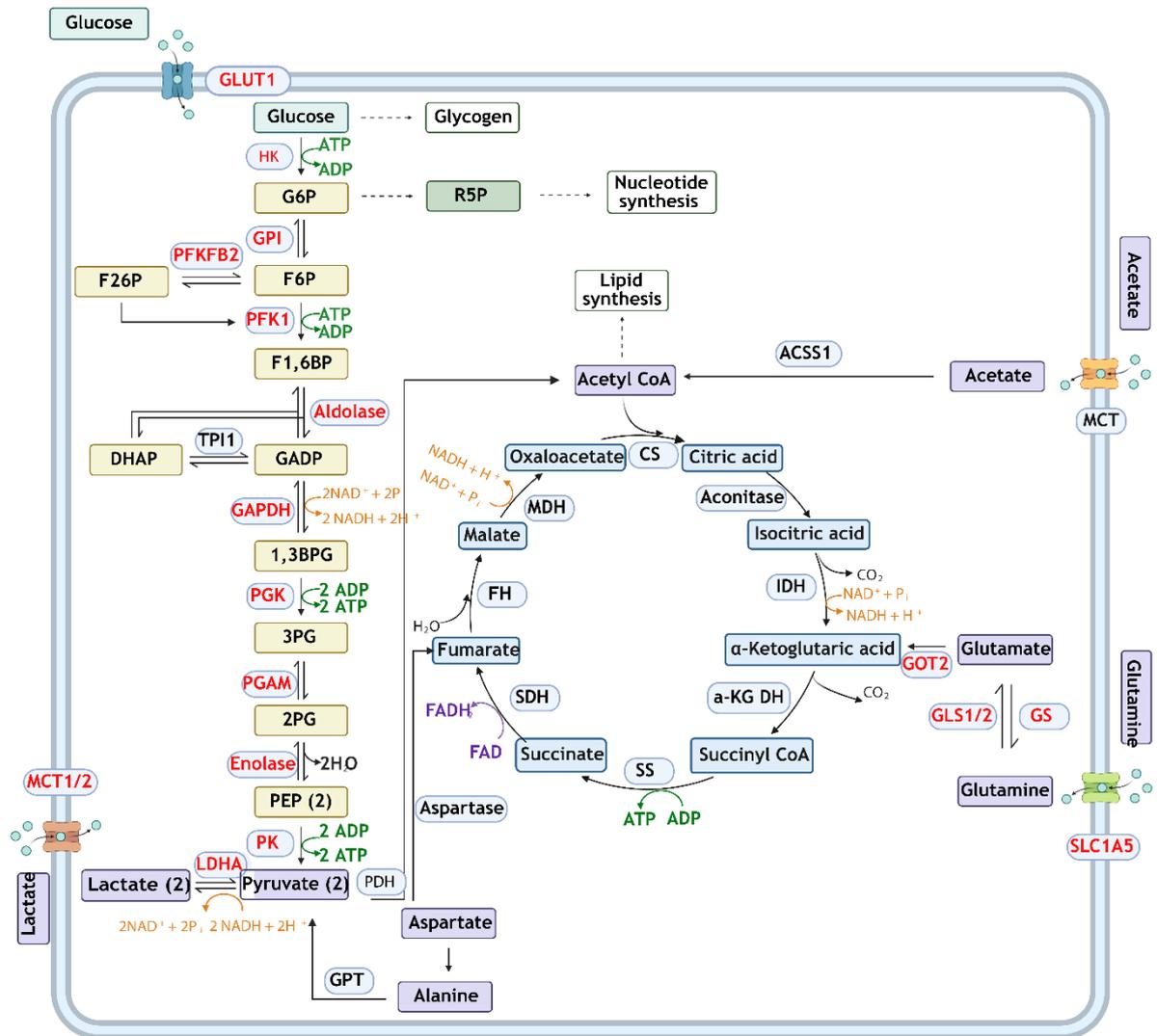


**Figure 1.7. Schematic representation of purine and pyrimidine biosynthesis.** Glucose can be derived into the PPP and be used for purine biosynthesis through the intermediary synthesis of IMP. Glycolysis intermediates can be used to sustain one carbon metabolism including serine and glycine biosynthesis, which can fill the folate cycle. GMP guanosine monophosphate, glucose-6-phosphate, HK hexokinase, PPP pentose phosphate pathway, R5P ribose 5-phosphate, PRPS1/2 phosphoribosyl pyrophosphate synthetase 1, PRPP phosphoribosyl pyrophosphate, PPAT phosphoribosyl pyrophosphate amidotransferase, IMP inosine monophosphate, IMPDH inosine-5'-

monophosphate dehydrogenase, AMP adenosine monophosphate, ADP adenosine biphosphate, AK3/4 Adenylate Kinase, Nt5e 5'-Nucleotidase Ecto, XMP xanthosine monophosphate, PHGDH phosphoglycerate dehydrogenase, PRA 5-phosphoribosylamine , PRPS pyrophosphate synthetase, PSPH phosphate ester hydrolysis, GSH glutathione. Created by BioRender.com.

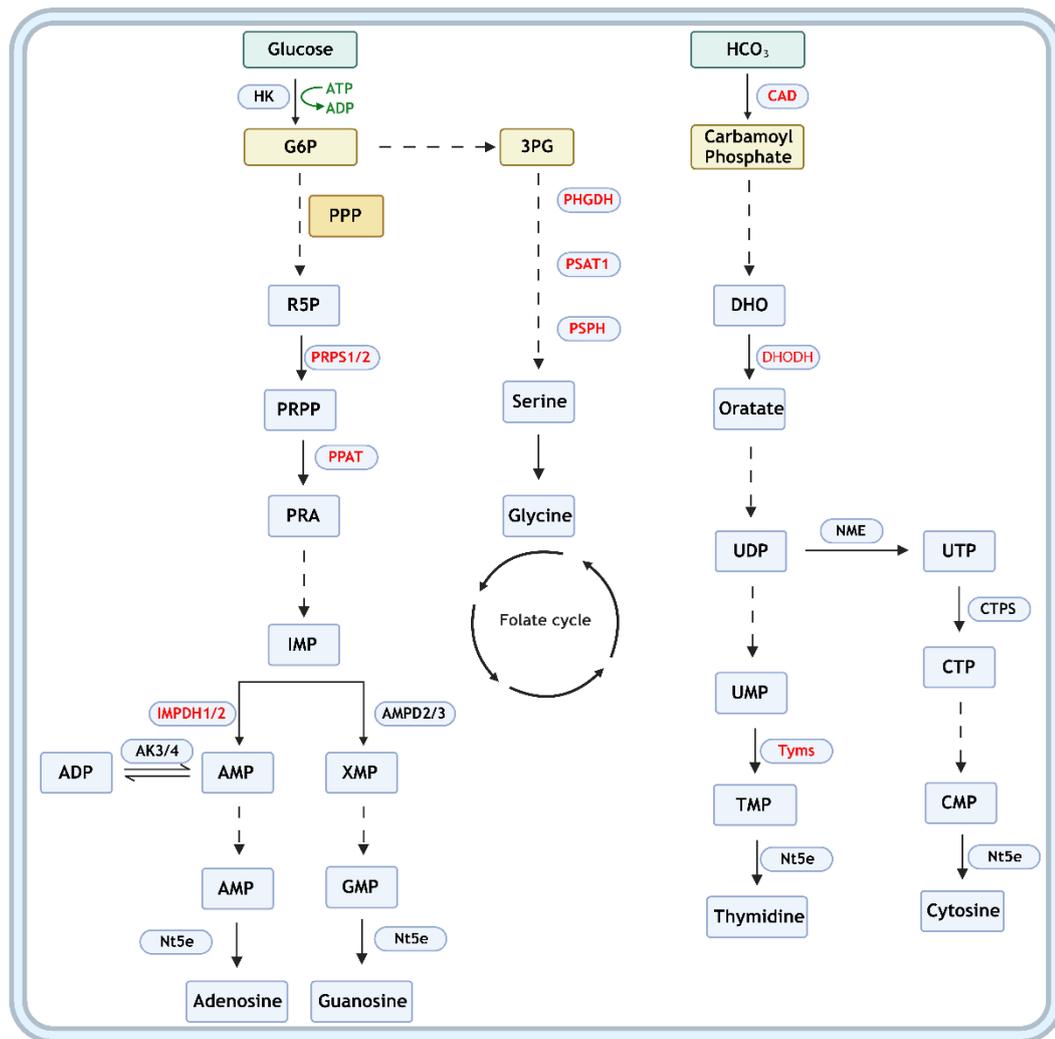
There are several regulators of glycolysis. For instance, MYC promotes glucose metabolism by upregulating glucose transporters and nearly all the glycolytic enzymes. It also upregulates glutamine addiction by promoting glutamine uptake and glutaminolysis, upregulating glutaminases 1 or 2 (*GLS1/2*), which will allow the incorporation of glutamate into the TCA cycle (Dong et al., 2020).

Additionally, MYC not only upregulates E2F transcription factor which regulates several gene expressions, not only cell proliferation genes of mid-to-late G1 phase (Kent and Leone, 2019), but also glycolytic genes. For instance, *PFKFB2*, which encodes an enzyme that is critical for energy production during glycolysis (Silvia Fernandez de Mattos, 2002). This enzyme is a well-known bifunctional enzyme with kinase and phosphatase activity, and it regulates the formation and degradation of fructose 1,6-biphosphate (F2,6BP). This metabolite allosterically regulates PFK1 which converts fructose 6-phosphate to fructose 1,6-biphosphate. This step determines the glycolytic flux (Figure 1.8) (Bartrons et al., 2018).



**Figure 1.8. Effect on MYC expression in glycolytic genes.** In this schematic glycolytic representation genes upregulated by MYC are shown in red. 1,3BPG 1,3-bisphosphoglycerate, F1,6-BP fructose 1,6- bisphosphate, F6P fructose 6-phosphate, GAPD glyceraldehyde-3-phosphate, GAPDH glyceraldehyde-3-phosphate dehydrogenase, GLS glutaminase, GOT2 glutamate oxaloacetate transaminase, G6P glucose-6-phosphate, GPI phosphoglucose isomerase, GPT glutamine pyruvate transaminase, GS glutamine synthetase, HK hexokinase, LDHA lactate dehydrogenase A, MCT monocarboxylate transporter, PDH pyruvate dehydrogenase, PEP phosphoenolpyruvate, PFK phosphofructokinase, PG phosphoglycerate, PGK phosphoglycerate kinase, PGAM phosphoglucomutase, PK pyruvate kinase, SLC solute carrier family, TCA tricarboxylic acid, PFKFB2 6-Phosphofructo-2-Kinase/Fructose-2,6-Bisphosphatase 2. Created with BioRender.com

Finally, *MYC* drives nucleotide biosynthesis by facilitating generation of phosphoribosyl pyrophosphate (PRPP) and by directly inducing enzymes involved in de novo nucleotide biosynthesis (Dong et al., 2020), which offers de novo purine and pyrimidine biosynthesis; by facilitating generation of serine and glycine, which offers one-carbon unit for folate cycle (Figure 1.9).

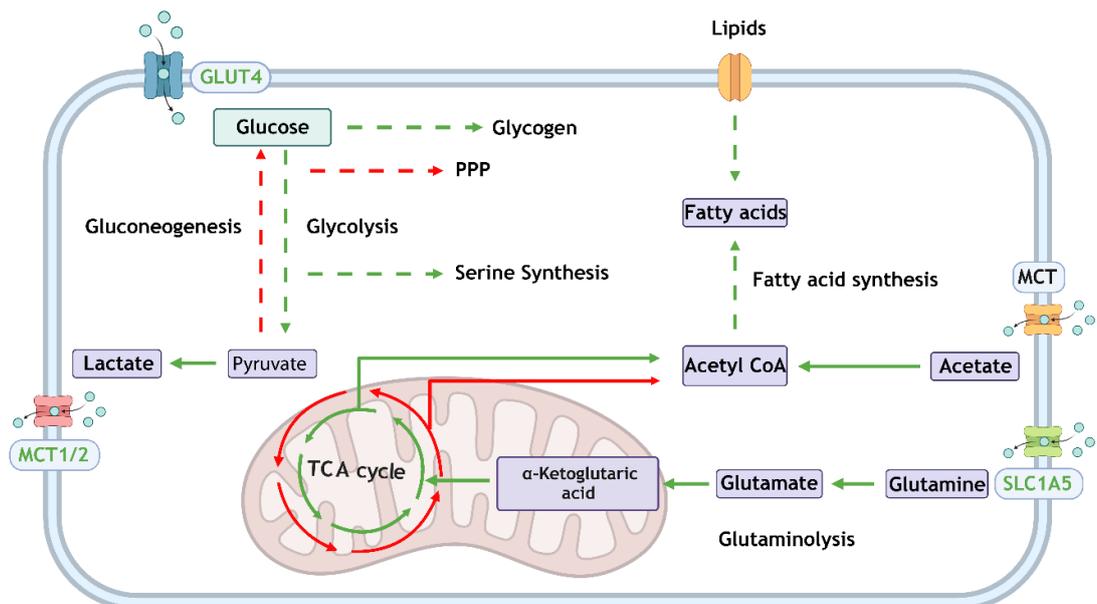


**Figure 1.9. Effect on MYC expression in nucleotide biosynthesis genes.** Schematic purine/pyrimidine biosynthesis representation where genes upregulated by MYC are shown in red. GMP guanosine monophosphate, glucose-6-phosphate, HK hexokinase, PPP pentose phosphate pathway, R5P ribose 5-phosphate, PRPS1/2 phosphoribosyl pyrophosphate synthetase 1, PRPP phosphoribosyl pyrophosphate, PPAT phosphoribosyl pyrophosphate amidotransferase, IMP inosine monophosphate, IMPDH inosine-5'-monophosphate dehydrogenase, AMP adenosine monophosphate, ADP adenosine diphosphate, AK3/4 Adenylate Kinase, Nt5e 5'-Nucleotidase Ecto, XMP xanthosine monophosphate, PHGDH phosphoglycerate dehydrogenase, PRA 5-phosphoribosylamine, PRPS pyrophosphate synthetase, PSPH phosphate ester hydrolysis, GSH glutathione, CAD carbamoyl-phosphate synthetase, DHO dihydroorotate, DHODH dihydroorotate dehydrogenase, UDP uridine diphosphate, UTP uridine triphosphate, UMP, uridine monophosphate, Tysms thymidylate synthase, TMP thymidine monophosphate, CDP cytosine diphosphate, CMP cytosine monophosphate, CTPS CTP Synthase 1, NME Nucleoside Diphosphate Kinase. Created by BioRender.com.

In addition, MTORC1 activity can be upregulated in tumour through KRAS-PI3K activation (Tian et al., 2019). mTORC1 promotes glycolysis by directly or indirectly (through the intermediate activation of HIF-1 $\alpha$  and MYC) several

glycolytic genes including *HK2*, *PFK* and *PKM2*, which are rate-limiting enzymes (Fan et al., 2021).

Additionally, hypoxia has been described to regulate involved in cell survival, autophagy, angiogenesis, glycolysis and metastasis. This pathway has also been described to promote glycolysis by upregulating glucose transport 4 (*GLUT4*) gene and promoting glucose uptake and it induces the expression of lactate dehydrogenase A (*LDHA*) gene to induce the synthesis of lactate and pyruvate dehydrogenase kinase 1 (*PDK1*) which inhibits pyruvate dehydrogenase (*PDH*) that incorporates pyruvate into the TCA cycle. Lactate is accumulated in the cell and can be excreted through *MCT4* transporter. Hypoxia is characterized by decreased flux through the TCA cycle in mitochondria. Moreover, under non-acute hypoxia electron transport chain activity is maintained, generating ROS which will in consequence end up activating adenosine monophosphate-activated protein-kinase (*AMPK*) to suppress ATP-consuming processes. Consequently, as a protective mechanism, glucose is diverted towards the serine synthesis pathway to overcome the loss of cellular antioxidant capacity, while pentose phosphate pathway activity and nucleotide synthesis are decreased. Hypoxia also promotes glycogenesis, which could provide a mechanism of energy storage to survive prolonged stress. Fatty acid synthesis and glutamine uptake is also upregulated through hypoxia (Figure 1.10) (Lee et al., 2020b).



**Figure 1.10. Schematic representation of changes in glucose metabolism due to the activation of hypoxia pathway.** Hypoxia promotes glycolysis by uptaking glucose from the extracellular space and converting it into lactate. It also upregulates lactate transport MCT1/2. Gluconeogenesis is inhibited as well as the entrance of pyruvate into the TCA cycle. However, it promotes the use of secondary nutrients to fill the TCA cycle such as glutamine and glutamate and their uptake. It also promotes the use of acetyl-coA for fatty acid and lipid metabolism. MCT1/2 monocarboxylate transporter 1, SLC solute carrier, GLUT 4 glucose transporter 4, TCA tricarboxylic acid. Created by BioRender.com.

Finally, *TP53* shows an inhibitory effect on glycolysis as directly stimulates oxidative phosphorylation. Therefore, deletion or function loss of P53 shifts metabolism from oxidative phosphorylation to glycolysis (Li et al., 2015).

Finally, neutral alterations described those metabolic alterations which happen to be dispensable for tumour growth and in contrast with transforming and enabling alterations, they are poor therapeutic targets. For instance, pyruvate kinase M2 (*PKM2*) is an isoform of the glycolytic enzyme pyruvate kinase that is expressed in most cancers and is regulated by oncogenic signalling. *PKM2* expression is not required for growth of breast tumours, liver tumours, leukaemia, or xenograft tumours from various human cancer cell lines in mice (Vander Heiden and DeBerardinis, 2017).

It is now understood that metabolic phenotypes in lung cancer are both heterogeneous and flexible, and they result from the combined effects of many different factors, including the tissue of origin and the acquired genetic mutations during the tumour development (Kim and DeBerardinis, 2019).

### **1.6.1 The effect of genomic alterations on lung cancer metabolism**

Mutations and alterations in the copy number of oncogenes and tumour suppressor genes can drive metabolic phenotypes in tumours arising in the same tissue (Kim and DeBerardinis, 2019). There are two mechanisms by which genetic alterations can cause metabolic heterogeneity. Firstly, the tissue where the oncogenic insult arises results in different metabolic strategies. For instance, *Myc*-induced murine liver tumours have an increased glucose glutamine catabolism, but in murine lung tumours *Myc* stimulates glutamine synthesis (Yuneva et al., 2012). *KRAS* activation and *P53* absence in NSCLC results in

branched-chain amino acids uptake, which is used as a nitrogen source, whereas pancreatic ductal adenocarcinomas have a decreased branched-chain amino acid uptake (Mayers, 2016). Moreover, different genetic insults in the same tissue can drive different tumour metabolic features. In lung cancer, many driver mutations in NSCLC have been associated with different metabolic adaptations. For instance, *KEAP1* mutant lung tumours are resistant to oxidative stress and depend on glutamine catabolism (Romero et al., 2017). Mutations and amplifications of *EGFR* in NSCLC have been associated with alterations in lipid synthesis, uptake, oxidation, and storage (Eltayeb et al., 2022). In SCLC, levels of achaete-scute homolog-1 (*ASCL1*) transcription factor define a metabolic group. *ASCL1*<sup>Low</sup>-SCLCs are linked to *MYC* dependent de novo guanosine nucleotide synthesis, through the activation of inosine monophosphate dehydrogenase-1 and -2 (Huang et al., 2018).

Combinations of genomic alterations can also drive specific metabolic signatures. For example, in NSCLC, tumours with mutations in *KRAS* and deletion of *LKB1*, express the urea cycle enzyme carbamoyl phosphate synthetase-1 (*CPS1*), which produces carbamoyl phosphate in the mitochondria altering nitrogen metabolism. Oncogenic *KRAS* alone or in combination with the deletion of other suppressor genes fails to show this phenotype (Kim et al., 2017). Deletion of tumour suppressor genes can result in distinct metabolic signatures. In cancer cells methylthioadenosine phosphorylase is frequently lost due to its proximity to the commonly deleted tumour suppressor gene, *CDKN2A*. Methylthioadenosine phosphorylase cleaves methylthioadenosine, which increases its intracellular concentration after methylthioadenosine phosphorylase deletion and inhibits arginine methyltransferase 5, revealing a potential vulnerability across multiple cancer lineages (Kryukov, 2016).

### **1.6.2 The effect of cell-of-origin on cancer metabolism**

Another potential factor affecting cancer metabolic heterogeneity and phenotype is the cell-of-origin of the tumour. The pure effect of cell-of-origin in cancer metabolism has been proved with leukaemia where it has been described that pyruvate dehydrogenase requirement for tumorigenesis depends on the tumour initiating cells (Buzzeo et al., 2007). Lesions found in the other tissue types but arising from different cell types can have distinct metabolic signatures

in some cancer types (Kim and DeBerardinis, 2019). This was first demonstrated in breast cancer where different histological subtypes have been defined and they have been described to arise from different cells (Perou, 2000). Basal breast tumours have a significant lower level of glutamine synthetase than luminal tumour cells, thus they are dependent on exogenous glutamine, while luminal cells express glutamine synthetase and are resistant to glutamine deprivation (Kung et al., 2011).

There are similar studies in lung cancer metabolism. Metabolic features in lung cancer-initiating cells have been described as essential for tumour development. Altered glycine decarboxylase activity in lung cancer-initiating cells is a critical factor for tumour development as it alters glycolysis and glycine/serine metabolism, leading to changes in pyrimidine metabolism to regulate cancer cell proliferation (Zhang et al., 2012). Moreover, ADC and SCC are the two predominant subtypes of NSCLC and, as explained above, they seem to arise from different cell types. SCC showed an elevated expression of *GLUT1* glucose transporter which increased glucose uptake and therefore, they showed a dependency on glycolysis while ADC are independent of glucose uptake and glycolysis to proliferate. This glucose uptake in SCC tumours correlated with a [<sup>18</sup>F]FDG uptake in position emission tomography (PET) imaging and uncovers significant implications in SCC diagnosis and therapeutic strategies (Goodwin et al., 2017). It has been described here not only a certain overlap between the cell types that can initiate ADC and SCC tumours, but they can also arise from different genetic alterations. Therefore, the effect of lung cancer cell-of-origin in metabolic phenotype and heterogeneity with independence of the genetic initial genetic alterations has never been studied.

## **1.7 Hypothesis, aims and objectives**

Given that each one of these cell types present a different epigenetic patterns (Kim, 2017) and considering that epigenetics alterations in cancer have been proved to be associated with specific metabolic features in multiple cancer types (Kim and DeBerardinis, 2019) it is reasonable to think that lung tumours arising from different lung cancer initiating cells may result in lesions with a distinctive metabolic signature. Therefore, the main aim of this project is to

establish the dependency of lung cancer metabolic heterogeneity on the cell-of-origin.

There are three main objectives in this research project:

1. To establish lineage-restricted organoids by sorting normal lung cells into their different cell-types, growing undifferentiated cells in 3D cultures and inducing malignant transformation in vitro. As described above there are five types of cancer-initiating cells in the lung epithelium: AT2, club, basal, NE cells and BASCs. Our aim is to isolate all of these cell types using specific cell surface markers, seed them in 3D and activate them in vitro to build lineage-restricted lung cancer organoids, which can then be used for metabolic characterisation of lung tumours.

*LSL-KRas<sup>G12D/+</sup> P53<sup>fl/fl</sup>* (KP) mouse model has been used widely globally to study lung cancer development but the effects of the cell-of-origin in this model have not been extensively studied. To effectively assess and understand the growth kinetics during tumour induction in KP mice, we will optimise the titre of cell-type-specific viruses based on the quantity of target cells within the epithelium and their accessibility to the viral vector.

2. The aim is to create lineage-restricted KP mouse models with similar growth kinetics to reduce the model variability and to isolate the effects of cell-of-origin on these lesions.
3. Finally, we will study and compare the effects of tumour-initiating cells on metabolism in the resultant in vivo and in vitro models using a wide range of techniques including RNA sequencing, proteomics, metabolomics, PET imaging, immunohistochemistry and flow cytometry.

## Chapter 2 Materials and Methods

### 2.1 Animal work

#### 2.1.1 Breeding and Maintenance of animals

All experimental work involving mice was conducted under license and regulations issued by the UK Home Office, in accordance with the Animals (Scientific Procedures) Act, 1986 and the European Directive 2010/63/EU and authorised by Animal Welfare and Ethical Review Board. Surplus mice were euthanised using schedule 1 protocols. Induced mice were sampled and imaged following PP4144283 licence guidance. Individually ventilated cages (IVC) with access to standard diet and water at all times and regular monitoring of general health were used for regular breeding and maintenance of mice. Tissue biopsies of ear notches from 4 weeks old mice by the Biological Services Unit at the CRUK Beatson Institute were acquired for identifications and genotyping purposes and were submitted to Transnetyx, Inc. (Cordova, TN, USA).

#### 2.1.2 SPC, CC10, CGRP and CMV Mouse Induction

Viruses were previously prepared by mixing the vector with MEM medium and 10mM of CaCl<sub>2</sub>. KP and Membrane-targeted tandem dimer Tomato/membrane-targeted green fluorescent protein (mTmG) mice were bred in-house under specific-pathogen free conditions. Typically, 8 to 10 weeks old KP mice, both males and females, were used for tumour induction by intranasal inhalation under isoflurane anaesthesia (induction 4% of anaesthetic in oxygen) using 5 x 10<sup>6</sup> pfu/ml Ad5-mSPC-Cre, 1 x 10<sup>8</sup> pfu/ml Ad5-CC10-Cre viral, 1 x 10<sup>8</sup> pfu/ml Ad5-CGRP-Cre or 1 x 10<sup>8</sup> pfu/ml Ad5-CMV-Cre (Viral Vector Core Facility, University of IOWA Health Care) vectors. They recovered on a heated bed. In order to develop KP mouse models and understand the number of cells targeted by the virus titres explained above, mTmG mice were induced following the same procedure using 1 x 10<sup>8</sup> pfu/ml Ad5-mSPC-Cre or Ad5-CC10-Cre viral vectors and 5 x 10<sup>6</sup> pfu/ml Ad5-mSPC-Cre. Mice were recovered on a heated bed. Mice were kept in cages of a maximum of 5 animals in the animal unit of the Beatson Institute. Animals were checked the day after the induction, and they were checked twice a week and weighted once a week and before every procedure.

### **2.1.3 Naphthalene optimisation**

KP mice, 8 to 10 weeks old, were intraperitoneal injected with 135 mg/kg naphthalene (Sigma-Aldrich) diluted in sterile corn oil (Sigma). Concentrations over and below 135 mg/Kg have also been tested previously to optimise the treatment. Control mice were injected only with corn oil. Animals were placed on a heated tab at 37° C for six hours after the injection and they were checked every hour. In order to optimize the naphthalene treatment concentration on KP mice, they were culled by CO<sub>2</sub> exposure 24 hours after the injection and lungs were dissected and placed in 4% paraformaldehyde for 24 hours. Samples were then placed in 70% ethanol. The tissue was paraffin-embedded and stained for CC10 and KT5. After optimisation, mice were treated with 135mg/Kg of naphthalene to deplete CC10 cells in the trachea epithelium.

### **2.1.4 KT5 Mouse Induction**

24 hours post-treatment mice were induced. Virus was previously prepared by mixing the vector with MEM medium and 10mM of CaCl<sub>2</sub>. Animals were intranasally induced while anaesthetised with isoflurane (induction 4% of anaesthetic in oxygen), using 1 x 10<sup>8</sup> pfu/ml Ad5-bK5-Cre viral vectors (Viral vector Core Facility - University of Iowa) and recovered on a heated bed. mTmG mice were induced following the same procedure using 1 x 10<sup>8</sup> pfu/ml Ad5-bK5-Cre viral vector. Mice were kept in cages of 2 animals in the animal unit of the Beatson Institute. Animals were checked the day after the induction and checked twice a week and weighted once and week.

### **2.1.5 Mouse Model Characterisation**

#### **2.1.5.1 MRI screening, tumour volume calculation and growth kinetic characterisation**

At 13, 18, 23 and 28 weeks after the induction, we performed axial, sagittal and coronal T1 and T2 MRI scans to identify individual tumours and measure their volume across these 28 weeks (Mediso Medical Imaging Systems, Hungary). During the procedure, mice were maintained under isoflurane anaesthesia (induction 4% of anaesthetic in oxygen; maintenance 1.5-2% in oxygen). Mice were scanned in a prone position and fixed with a tape. Their temperature and

heart rate were monitored during the process. Animals were warmed during and after the procedure to ease the recovery. The images were viewed and analysed using VivoQuant 4.0 (Invicro). VivoQuant is a software which provides 3D segmentation tools that allow the user to define a volume within an image for which quantity and quality can be characterised. Tumours were identified in the lungs and defined as regions of interest (ROIs). Using the Otsu thresholding 3D segmentation tool, which identifies variations of intensities within one ROI, the volume of the tumour was calculated. Growth curves for individual tumours were calculated using the exponential Malthusian growth equation in GraphPad version 9.3.1, which assumes the object keeps growing indefinitely and follows the same equation. This growth kinetic is represented by the following equation:

$$N(t) = N(0) * e^{Rt}$$

In this data, t value represents a specific time point after induction, N(t) is a specific tumour volume on that precise time point. Moreover, N(0) is the tumour volume at an initial time point and R is the constant rate of population increase. An example to properly define this rate will be bacteria grown in a lab. Bacteria reproduce by binary fission thus, if we place 1 bacterium in a flask on  $t_0$ , on  $t_1$  there will be 2 bacteria and 4 on  $t_2$ . Therefore, the constant rate of population increase (R) is defined as the number of organisms added in each generation. In this piece of data, R would be the number of cells added in each generation. However, we are not measuring cell number but tumour volume. Therefore, the constant rate is the volume increased in each measurement. This rate differs from the doubling time, which is the time needed for the tumour volume to double.

This equation returns the doubling time value, which is the time needed for the object or the tumour to double, and the constant rate which describes the volume increased in each measurement. These values were compared between the Ad5-mSPC-Cre and the Ad5-CC10-Cre tumours by running a two-tailed nested t-test. Individual tumour latency was calculated by counting the weeks until a tumour was identifiable in the MRI scans. Both groups were again compared by running a two-tailed nested t-test. The total number of tumours was calculated by summing individual tumour volume per mouse and both groups' values were compared using a t-test.

### **2.1.5.2 Survival analysis**

To generate Kaplan-Meier curves for survival analysis, overall survival data was plotted using GraphPad Prism (v9.5.1) and a Log-rank (Mantel-Cox) test was conducted to determine differences in the survival between these groups. Clinical onset was calculated as the time between birth and development of the first tumours seen in MRI imaging. Tumour progression was tracked through calculating the difference between clinical onset and clinical endpoint. Mice culled due to pathologies unrelated to genotype were not included in the analysis.

### **2.1.6 Histology**

Histological blocks and stains were performed by the Core Histology Service at the CRUK Beatson Institute. Lung tissues were fixed in fresh 10% neutral buffered formalin for 48 hours at room temperature, transferred to 70% ethanol and they were processed to formalin fixed paraffin embedded sections, prior to processing for haematoxylin and eosin (H&E) staining and immunohistochemical staining. All H&E, Elastin Van Gieson (EVG) and immunohistochemistry (IHC) staining took place on 4 µm cut from formalin fixed paraffin embedded sections tissue blocks, mounted on slides (VWR) which had previously been ovened at 60°C for 2 hours. To complete the H&E, EVG and IHC staining sections were rinsed in tap water, dehydrated through graded ethanol solutions and placed in xylene. The stained sections were coverslipped in xylene using DPX mountant (SEA-1300-00A, CellPath).

### **2.1.7 H&E staining**

H&E staining took place on a Leica autostainer (ST5020) where sections were dewaxed in xylene, taken through graded alcohols, washed in water and stained with haematoxylin z for 13 mins. Sections were washed in tap water, differentiated in 1% acid alcohol, washed and the nuclei blu'd in scotts tap water substitute (in-house). After washing the sections were stained with Putt's Eosin (in-house) for 3 minutes.

### 2.1.8 Immunohistochemistry

The following antibodies were stained on a Leica Bond Rx autostainer, Keratin 5 (905501, Biolegend), and p63 (12143-1-AP, Proteintech). All formalin fixed paraffin embedded sections underwent on-board dewaxing (AR9222, Leica) and epitope retrieval using ER2 solution (AR9640, Leica) for 20 minutes at 95°C. Sections were rinsed with Leica wash buffer (AR9590, Leica) before peroxidase block was performed using an Intense R kit (DS9263, Leica) for 5 minutes. Sections were rinsed with wash buffer before application of primary antibodies at an optimised dilution (Keratin 5, 1/1500; p63, 1/200) for 30 minutes. Sections were rinsed with wash buffer, and all had rabbit envision secondary antibody (K4003, Agilent) applied for 30 minutes. Sections were rinsed with wash buffer and visualised using DAB in the Intense R kit.

The following antibodies were stained on an Agilent AutostainerLink48, TTF1 (ab133638, Abcam) and Uteroglobin/CC10 (ab213203, Abcam). All sections were loaded into an Agilent pre-treatment module to be dewaxed and undergo heat induced epitope retrieval (HIER) using a specific target retrieval solution (TRS) where all sections were heated to 97°C for 20 minutes. Sections for TTF1 staining underwent retrieval using High TRS (K8004, Agilent) and sections for Uteroglobin/CC10 were retrieved using Low TRS solution (K8005, Agilent). After HIER sections were washed thoroughly with flex wash buffer (K8007, Agilent) prior to being loaded onto the Agilent Autostainer link48. The sections underwent peroxidase blocking (S2023, Agilent) for 5 minutes and rinsed with flex buffer. Primary antibody staining was at an optimised dilution (TTF1, 1/1000; Uteroglobin/CC10, 1/4000) for 30 minutes. Following staining sections were rinsed with flex buffer before application of rabbit envision secondary antibody for 30 minutes. Flex buffer rinsing took place before applying Liquid DAB (K3468, Agilent) for 10 minutes. After staining sections were washed in tap water and counterstained with haematoxylin z (RBA-4201-00A, CellPath), washed in tap water, differentiated in 1% acid alcohol, washed and the nuclei blu'd in scotts tap water substitute (in-house).

EVG staining was performed manually where sections were dewaxed in xylene, taken through graded alcohols and washed in water. After washing sections were

staining for EVG using a kit (RRSK11-500, Atom Scientific) where the instructions were strictly followed.

**Table 2.1. List of antibodies, autostainers and antigen retrieval methods used for IHC staining.** TTF1 Transcription Termination Factor 1), HIER heat induced epitope retrieval, EVG Elastin van Gieson.

Antibody	Dilution	Clone	Company	Autostainer	Antigen retrieval method, time
Keratin 5	1/1500	905501	Biologend	Leica Bond Rx	ER2, 20 min
p63	1/200	12143-1-AP	Proteintech	Leica Bond Rx	ER2, 20 min
TTF1	1/1000	133638	Abcam	Agilent AutostainerLink48	HIER, 20min
Uteroglobulin/CC10	1/4000	213203	Abcam	Agilent AutostainerLink48	HIER, 20min
EVG staining	-	RRSK11-500	Atom Scientific	-	-
Rabbit envision secondary antibody	-	K4003	Agilent	-	-

### 2.1.9 Histological Analysis

IHC scanned images were analysed using HALO™ v.3.6.4134. For staining quantification both the CytoNuclear FL v2.0.12 module was applied. For quantification of tumour cell staining, the CytoNuclear module was applied for to the selected tumour annotations. For clinical subtype classification, the presence or absence of a lung adenocarcinoma pattern was noted together with the presence or absence of broken elastin fibres. Graphs were made using GraphPad Prism v.9.5.1.

### **2.1.10 DNA, RNA and Protein Isolation**

Ad5-mSPC-Cre, Ad5-CC10-cre and Ad5-CMV-Cre mice were culled by CO<sub>2</sub> exposure and tumours between 3-6 mm in diameter were dissected and cut in half. The first half was used for DNA, RNA and protein isolation using the AllPrep DNA/RNA/Protein QIAGEN isolation Kit (QIAGEN). Samples were placed in a Precellys CK14 2 ml tubes with beads (Bertin Technologies) and they were homogenized at 0 °C in the Precellys 24 homogenizer (Bertin technologies) with 350 µl of RLT buffer (QIAGEN), provided in the QIAGEN isolation kit. To ensure a correct homogenization, three cycles of 30 s with a 30 s gap between them were run.

Homogenized samples were centrifuged for 3 minutes at full speed. 100 µl of this supernatant was placed in a Molecular Biology 100 µl Eppendorf tube (ThermoFisher Scientific) and they were brought for sonication. Bioruptor sonication system (Diagenode Belgium) at 4 °C was used. Samples were sonicated for 10 minutes at 30 seconds intervals. Samples were agitated for 30 minutes and centrifuged at 16,000 g for 5 minutes. This supernatant was stored at -80 °C until further use.

The remaining 250 µl were used for RNA and DNA isolation following the DNA/RNA/Protein QIAGEN isolation Kit procedure. Samples were stored at -80 °C until further use.

### **2.1.11 Metabolite Isolation**

Ad5-mSPC-Cre, Ad5-CC10-cre and Ad5-CMV-Cre mice were culled by CO<sub>2</sub> exposure and tumours between 3-6 mm in diameter were dissected and cut in half. The first half has been sent to the metabolomics department for targeted metabolomics analysis. Samples were homogenized at 20 mg/ml with the extraction polar solvent (50% methanol (Sigma-Aldrich), 30% acetonitrile (Sigma-Aldrich) and 20% water) approximately at 0 °C in the Precellys 24 homogenizer (Bertin Technologies). To ensure a correct homogenization, three cycles of 30 s with a 30 s gap between them were run. Homogenized samples were centrifuged in the Precellys CK14 2 ml tubes with beads (Bertin technologies) at 16,000 g for 10 minutes at 4 °C. The supernatant was collected and frozen down at - 80 °C for

a week. On the day of the LC-MS analysis samples were thawed on ice and centrifuged at 16,000 g for 10 minutes at 4°C. The supernatant was again collected and transferred to glass HPLC vials.

### **2.1.12 Metabolomic analysis of Tumours and Normal lung Tissue**

Tissue extracts, 5 µl, were injected in a Q Exactive Plus Orbitrap mass spectrometer (Thermo Scientific, Waltham, MA, USA) coupled with a Thermo Ultimate 3000 HPLC system. The HPLC setup consisted of a ZIC-pHILIC column (SeQuant, 150 x 2.1 mm, 5 µm, Merck KGaA, Darmstadt, Germany), with a ZIC-pHILIC guard column (SeQuant, 20 x 2.1 mm) and an initial mobile phase of 20% 20mM ammonium carbonate, pH 9.2, and 80% acetonitrile. Metabolites were separated over a 15 minutes mobile phase gradient, with an acetonitrile content up to 20%, at a flow rate of 200 µL/min and a column temperature of 45°C. The total analysis time was 25 minutes. All metabolites were detected across a mass range of 75-1000 m/z at a resolution of 70,000 (at 200m/z), with electrospray (ESI) ionization and polarity switching to enable both positive and negative ions to be determined in the same run.

A pooled sample comprising a mixture of all sample extracts was analysed using the same HPLC conditions but running the mass spectrometer in single ionization mode and using data dependent fragmentation (ddMS2) for improving the confidence in metabolite identification. This was done for both positive and negative ionisation modes. The pool sample was also used as a quality control and was repeatedly measured every ~10 sample injections throughout the acquisition of the randomised biological extracts. Data were acquired with Thermo Xcalibur software.

The data were analysed using Compound Discoverer software (Thermo Scientific v3.3). Unknown compound detection and grouping of compound adducts was carried out across all samples (mass tolerance 5 ppm, RT tolerance 0.2 min, Peak rating threshold 6 in a minimum of 6 data files). Missing values were filled using the software's Fill Gap feature (mass tolerance 5 ppm, S/N tolerance 1.5). Within Compound Discoverer, data were corrected for batch effects using the QC replicate injection data and a QC-based area correction regression model

(Linear, 50% QC coverage, 30% RSD). Feature identification was achieved by matching the mass and retention time of observed peaks to an in-house database generated using metabolite standards (mass tolerance 5 ppm, RT tolerance 0.5 min). Annotations were further confirmed using mzCloud (ddMS2) database search (Precursor and fragment mass tolerance of 10 ppm, match factor threshold 60).

### **2.1.13 Bioinformatic Analysis of Metabolomic data**

Targeted metabolomic data was analysed with Skyline v22.2.0.255 (MacCoss Lab. Department of Genome Science, University of Washington). A total of 84 metabolites from 11 pathways were analysed, including: glycolysis, pentose phosphate pathway, TCA cycle, urea cycle, methionine cycle, purine, redox pathway, amino acid pathways, fatty acid oxidation and pyrimidines pathway. Each compound peak was identified in each sample within a 2 minutes explicit retention time window. Blank and standards samples were used to ensure the correct peak identification. Peak area was plotted using GraphPad Prism v.9.5.1. Metabolic pathway analysis was run with Metaboanalyst 5.0 with Joint Pathway analysis feature. Hypergeometric test was chosen as enrichment analysis, the topology measure was Degree-Centrality measures the number of links that connect to a node. For integration methods, tight integration by combining queries was chosen in which genes and metabolites are pooled into a single query and used to perform enrichment analysis within their "pooled universe".

Unpaired parametric multiple t tests, which performs several t-test at once, were run to understand the statistical differences of these metabolites between CC10 and SPC samples and S1 and ACE tumours with a false discovery rate of a 5%. This method will decide which p value is small enough to be considered significant with no more of 5% of false positive results.

### **2.1.14 RNA Sequencing of tumours and normal lung tissue**

RNA sequencing was conducted by Graeme Clark from the Molecular Technology Services at the CRUK Beatson Institute. Quality control of all RNA samples was performed (Agilent TapeStation 4200, High Sensitivity RNA screentape), and only those samples showing RIN values >8 were processed. RNA concentrations were

determined by Qubit Fluorometer using the Qubit RNA Broad Range assay (both Thermo Fisher), with 1 ug of total RNA used as initial input. Libraries were then prepared using manufacturers standard procedures (Illumina Stranded mRNA), with IDT for Illumina RNA UD Indexes used to index libraries. Post library QC was then performed using High Sensitivity D1000 screentape (Agilent) for library sizing and profiling and quantified using Qubit High Sensitivity DNA assay. Libraries were then pooled equimolar, to a final concentration of 4 nM prior to sequencing. The library pool was then sequenced on an Illumina NextSeq 500 instrument, on a High-Output 150 cycle run with paired-end 74 bp read length.

### **2.1.15 Bioinformatic analysis of tumours and normal lung tissue**

Processing of raw RNAseq data was performed by Robin Shaw of the Bioinformatics Core Facility at the CRUK Beatson Institute. The RNA-seq preprocessing workflow involved steps aimed at preserving the quality and reliability of the sequencing data. In the initial stage, we demultiplexed Illumina sequencer results from three flowcells, resulting in 70 samples that were converted into fastq files. These files underwent an examination for potential contamination using Fastq\_screen. Subsequently, Fastp was employed to trim adapter sequences and remove low-quality bases from the reads. To assess the data's quality, FastQC was applied both prior to and after the trimming process.

The strandness of pair-end reads was determined through RSeQC, and the reads were then aligned to the mouse reference genome (Ensembl 39.109) using Hisat2. Count tables were generated using FeatureCounts, and a comprehensive quality control (QC) report was compiled using MultiQC.

Notably, the FastQ Screen analysis did not reveal any contamination issues. Among the samples, S51 and S62 from the SPC\_Tumour group exhibited the lowest read counts, falling below 5 million reads. Additionally, samples S51, S06, and S62 displayed a noticeable spike (>3%) in insert size at 35 base pairs. It is worth noting that each base position in the reads maintained a reasonable quality with a Phred score exceeding 20. Furthermore, the data exhibited low levels of over-represented sequences, with each sample accounting for less than 0.25% of such sequences.

For further analysis, filterByExpr function from edgeR library to filter out lowly expressed genes was applied. Several outliers S18, S17, S20, S06, S28, S19 and S11 were detected by creating a PCA plot using variance stabilizing transformation (VST) gene counts. These samples were manually grouped as a “subgroup”. The CC10 vs SPC and tumour vs normal tissue comparisons were performed with and without the bias effect of the “subgroup”. For pathway enrichment analysis gene set enrichment analysis was performed. Gene set enrichment analysis (GSEA) looks for gene sets with relatively high ranks of fold-change (or p-value), while enrichment analysis looks for gene sets that are overrepresented in genes with  $p < 0.05$  (and fold-change  $> 1.5$ ). GSEA is more sensitive than other pathway enrichment analysis and can be used even when there are no significant genes. The pathways used for this analysis were cancer hallmark from the Molecular Signature Database (MsigDB) and metabolic related pathway from Kyoto Encyclopaedia of Genes and Genomes (KEGG).

For S1, S2 and acetate group analysis, filterByExpr function from edgeR library to filter out lowly expressed genes was applied. An outlier, S26, was detected by creating a PCA plot using variance stabilizing transformation gene counts and was subsequently excluded from subsequent analysis. Tumour subtypes were annotated by clustering the normalized gene counts of signature genes (S1, S2, ACE, FDG) into three clusters (ACE, S1, S2) using k-means.

S1 and S2 gene sets were shared to us by Dr Melissa R Junttila who had previously defined them (Daemen et al., 2021). [ $^{18}\text{F}$ ]FDG and [ $^{11}\text{C}$ ]acetate gene signatures were obtained from imaging Ad5-CMV-Cre KP mice tumours in our research group although they have not been tested in human yet. These gene signatures indicate the uptake preference of murine tumours. As a result of imaging a large number of tumours from KP mice induced with the CMV vector, they identified two tumour types in the murine lung: acetate or FDG avid tumours. From these tumours analysis a gene signature for both have been identified. They both were included in the analysis for a better understanding of murine lung transcriptomic profile.

Differential gene expression analysis was performed using Deseq2 with a model design that included Tumour\_type and Sex. The resulting fold-changes were visualized through volcano plots, and the genes that overlapped among pairwise

comparisons were represented using Venn diagrams. To explore functional aspects, Gene Set Enrichment Analysis (GSEA), utilizing MsigDB, KEGG and Gene ontology biological process (GO BP) gene sets and the EnrichmentBrowser R package.

Principal component analysis (PCA) was performed with factoextra R package after differential gene expression to reduce dimensionality in the dataset, and to allow visualisation of the variability between different sample groups. The sample groups were set as the variable of interest and a confidence threshold was applied to each group. This allowed an elliptical to be created around each group representing the mean and 95% confidence interval for each sample group. The mean for each group lies in the centre of each elliptical, with the outline the 95% CI.

### **2.1.16 Proteomics**

Five SPC and CC10 samples, four ACE samples and three S1 samples were randomly selected for proteomic analysis. Prior to digestion samples were precipitated with 5 x volume of sample of ice-cold acetone (Merck), overnight at -20°C, to remove proprietary QIAGEN kit buffer. Precipitated samples were spun at 15,000 x g for 10 mins at 4°C and supernatant discarded. A further 2 x 300 µL of ice-cold acetone was used to ensure complete removal of any buffer, this was again spun at 15,000 x g for 10 mins at 4°C. 50 µg protein lysate were then re-suspended in 40 µL of 0.2 M HEPES buffer (Gibco), reduced with DTT (Sigma) for 1 hour at room temperature, then alkylated in the dark for 30 mins at room temperature with IAA (Sigma) to a final concentration of 5 mM and 50 mM, respectively. Samples were then digested using a two-step digestion, firstly with endoproteinase Lys-C (ratio 1:33 enzyme:lysate, Alpha laboratories) for 1 hour at room temperature then with trypsin (ratio 1:33 enzyme:lysate, Promega) overnight at 37°C. Once digested, peptide samples were then labelled with 2 x TMT 16plex reagent kit (Thermo Scientific), checked for complete incorporation of the TMT labels, mixed and cleaned using Sep-Pak columns (Waters).

800 µg (16 samples at 50 µg mixed together in a single sample) digested sample was fractionated using reverse phase chromatography at pH 10. Solvents A (98% water (Merck), 2% ACN (Merck)) and B (90% ACN (Merck), 10% water (Merck))

were adjusted to pH10 using ammonium hydroxide (Merck). Samples were run on an Agilent 1260 Infinity II HPLC where they were manually injected using a Rheodyne valve. The samples were separated over a two-step gradient, 2-28% Solvent B in 39 mins then 28-46% Solvent B in 13 mins. The column was washed for 8mins at 100% Solvent B followed by a re-equilibration for 7 mins. Total run time was 76 mins and flow rate was set to 200  $\mu$ L/min. The samples were collected into 21 fractions.

Peptide samples were run on the Thermo Scientific Orbitrap HF mass spectrometer (Thermo Scientific) coupled to an EASY-nLC II 1200 chromatography system (Thermo Scientific). Samples were loaded onto a 50 cm fused silica emitter (packed in-house with ReproSIL-Pur C18-AQ, 1.9  $\mu$ m resin, Dr Maisch) which was heated to 55 °C using a column oven (Sonation). Peptides were eluted at a flow rate of 300 nl/min over three optimised two-step gradient methods for fractions 1-7, 8-15 and 16-21 utilizing solvents A, 0.1% formic acid (Sigma) and Solvent B, 80% ACN (Merck):0.1% FA (Sigma). Step one was commenced for 113 mins and step two for 37 mins. For fractionated samples 1-7 the % of solvent B was 2-12% at step one and 20% at step two. For fractions 8-14 the % of B was 2-16% at step one and 25% at step two and for fractions 15-21 the % B was from 3-22% at step one and 33% at step two. Peptides were electrosprayed into the mass spectrometer using a nanoelectrospray ion source (Thermo Scientific). An Active Background Ion Reduction Device (ESI Solutions) was used to decrease air contaminants.

Data was acquired Xcalibur software (Thermo Scientific) in positive mode utilising data-dependent acquisition (DDA). Full scan mass (MS1) range was set to 375-1400 m/z at 60,000 resolution. Injection time was set to 20 ms with a target value of 3E6 ions. HCD fragmentation was triggered for MS2 analysis. MS2 injection time was set to 50 ms with a target of 1E5 ions and resolution of 45,000. Ions that have already been selected for MS2 were dynamically excluded for 45 s.

### **2.1.17 Proteomic Data Analysis**

MS raw data was processed using MaxQuant software version 1.6.14.0 and searched with the Andromeda search engine against the Uniprot *Mus Musculus*

database (Swiss-prot, 17,185 entries). Data was searched with multiplicity set to MS2 level TMT16plex. First and main searches were done with a precursor mass tolerance of 20 ppm for the first search and 4.5 ppm for the main. MS/MS mass tolerance was set to 20 ppm. Minimum peptide length was set to 7 amino acids and trypsin cleavage was selected allowing up to 2 missed cleavage sites. Methionine oxidation and N-terminal acetylation were selected as variable modifications and Carbimidomethylation as a fixed modification. False discovery rate was set to 1%.

MaxQuant output was processed using Perseus software version 1.6.15.0. Briefly, 'Reporter corrected intensity' columns were divided by sum, then processed through the following equations:

$$\begin{aligned} & [\textit{geometric mean of ALL pools}] \div [\textit{TMT group specific pool MEAN}] \\ & = \textit{correction factor} \end{aligned}$$

To create a normalisation correction factor for multiple TMT groups. The intensity values are then transformed  $[1e12 \cdot (x)]$  then multiplied by correction factor.

$$\begin{aligned} & [\textit{Reporter Corrected Intensity}] \times [\textit{correction factor}] \\ & = \textit{Normalised Intensity Value} \end{aligned}$$

Statistical analysis was then performed on the normalised data. Unpaired parametric multiple t tests, which performs several t-test at once, were run to understand the statistical differences of these protein between CC10 and SPC samples and S1 and ACE tumours with a false discovery rate of a 5%. This method will decide which p value is small enough to be considered significant with no more of 5% of false positive results.

### **2.1.18 PET Imaging**

PET imaging with  $[^{11}\text{C}]$ acetate and  $[^{18}\text{F}]$ FDG was performed at 75% of median life span for Ad5-mSPC-Cre and Ad5-CC10-Cre mice using a Mediso nanoScan PET/magnetic resonance imaging (MRI) (T1) scanner. During the procedure, mice were maintained under isoflurane anaesthesia (induction 4% of anaesthetic in

oxygen; maintenance 1.5-3% in oxygen). Mice were scanned in a prone position and fixed with a tape. Their temperature and heart rate were monitored during the process. Animals were warmed during and after the procedure to ease the recovery. Between 200 and 250 MBq of [<sup>11</sup>C]acetate was injected via the cannulated tail vein first, followed by the whole body-T1-weighted MRI and the chest T2-weighted MRI, which were performed for the anatomical registration. PET images with [<sup>11</sup>C]acetate were performed acquired between 80 and 100 minutes post injection. The following date, between 10 and 15 MBq of [<sup>18</sup>F]FDG was injected via the tail vein and PET images for [<sup>18</sup>F]FDG were acquired between 80 and 100 minutes post-injection. The images were reconstructed using the whole-body reconstruction protocol and Tera-Tomo 3D iterative reconstruction method for both tracers, and further analysed using VivoQuant software. Tumours were identified in the lungs from the MRI and co-registered for both tracers first. For PET image quantification, ROIs were manually drawn over the tumours to obtain the mean and maximum values of activity concentration. SUVmax and SUVmean values for [<sup>11</sup>C]acetate and [<sup>18</sup>F]FDG for each tumour were then calculated using the formula:

$$\text{SUV} = \text{ROI activity concentration (Bq/ml)} / \text{mouse weight (g)} \times \text{injected activity (Bq)}.$$

## 2.2 Tissue culture

### 2.2.1 Lung Tissue Dissociation

Induced mTmG mice were culled by CO<sub>2</sub> overdose. To remove red blood cells, the right ventricle of the heart was perfused with 7-10 ml of sterile phosphate-buffered saline (PBS) before dissecting the lungs. To isolate cells, dissected tissue was digested twice for 45 minutes, using a different combination of enzymes for each digestion. The first reaction contains 4 U/ml elastase (STEMCELL technologies), 1 U/ml dispase (STEMCELL technologies), 200 µg/ml DNase (Merck) and 10 µM of RHO/ROCK (STEMCELL technologies) pathway inhibitor in RPMI medium (Gibco). The released cells were collected to avoid an overexposure to the enzymes. The remaining tissue was cut into small (less than 1 mm) pieces and digested in the same medium containing 25 µg/ml liberase (Roche), 200 µg/ml and 10 µM of RHO/ROCK pathway inhibitor (STEMCELL

technologies) (Figure 2). Remaining red blood cells were removed using red blood cells lysis buffer (Invitrogen). Released cells were washed through a 70 µm cell strainer (StarLab) to obtain a single cell suspension, and they were kept on ice.

### **2.2.2 Trachea Tissue Dissociation**

Induced mTmG mice were culled by CO<sub>2</sub> overdose. To remove red blood cells, the right ventricle of the heart was perfused with 7-10 ml of sterile PBS. Connective tissue was snipped away to allow the trachea to be dissected out. Trachea was cut and placed in a 24-well dish (Corning) with 500 µl of 50 U/ml Dispase (Sigma-Aldrich) solution at room temperature for 40 minutes. After this time trachea is moved to a fresh well containing 500 µl of PBS and inject 500 µl of PBS through trachea using a 25G needle and 1 ml syringe. Trachea was flushed at least five times reusing the volume in the well. All the detached cells were transferred to an Eppendorf and spined. Cells were resuspended in 100 µl 0.25% Trypsin/EDTA (Gibco) for 2 minutes maximum. This final cell suspension was kept on ice until further use.

### **2.2.3 Extracellular Cell Staining**

Single cell suspension obtained from Methodology sections 1.2.1 and 1.2.2 was centrifuged at 1000 rpm for 5 minutes and resuspended in sterile PBS. Cells were stained with 1:200 Yellow Zombie viability dye for 20 minutes at room temperature. Cells were then centrifuged at 1000 rpm for 5 minutes before resuspended in 50 µl Fc Blocking Buffer and incubated at 4°C for 20 minutes at room temperature. Fc Blocking Buffer was prepared by diluting TruStain FcX™ anti-mouse CD16/32 (101320, Biolegend) 1:50 in FACS buffer (PBS supplemented with 1% foetal bovine serum (FBS) (Gibco) and gentamicin). After this time, cells were centrifuged at 1000 rpm for 5 minutes and resuspended in 1 ml of FACS buffer. Conjugated antibodies were added to the suspension for 30 minutes in the dark at room temperature. Fluorescence Minus One controls were also prepared by staining the same lung murine cells and they were used to ensure appropriate gating. Unstained control was also provided. Finally, all samples were filtered using a 70 µm filter into 5 ml Corning™ Falcon® Round-Bottom Tubes with cell strainer caps (10585801, Fisher Scientific) for acquisition.

**Table 2.2.** List of antibodies used in flow cytometric analysis of lung cells. EpCAM epithelial cell adhesion molecule, MHC major histocompatibility complex, NGFR neural growth factor receptor.

Antigen	Fluorochrome	Dilution	Stock concentration	Clone	Catalogue	Source
CD31	APC	1:200	0.2mg/ml	MEC13.3	102510	BioLegend
CD45	APC	1:200	0.2mg/ml	30-F11	103112	BioLegend
CD326 (EpCam)	PE-Cy7	1:100	0.2mg/ml	G8.8	118216	BioLegend
MHC Class II	eFluor450	1:200	0.2mg/ml	M5/114.15.2	48-5321-82	Invitrogen
Ly-6A/E Sca-1	APC-Cy7	1:200	0.2mg/ml	D7	560654	BD Biosciences
CD24	PE	1:300	0.2mg/ml	M1/69	101807	BioLegend
CD24	Brilliant Violet 421	1:300	50µg/ml	M1/69	101825	BioLegend
CD56	Brilliant Violet 711	1:100	0.2mg/ml	809220	748099	BD Biosciences
NGFR	Alexa Fluor 488	1:200	0.2mg/ml	B-1	Sc-271708	Santa Cruz Biotechnologies
CD16/32	None	1:1000	0.5mg/ml	ST7011E	156604	BioLegend
Zombie Yellow	Zombie Yellow	1:200	-	-	423104	Biologend

## 2.2.4 Compensation and flow cytometry analysis

Cells were run in the cell cytometer in FACS buffer. The flow cytometer used was BD LSRFortessa™ Z6102 (BS Biosciences) and the results were analysed using FlowJo software. Compensation control preparation involves a 1.4 ml U-bottom FACS tubes (Micronic) were set up for each antibody utilised in each experiment. To each tube, 1 ml FACS buffer and 2 µl of antibody was added to generate a single stain control, together with a single drop of UltraComp eBeads™ Compensation Beads (Invitrogen). Each drop of beads contains two populations: a positive population that will capture any mouse antibody and a negative population that will not react with antibody, making each sample its own unstained control. These samples were vortexed for a minimum of 40 seconds to 1 minute. ArC Amine Reactive Compensation Beads (Thermo Fisher Scientific)

were used to prepare the compensation control for the Yellow Zombie viability dye. To each tube, 1 ml FACS buffer, 3 µl of Yellow Zombie viability dye and a drop of amine reactive beads (component A) which bind any of the amine-reactive dyes and provide a positive signal and a drop of the ArC negative beads (component B) have no reactivity and provide a negative compensation control.

Following compensation, sample acquisition was carried out on LSRFortessa™ (BD Bioscience) using the DIVA software. Data analysis was performed using Flowjo software v10.8.1 (Flowjo, LLC)

### **2.2.5 Lung epithelial and mesenchymal cell sorting**

Lung epithelial cells and mesenchymal cells isolation was performed following protocols detailed in section 2.2.1. Cells were stained following the steps detailed in section 2.2.3 and compensation controls were prepared following the steps in section 2.2.4. Cell sorting was carried out by Tom Gilbey and Yi-Hsia Liu from the Beatson Advanced Imaging Resource (BAIR) and the flow cytometry central services. Cells were brought to a concentration of 5-6 million cells/ml of FACS buffer. BD FACSAria™ III cell sorter (BD Biosciences) was used for the sorting. The nozzle size was 100µm and the sheath pressure was 20 psi. Mesenchymal cells were collected in DMEM supplemented with 10% FBS, 4 mM glutamine (Gibco) and 1 mg/ml of gentamicin (Gibco). Lung epithelial cells were collected in DMEM-Ham's F-12 media (Gibco) supplemented with 15 mM HEPES (Gibco), 8 mM L-Glutamine, 1x penicillin/streptomycin (Gibco), 10% FBS and 5 µg/mL ITS (Gibco).

### **2.2.6 Mesenchymal cell isolation**

Mice 4 week old were culled by CO<sub>2</sub> exposure. Lungs were inflated with 1 ml of RPMI medium containing 200 µg/ml of DNase (Sigma), 4 U/ml of elastase (STEMCELL technology), 1 U/ml of dispase (STEMCELL technologies) and 1 mg/ml of gentamicin (Gibco). Lungs were removed and placed in a 3.5 cm petri dish (Corning) containing the same RPMI medium (Gibco) for 45 minutes at 37°C. After this time, tissue was minced in small tissues using a scalpel and this cell suspension was strained through a 100 µm strainer (StarLab). The remaining pieces of tissues were placed in a 6cm petri dish (Corning) containing RPMI

medium with 25 µg/ml of liberase (Roche), 200 µg/ml of DNase (Sigma) and 1mg/ml of gentamicin for 45 minutes at 37°C. This cell suspension was again strained through a 100 µm strainer. Cells were centrifuged at 1000 rpm for 5 minutes at 4°C and resuspended in 1x Red Blood Cell lysis buffer (Invitrogen). Cells were incubated for 5 minutes at room temperature and centrifuged again at 1000 rpm for 5 minutes at 4°C. Cells were resuspended in 1x PBS containing 1 mg/ml of gentamicin. Released cells were washed through a 70 µm cell strainer to obtain a single cell suspension, and they were kept on ice. This cell suspension was stained with 1:200 dilution of Zombie Yellow Fixable Viability dye (BioLegend) and incubated for 20 minutes at room temperature. Cell suspension is centrifuged and resuspended in 1x PBS containing 1mg/ml of gentamicin. Cells were stained with 1:200 dilution of Rat monoclonal anti-CD45 APC, 1:200 dilution of Rat monoclonal anti-CD31 APC and 1:100 Rat monoclonal anti-CD326 (EpCAM) PE/ Cy7 (Table 2.2). Alive triple negative cells for these three cell surface markers were sorted following protocol detailed in section 2.2.5. Sorted cells were cultured in a 10cm petri dish containing DMEM 21969-035 (Gibco) with 10% FBS, 4 mM glutamine and 1 mg/ml of gentamicin.

### **2.2.7 Endothelial Cells Isolation**

Wild-type (WT) mice, 20 days old, were culled by CO<sub>2</sub> exposure. Lungs were removed and diced as fine as possible using a scalpel and placed in a 10 cm petri dish. Tissue was digested in 1 x PBS supplemented with 10 mg/mL type II collagenase (STEMCELL technologies) and 20 µg/mL DNase I (Sigma) for 40 minutes at 37°C. After this time, tissue was further disassociated using a 5 ml pipette and strained through a 100 µm and a 40 µm strainers (StarLab). Equal volume of FBS was added to quench collagenase activity. Sample was centrifuged at 1000 rpm for 10 minutes at 4°C and washed once with 10 ml of 1x PBS. Cells were resuspended in 5 ml of Advanced DMEM (Gibco), and this cell suspension was laid very carefully over 5 ml Histopaque 1077 (MP biomedical). This suspension was centrifuged at 1000 rpm for 23 minutes at 4°C. The cloudy interface was removed, and cells were centrifuged again at 1000 rpm for 10 minutes at 4°C. The pellet was resuspended in 1ml of 1 x PBS and centrifuged again. The pellet was resuspended with 80 µL of PBS and 20 µL of CD31-conjugated microbeads (Milteny Biotec) and incubated for 15 minutes at 4°C. Cells are washed with 1 ml of 1 x PBS and resuspended in 500 µL of 1 x PBS. This

cell suspension was placed in a LS column in a magnetic field of a MACS separator and the column (Miltenybiotec) was washed with 1x PBS to allow unlabelled cells to go through. The column was removed from the magnet and placed in a collection tube. 5 ml of 1x PBS was pipetted into the column so labelled cells can go through and collected into the tube. Collected cells were centrifuged and placed at 1000 rpm for 5 minutes at 4°C and resuspended with EC growth media (Advanced DMEM supplemented (Gibco) with 100 µg/mL heparin (STEMCELL technologies), 100 µg/mL endothelial cell growth supplement (ECGS) (Merck Millipore), 20% FBS, 1 X glu-pen-strep (Gibco), 25 mM HEPES (Gibco)) and plated on 0.1% gelatin-coated (Merck Millipore) plates (Corning).

### 2.2.8 Organoids Culture

Uninduced membrane-targeted tandem dimer Tomato/membrane-targeted green fluorescent protein - LSL-*KRas*<sup>G12D/+</sup> *P53*<sup>fl/fl</sup> (MTMGKP) mice were used to sort the target cell populations and grow organoids. The sorter used was BD Aria Sorter Z6001 (BD Bioscience). 200,000 of sorted tumour initiating cells were resuspended in 100 µl of MTEC 3D media, which consists of DMEM-Ham's F-12 media (Gibco) supplemented with 15 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (Gibco), 8 mM L-Glutamine (Gibco), 1 x penicillin/streptomycin (Gibco), 10 µg/mL insulin (Sigma-Aldrich), 5 µg/mL 1 x-insulin/transferrin/selenium (ITS) (Gibco), 0.1 µg/mL Cholera Toxin (Sigma-Aldrich), 25 ng/mL EGF (Merck), 25 ng/mL bFGF (STEMCELL technologies), 30 µg/mL bovine pituitary extract (Gibco) and 5% FBS. Ad5-mSPC-Cre was added to this mix to a final concentration of 6 x 10<sup>7</sup> pfu/ml. The cells were incubated for 1 h at 37°C, 5% CO<sub>2</sub>, to allow for transformation. Cells were washed in PBS two to four times, to ensure virus particles were fully removed and were counted. Between 10,000 - 50,000 of these transformed cells were resuspended in 100 µl aliquots of Distal 3D media. This second media consists of DMEM-Ham's F-12 media supplemented with 15 mM HEPES, 8 mM L-Glutamine, 1x penicillin/streptomycin, 10% FBS and 5 µg/mL ITS. 50,000 - 100,000 support cells (typically mouse endothelial cells, obtained following protocol explained in Methodology section 1.2.5) were added to this mix. Equal volume of Matrigel was added to the media and 200 µl of this mix was seeded to the upper chamber of a 24 well 0.4µm pore transwell plate (Costar). The plate was placed in the incubator for 15 minutes to allow the Matrigel reduced growth factor (Matrigel)

(Corning) to solidify and 500  $\mu$ l of Distal 3D media was added to the lower chamber of the well, and the media was changed every couple of days.

## 2.3 Statistics

All statistical tests except those used in bioinformatics analysis of RNAseq data, which was performed in RStudio version 2023.03.0, were performed using GraphPad Prism (v9.5.1). For analysis of Kaplan Meier survival curves, the Log-rank (Mantel-Cox) test was employed. For statistical analysis comparing two groups of data unpaired t-test was used. For comparing more than two groups of data generated from in vivo experiments, like tumour volume, total tumour burden, growth rates and doubling time, together with some data obtained from the flow cytometry data was analysed using the ordinary one-way ANOVA for multiple comparisons for mean of each column was compared with the mean of every other column. PET imaging, data generated from flow cytometry and proteomics and targeted metabolomic data was analysed using the multiple unpaired t-test between two groups. Fisher exact test was used to compared categorical data. Statistical tests are mentioned in each figure legend. P values are shown as \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  and \*\*\*\* $p < 0.0001$ .

## Chapter 3 Optimising lineage-specific lung cancer organoid cell culture protocol.

### 3.1 Introduction

Metabolic alterations are considered to be a hallmark of cancer, playing a pivotal role in the uncontrolled growth and survival of tumour cells (Hanahan and Weinberg, 2011). For years, 2D cell culture has been the preferred model for studying cancer metabolism. However, 2D culture has several limitations, as it oversimplifies the complexity of tumour heterogeneity and tumour microenvironment (Dragic et al., 2022). One significant example of the limitations of 2D culture is observed in the *KRas*-driven NSCLC mouse model. Tumour cells in these models display distinct metabolic preferences when grown in 2D culture compared to in vivo conditions. In 2D culture, these cells prefer glutamine as a nutrient source, while in the tumour microenvironment, they rely on glucose-derived carbon to fuel the TCA cycle, due to the unphysiological and increase glutamine concentration in culture media. This discrepancy highlights the critical role of the tumour microenvironment in shaping cancer cell metabolism (Davidson et al., 2016).

The microenvironment of solid tumours is characterized by several factors which are not adequately addressed in 2D cultures. For instance, the dysfunctional blood flow, which alters perfusion, oxygen and nutrient availability. As a result, cancer cells must adapt their metabolic pathways to ensure survival under nutrient conditions which differ from the physiological ones. Cell proliferation is heavily dependent on nutrient availability, prompting cancer cells to undergo metabolic adaptations to cope with changes in the tumour microenvironment (Vander Heiden and DeBerardinis, 2017). Moreover, the tumour microenvironment is also modulated by cell-to-cell interactions, nutrient exchanges between cancer and stroma cells, mechanical constraints due to a modified extracellular matrix, and the nutrient consumption by non-tumour cells in the tumour microenvironment (Dragic et al., 2022). For example, tumour-infiltrating myeloid cells which consume around 30% of the total available glucose even though they are less abundant than cancer cells (Reinfeld et al., 2021).

To bridge the gap between 2D cultures and physiological conditions, researchers have turned to organoids 3D cultures in the recent years. Organoids are generated from tissue-derived adult stem cells embedded into a 3D matrix, typically Matrigel, allowing the cells to grow into self-organizing 3D structures that resemble the architecture of organs (Lancaster and Knoblich, 2014, Drost and Clevers, 2018). In some cases, organoids are seeded with stromal cells, providing a more realistic tissue stiffness representation of the tumour microenvironment. Moreover, cancer-derived organoids, developed from adult stem cells after genetic modification, have become valuable tools in cancer research. They maintain the original tumour features, the genetic and cellular heterogeneity present in the original tumour and offer a simpler and more accessible alternative to complex in vivo models commonly used in cancer metabolism research (Lancaster and Knoblich, 2014, Drost and Clevers, 2018). Furthermore, 3D models more closely resemble the in vivo environment as they can be co-cultured with mesenchymal, immune or endothelial cells, which are normally present in the tumour, but at the same time, simplifying the complexity of the tumour microenvironment in in vivo models, reducing confounding factors associated with in vivo models (Dragic et al., 2022).

Despite their advantages in cancer metabolism research, 3D organoid cultures also present some challenges. They are more expensive, time consuming and technically demanding compared to 2D cell culture. Obtaining a confluent culture or a sufficient number of cells for most techniques can be more challenging and time consuming than in 2D cultures. Moreover, co-culturing cancer cells with stromal cells can lead to difficulties in precisely controlling the location of different cell types and may complicate data analysis (Temple et al., 2022). Standardization of 3D culture techniques remains an ongoing challenge across laboratories. Factors such as cell seeding concentration, culture length, and seeding technique can influence the culture's outcome. In the case of spheroids, the seeding method can affect nutrient diffusion within the organoid, impacting the metabolic behaviour of the cells (Shreya Raghavan 2016).

One of the most challenging complications of 3D culture is the biochemical analysis. Most of the standardised analysis and protocols are designed for 2D cultures and they need to be adapted to 3D cultures. This adaptation, however, faces some difficulties. Firstly, nutrient diffusion is altered by the presence of

Matrigel, which reduces culture permeability, leading to reduced nutrient access for cells. Additionally, quantification and normalization of results in 3D cultures are more challenging compared to 2D cultures. For instance, 2D results tend to be normalised by the cell number, but the cell counting is a challenging task in 3D cultures as the organoid organisation itself, the presence of the Matrigel or the presence of support cells in the culture may make the use of microscopy difficult for this task. At the same time, the extensive cell-to-cell and cell-to-matrix interactions make it difficult to obtain a single-cell suspension for automated cell counting, which is commonly used in 2D cultures for normalization (Temple et al., 2022, Shreya Raghavan 2016). To overcome this problem, attempts have been made to count for the total amount of protein in the sample to give an estimation of the total number of cells in culture. However, Matrigel in culture can be misleading for this measure. Other have attempted to quantify “house-keeping” genes such as *GAPDH* or *B-ACTIN*, as they are constitutively expressed in most cell types. This is a better approximation than quantifying the total protein amount as the expression of the housekeeping gene and the expression of the gene of interest are equally affected by the culture conditions. The problem arises when cells face hypoxic or stressful conditions when house-keeping genes conditions are affected (Temple et al., 2022). Despite of its challenges, tumour organoids have been generated from a number of organs from both mouse and human.

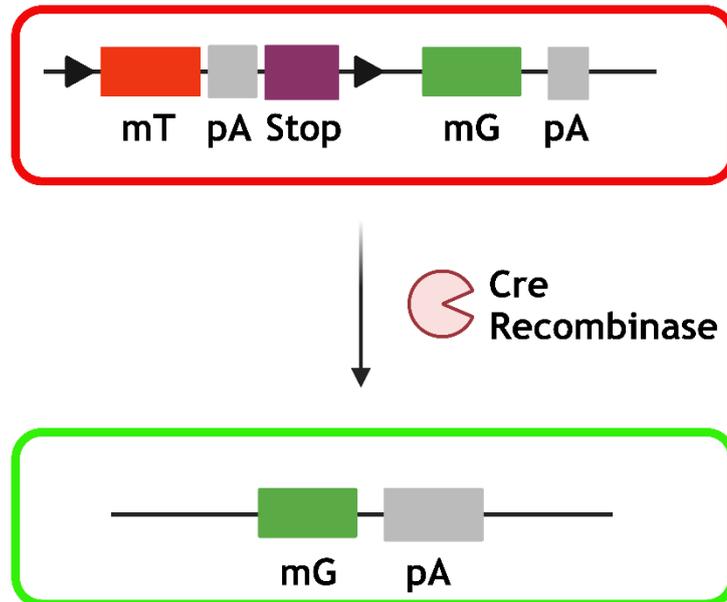
In the context of lung cancer research, lung tumour organoids (LTOs) have emerged as a valuable alternative to 2D in vitro models. LTOs accurately recapitulate the intertumoral and intratumoral heterogeneity of lung cancer, retaining the histology and genetic characteristics of the original cancer tissue while maintaining epithelial organization (Kim et al., 2019). As described in chapter 1, lung cancer can arise from different cell types in the lung epithelium, and they can determine the histology subtypes and certain tumour characteristics. There have been numerous attempts to grow organoids from each of these cells, not only for cancer research, but also for basic and respiratory disorders research.

There are several organoid systems which have been developed by different cell types from murine lungs. For instance, Jason R. Rock, 2009 isolated murine basal cells expressing NGFR and used them to study differentiation pathways in the

trachea, to screen drugs for controlling basal cell proliferation and to understand the effects of differentiated basal cells in patients suffering from asthma, chronic obstructive pulmonary disease and cystic fibrosis (Barkauskas et al., 2017). Another example are club cells, can be isolated from murine lungs based on CD24<sup>low</sup>, CD104 and CD49f expression, seeded in a 3D culture to investigate respiratory disorders such as asthma and chronic obstructive pulmonary disease (McQualter et al., 2010). Finally, alveolar cells, including AT2 cells and BASCs, isolated based on HTII280 or Sca-1 expression, respectively, have been used to study emphysema and idiopathic pulmonary fibrosis and to improve the understanding of progenitor cell function (Barkauskas et al., 2017) and to improve the knowledge of the progenitor cell function (Louie et al., 2022).

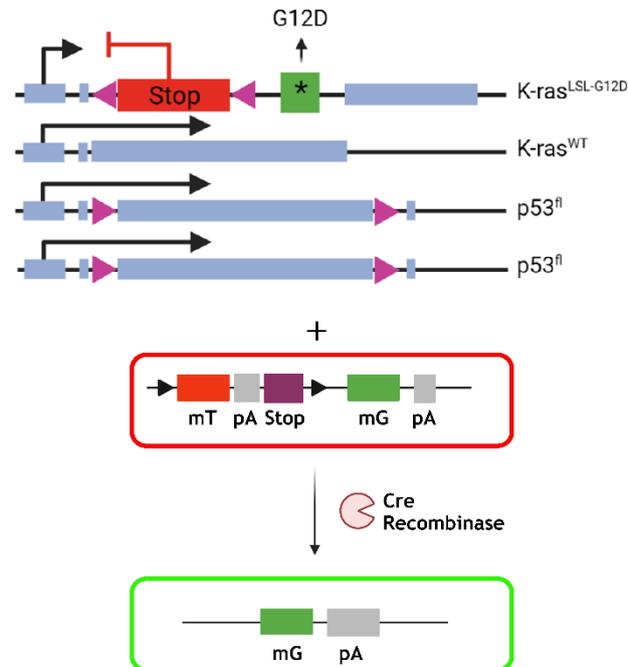
In addition to lung cell types derived from normal lung tissue, LTOs can be generated from GEMMs harbouring common gene mutations found in human lung cancer (DuPage et al., 2009). These GEMMs provide a valuable platform for studying the effects of specific genetic alterations on cancer metabolism. One popular genetic engineering system used in LTO generation is the *Cre/loxP* system, which allows conditional mutagenesis in mice.

This system involves inserting *loxP* sequences surrounding a gene of interest, with Cre recombinase being brought into the cells by virus vectors controlled by tissue-specific or temporally regulated promoters. Upon activation, CRE recombinase excises the *loxP*-flanked genes, generating knockouts or switching on genetic mutations (Hua Gu, 1994). mTmG mouse model is an example of this genetic modification system. This mouse model cells express membrane-targeted tdTomato (mT) emitting red fluorescence prior to CRE excision, and membrane-targeted enhanced green fluorescent protein (EGFP) emitting green fluorescence following CRE excision. This allows the visualization and distinction of recombined and non-recombined cells (Muzumdar et al., 2007). In this mouse model, the *loxP*-gene-*loxP* sequence is placed in the *Rosa26* locus. N-terminal membrane-tagged version of mT and a stop coding were inserted surrounded by two *loxP* sequences. N-terminal membrane-tagged version of EGFP was inserted distal to the second *loxP* sequence. This construction was cloned into an expression vector containing a CMV  $\beta$ -actin enhancer-promoter to guarantee a strong expression of this construction (Figure 3.1) (Muzumdar et al., 2007).



**Figure 3.1. Schematic representation of mTmG reporter gene set in mice.** Adapted from ref15. Based on (Muzumdar et al., 2007) mT reporter gene is excise after the expression of CRE recombinase as it is surrounded by *LoxP* sequences (triangle shapes). This drives the expression of mG gene which codifies for EGFP. pA polyadenylation sequences, mT membrane-targeted tandem dimer Tomato, mG membrane-targeted enhanced green fluorescent protein. Created by BioRender.com

Most importantly, these adult stem cells can be isolated from GEMMs harboring common gene mutations found in human lung tumours, such as *KRas*<sup>G12D</sup> or *P53* deletion. An example of this is Dost et al., 2020 using *K-Ras*<sup>LSL-G12D/+</sup> and *K-Ras*<sup>LSL-G12D/+</sup>; *P53*<sup>fl/fl</sup> AT2 activated cells (Dost et al., 2020) or Kim et al., 2005 using *K-Ras*<sup>LSL-G12D/+</sup> BASCs activated cells (Kim et al., 2005). In this project, we use a GEMMs which combines the mTmG mouse model with *K-Ras*<sup>LSL-G12D/+</sup>; *P53*<sup>fl/fl</sup> mouse model (MTMGKP). This mouse model consists of a *loxP-Stop-loxP* cassette inserted into intron 0 upstream of the transcriptional start site, which prevents the transcription of a *K-RAS*<sup>G12D</sup> allele before activation. This is an activating mutation at codon 12 of the *K-RAS* gene which exchanges a glycine for an aspartic acid. The *P53*<sup>fl/fl</sup> allele encodes the wild type *P53* genes, both surrounded by *loxP* sites flanking exons 2 and 10. Upon CRE activation both alleles of *P53* gene are deleted (Jackson et al., 2005).



**Figure 3.2. Schematic representation of both KP and mTmG systems combination in the MTMGKP mouse model.** Adapted from (Muzumdar et al., 2007). MTMGKP GEMM is characterised by the recombination of wild-type KRAS into KRAS<sup>G12D</sup> and the deletion of P53 after the expression of CRE recombinase, the excision of *LoxP* sequences (triangle shapes) and the suppression of the stop codon in the KRAS gene. mT reporter gene is excised after the expression of CRE recombinase as it is surrounded by *LoxP* sequences (triangle shapes). This drives the expression of mG gene which codifies for EGFP. Created by BioRender.com.

CRE recombinase is brought into the cell by virus vectors and in this case its expression is controlled by a lung cell-specific promoter, which allows us certain control over the cell type that is activated. In this project, we used AdenoCre virus vectors provided by University of Iowa Viral Vector Core Facility. We used Ad5-mSPC-Cre, which will only express Cre recombinase in AT2 and BASCs cells; Ad5-CC10-Cre, which will activate only club cells and BASCs; Ad5-KT5-Cre, which will express Cre on basal cells, and Ad5-CGRP-Cre, which activated neuroendocrine cells only (Chen et al., 2014).

### 3.1.1 Hypothesis and Aims

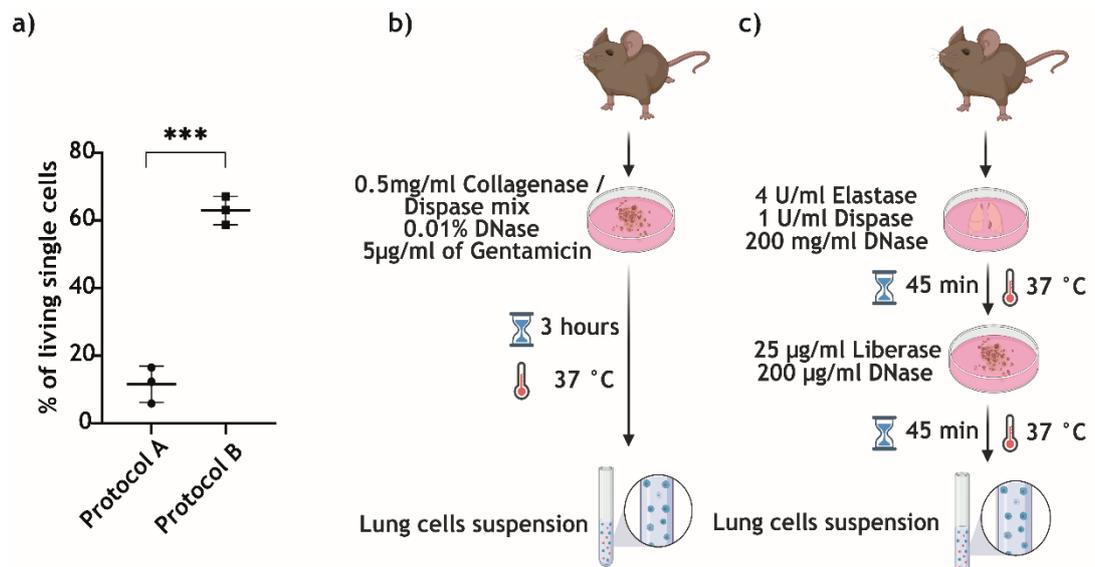
The aim of this chapter is to develop a protocol for creating lineage-restricted organoids, specifically LTOs. The approach involves sorting normal lung cells-of-origin into their distinct cell types, inducing malignant transformation *in vitro*, and subsequently growing the activated cells in 3D cultures. The primary objective of this protocol is to establish a more controlled environment that

closely resembles *in vivo* conditions, allowing for the assessment of the contribution of lung cancer cell-of-origin to metabolic heterogeneity.

## 3.2 Results

### 3.2.1 Epithelial lung cell isolation

To achieve a single cell suspension from a murine lung, two tissue digestive protocols were tested. Protocol A was regularly used in our research group for tumour dissociation. It consists in a three-hour incubation of a previously dissected lung in a PBS-based solution containing 0.5 mg/ml collagenase/dispase mix, 0.01% DNase, 5 µg/ml of gentamicin, which prevents the growth of both gram-positive and gram-negative bacteria *in vitro*, and 10 µM of RHO/Rock pathway inhibitor, which prevents the stem cell apoptosis when lungs are disassociated to single cells (Barcelo et al., 2023) (Figure 3.3b). Protocol B was adapted from (Nakano et al., 2018) and it consists of two separate incubations with a different enzyme combination (Figure 3.3a). In both cases, cells were stained with Zombie Yellow fixable dye as described. Protocol B showed a major increase in the percentage of living cells (Figure 3.3a), so it was used for all subsequent experiments.



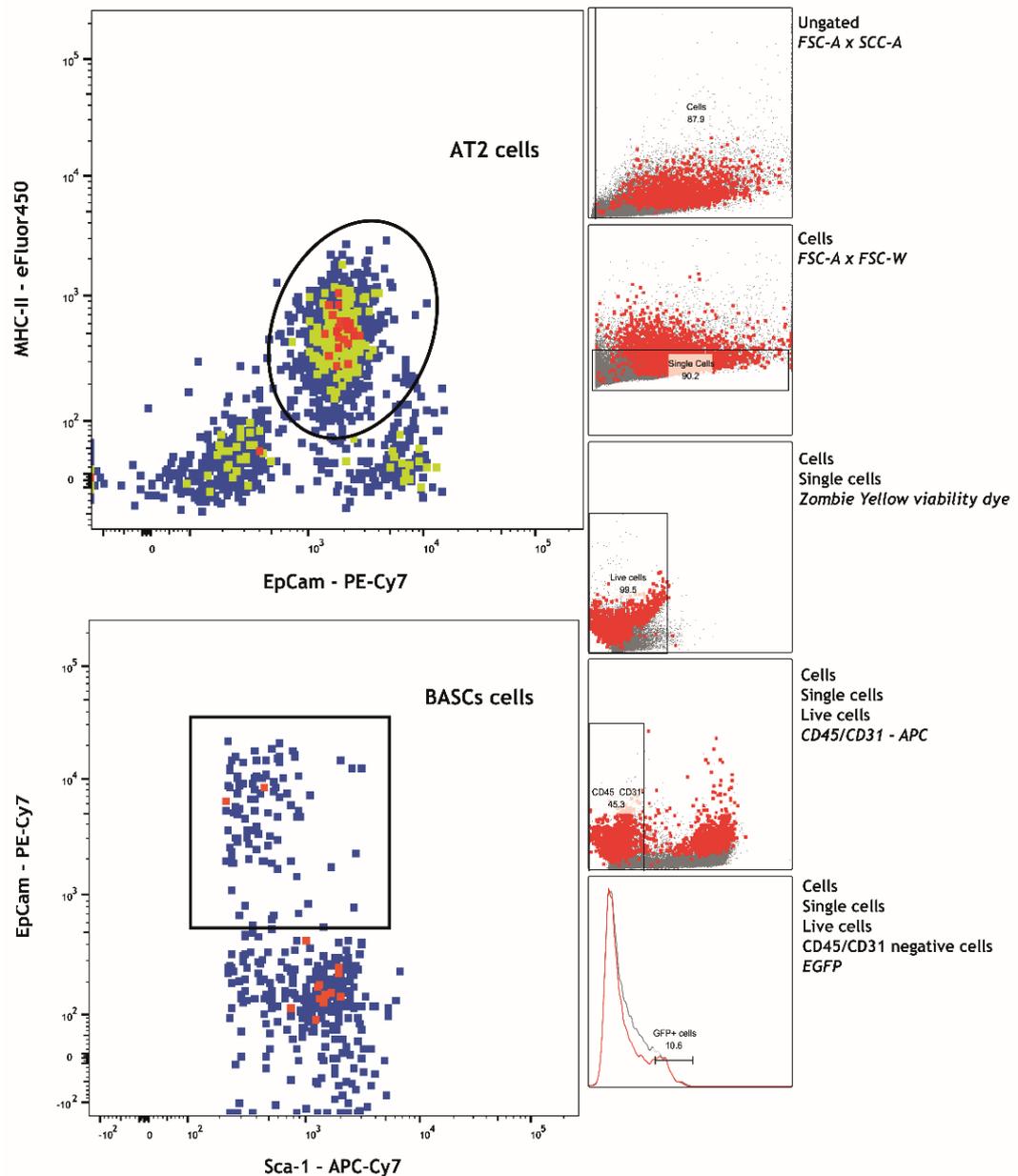
**Figure 3.3.** Lung dissociation protocol B shows a significant increase in the percentage of living cells. (a) Dot plot of living cells percentage, (b) schematic summary of protocol A, (c) schematic representation of protocol (b). Each data point in figure (a) represents the result of an experimental trial ( $n=3$  for each protocol). Error bars in (a) are generated from mean  $\pm$ SD. Statistical analysis using unpaired two-tailed t-test performed in GraphPad Prism was used to

determine a statistically significant difference between the two protocols. \*\*\*  $p \leq 0.001$ . Created by BioRender.com

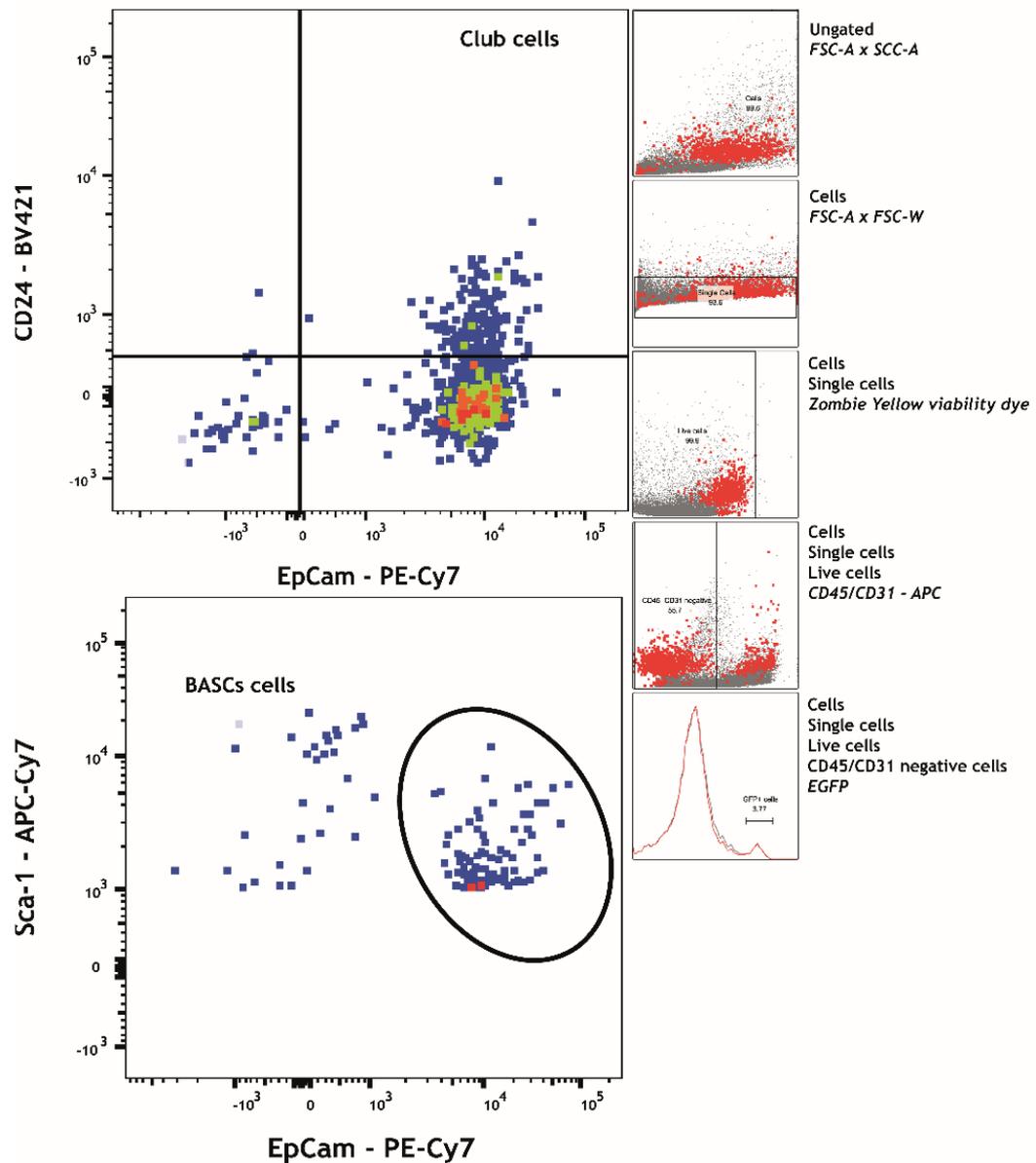
### 3.2.2 Detection of lung cancer AT2, BASCs, Club and neuroendocrine cells with flow cytometry

AT2, BASCs, club and neuroendocrine cells are distributed all over the murine airways (Rock and Hogan, 2011). They express specific cell surface markers which allow us to identify and isolate them (Barkauskas et al., 2017). They are negative for CD31 and CD45 markers of endothelial cells (Lertkiatmongkol et al., 2016) and leukocytes (Josef M. Penninger, 2001), respectively. They are all epithelial cells, thus they all express epithelial cell adhesion molecule (EpCAM) (Trzpis et al., 2007). AT2 cells are defined by the expression of major histocompatibility complex class II (MHCII) (Hasegawa et al., 2017). BASCs isolation is based on the expression of sca-1 (Lee et al., 2014). Club cells isolation is based on the surface expression CD24 (McQualter et al., 2010) and, neuroendocrine cells are defined by the expression of CD56 (Yang et al., 2019).

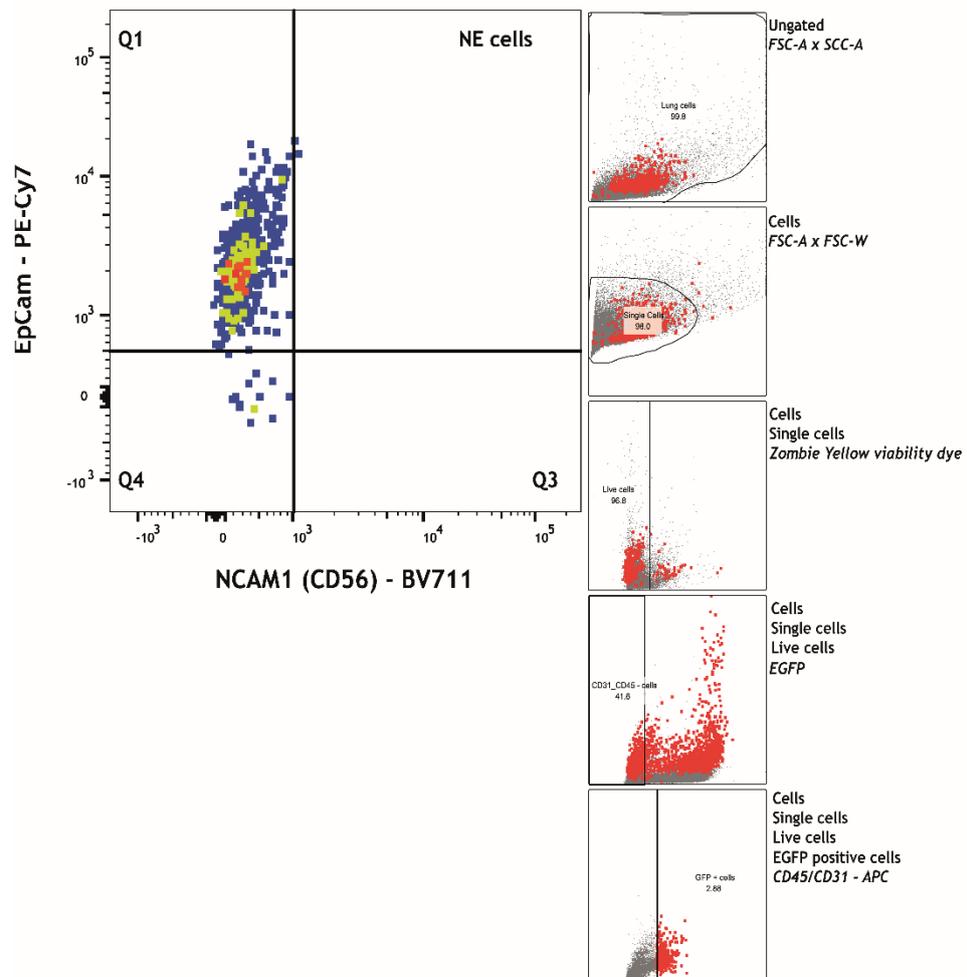
To develop the organoid protocol, cells-of-origin from MTMGKP mice need to be isolated based on their expression of surface markers, activated in vitro to express *KRas* G12D mutation, delete *P53* and express EGFP. This activation is conducted using the Ad5-mSPC-Cre, Ad5-CC10-Cre and Ad5-CGRP-Cre adenovirus vectors described in section 3.1 which target AT2 and BASCs cells, club and BASCs cells and neuroendocrine cells, respectively (Chen et al., 2014). EGFP signal will be used in the future to visualise these organoids. Given the knowledge of specific cell-of-origin cell surface markers available in the literature, we used the mTmG mouse model described in section 3.1 to confirm these viruses are targeting the expected cell types and to test their specificity. Mice were induced with  $10^8$  pfu/ml with each of these vectors which will induce the expression of *EGFP* in either AT2 and BASCs, club and BASCs cells or neuroendocrine cells following the protocol described in section 2.1.3. Mice were culled a month after the induction and lungs were digested following the protocol B. Cells were stained as explained and appropriate controls as explained were applied. The resultant green fluorescence signal can be detected by flow cytometry and data was analysed using FlowJo.



**Figure 3.4. Gating strategy for lung epithelial cells panel used to identify AT2 cells and BASCs.** Alive cells were gated due to their negative staining for zombie yellow viability dye. Then, blood and immune cells were distinguished from epithelial cells due to their expression of CD45 and CD31 (APC staining). EGFP positive cells were identified based on their slightly increase in green fluorescence. Finally, both AT2 cells and BASCs were identified based on the expression of EpCam (PE-Cy7) an epithelial cell marker. AT2 cells are identified based on the expression of MHC-II (eFluor450), and BASCs cells are isolated based on the expression of Sca-1 (APC-Cy7). EGFP enhanced green fluorescence protein, EpCAM epithelial cell adhesion molecules, MHC major histocompatibility class. Grey dots represent ungated cells. Red dots represent gated cells.



**Figure 3.5. Gating strategy for lung epithelial cells panel used to identify club cells and BASCs.** Alive cells were gated due to their negative staining for zombie yellow viability dye. Then, blood and immune cells were distinguished from epithelial cells due to their expression of CD45 and CD31 (APC staining). EGFP positive cells were identified based on their slightly increase in green fluorescence. Finally, both club cells and BASCs were identified based on the expression of EpCam (PE-Cy7) an epithelial cell marker. Club cells are identified based on the expression of CD24 (BV421), and BASCs cells are isolated based on the expression of Sca-1 (APC-Cy7). EGFP enhanced green fluorescence protein, EpCAM epithelial cell adhesion molecules, BV brilliant violet. Grey dots represent ungated cells. Red dots represent gated cells.



**Figure 3.6. Gating strategy for lung epithelial cells panel used to identify neuroendocrine cells.** Alive cells were gated due to their negative staining for zombie yellow viability dye. Then, blood and immune cells were distinguished from epithelial cells due to their expression of CD45 and CD31 (APC staining). EGFP positive cells were identified based on their slightly increase in green fluorescence. Finally, both NE cells were gated based on the expression of EpCam (PE-Cy7) an epithelial cell marker and the expression of CD56 (BV711). EGFP enhanced green fluorescence protein, EpCAM epithelial cell adhesion molecules, BV brilliant violet. Grey dots represent ungated cells. Red dots represent gated cells.

Flow cytometry data shows that 55.2% of EGFP lung cells from Ad5-mSPC-Cre induced mice were positive for EpCAM and MHCII (Figure 3.4 and Table 3.1), which has been defined as AT2 cells and 9.5% was positive for EpCAM and Sca1, which are defined as BASCs (Figure 3.4 and Table 3.1). 23.2% of GFP lung cells from Ad5-CC10-Cre induced mice were positive for EpCAM and CD24 or club cells (Figure 3.5 and Table 3.1) and 58.1% was positive for EpCAM and Sca1 (Figure 3.5 and Table 3.1). Finally, flow cytometry data showed that only 0.43% of GFP positive cells from mice induced with Ad5-CGRP-Cre were positive for CD56 (Figure 3.6 and Table 3.1).

Table 3.1. Summary of the percentage of each cell type identified in every experiment for each mouse model used. E: experimental replicate.

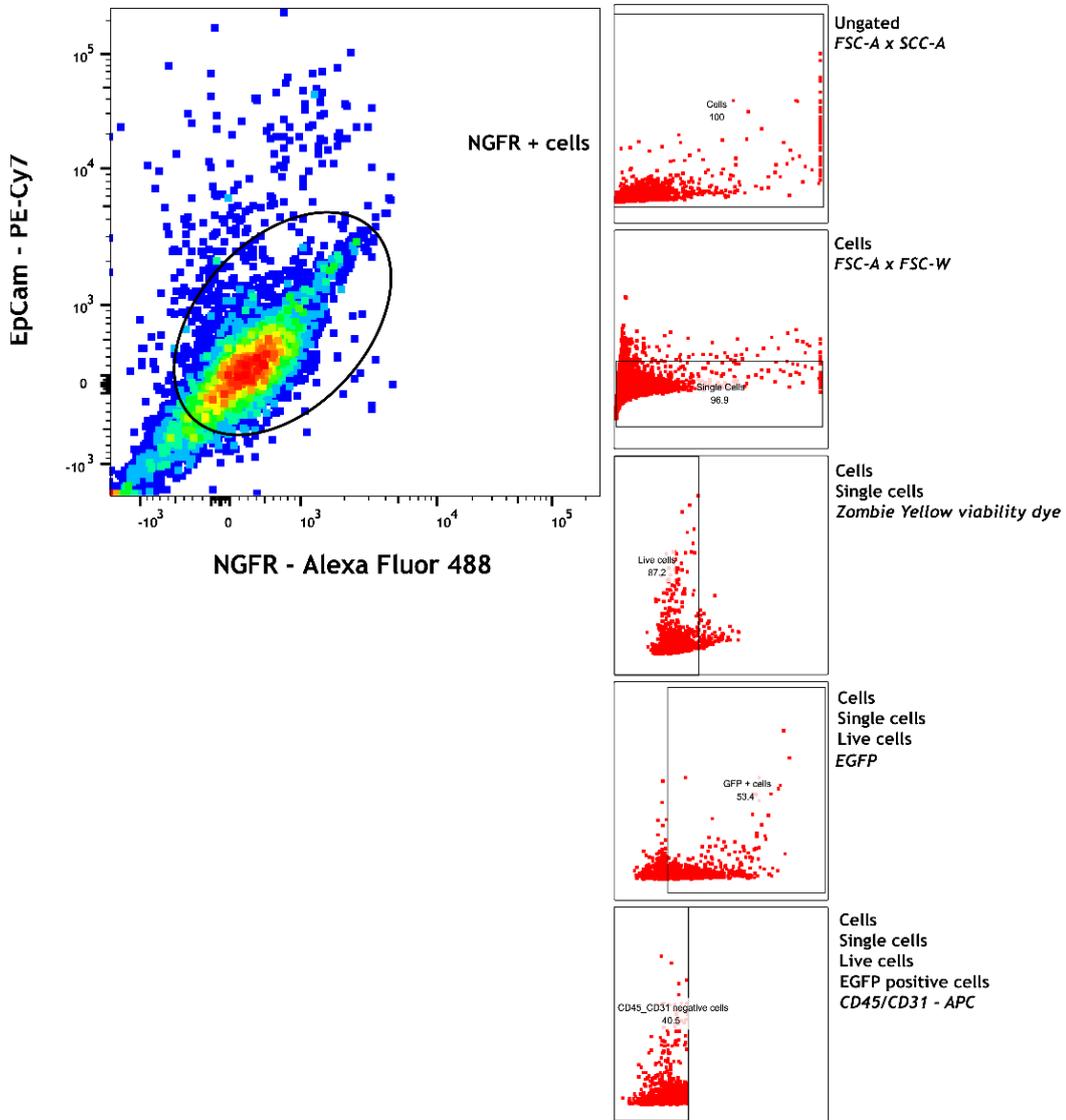
Virus	Ad5-mSPC-Cre		Ad5-CC10 Cre		Ad5-CGRP-Cre
Cell type	AT2 cells	BASCs	Club cells	BASCs	NE cells
E1: % of EGFP cells	55	10.6	27.9	57.9	0.1
E2: % of EGFP cells	50.2	7	17.5	73.1	0.35
E3: % of EGFP cells	60.5	11	24.2	43.3	0.49
Mean (%)	55.2	9.5	23.2	58.1	0.43
Total (%)	64.7		81.3		0.43

To generate lineage-restricted organoids from different lung cell types, we initiated the assessment of AdenoCre virus vectors' cell specificity using the mTmG reporter mouse model. As explained in section 3.1, cells expressing the Cre recombinase emit green fluorescence due to the GFP expression upon activation. This fluorescence can be detected through flow cytometry, providing valuable insights into cell specificity. Among the AdenoCre virus vectors tested, Ad5-mSPC-Cre demonstrated to be the quite specific, with an average of 64.7% of EGFP positive cells defined as AT2 cells and BASCs. Ad5-CC10-Cre also exhibited high specificity for club cells BASCs cells. Ad5-CGRP-Cre specificity proved to be lower than expected as a very few amounts of NE cells were detected.

### 3.2.3 Detection of lung cancer basal cell-of-origin with Flow Cytometry

Basal cells are located in the murine trachea (Kyung U. Hong, 2003). We reached out to Dr Christine Brainson from the University of Kentucky for collaboration and support in cell isolation and organoid culture. She shared with us a trachea tissue dissociation protocol described in section 2.2.2, as the isolation method described was not efficiently isolating basal cells mTmG mice induced with  $1 \times 10^8$  pfu/ml of Ad5-KT5-Cre, which targets basal cells specifically, were induced

intranasally with Ad5-KT5-Cre. This GFP signal was detected by flow cytometry. This data was analysed by FlowJo. We saw that 87.4% of GFP cells were positive for EpCAM and NGFR being these cells defined as murine basal cells, displaying a good specificity of Ad5-KT5-Cre virus for basal cells (Figure 3.7 and Table 3.2).



**Figure 3.7. Gating strategy for lung epithelial cells panel used to identify basal cells.** Alive cells were gated due to their negative staining for zombie yellow viability dye. Then, blood and immune cells were distinguished from epithelial cells due to their expression of CD45 and CD31 (APC staining). EGFP positive cells were identified based on their slightly increase in green fluorescence. Finally, both basal cells were gated based on the expression of EpCam (PE-Cy7) an epithelial cell marker and the expression of NGFR (Alexa Fluor 488). EGFP enhanced green fluorescence protein, EpCAM epithelial cell adhesion molecules. Grey dots represent ungated cells. Red dots represent gated cells.

**Table 3.2.** Summary of the percentage of basal cell identified in every experiment for Ad5-KT5-Cre mouse model.

<b>Virus</b>	<b>Ad5-KT5-Cre</b>
<b>Cell type</b>	Basal cells
<b>E1: % of EGFP cells</b>	98.7
<b>E2: % of EGFP cells</b>	78
<b>E3: % of EGFP cells</b>	85.6
<b>Mean (%)</b>	87.4

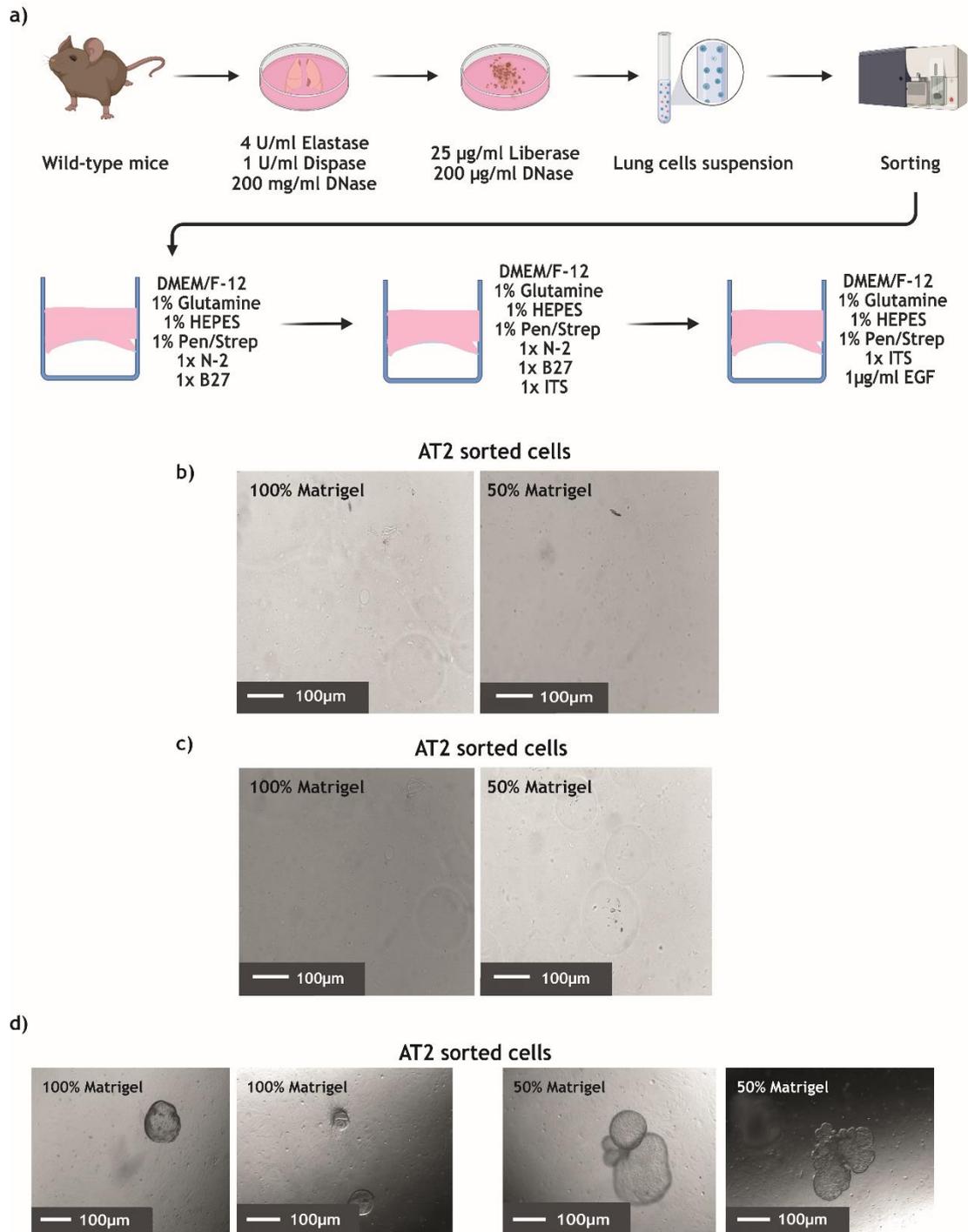
### **3.2.4 Generation of lung organoids from lung epithelial cells**

The Beatson Institute and Owen Sansom's research group is known for working with murine colon cancer organoids culture. Given the lack of knowledge and experience of our research group in working with 3D cultures and the need of setting up the simplest protocol as possible, we first attempted to adapt the crypt colon organoids culture to lung epithelial cells.

Murine lungs were dissected following the protocol B described above. The resultant single cell suspension stained for AT2 surface markers and these same cells were sent for sorting by selecting CD45 and CD31 negative cells and EpCAM and MHCII positive cells. Cells were collected in medium 1 which consists of DMEM F/12 with 1% glutamine, 1% Penicillin/Streptomycin and 1% HEPES and 0.05% bovine serum albumin (BSA) and strained through a 70  $\mu$ m filter. Then, 500,000 cells were resuspended in 50  $\mu$ l of Matrigel (100%) or in 25  $\mu$ l of Matrigel mixed with 25  $\mu$ l of medium 1 (50%) and seeded on a 24-well plate well. The plate was placed at 37°C upside down for 10 minutes for the Matrigel to solidify. Finally, 500  $\mu$ l of medium two, which consists of medium one with 1x N2, 1x B27 and 40 ng/ml of recombinant murine Noggin were added to the well (Figure 3.8a). In all cases, no organoids were seen in culture regardless of the Matrigel stiffness as seen in figure 3.8b.

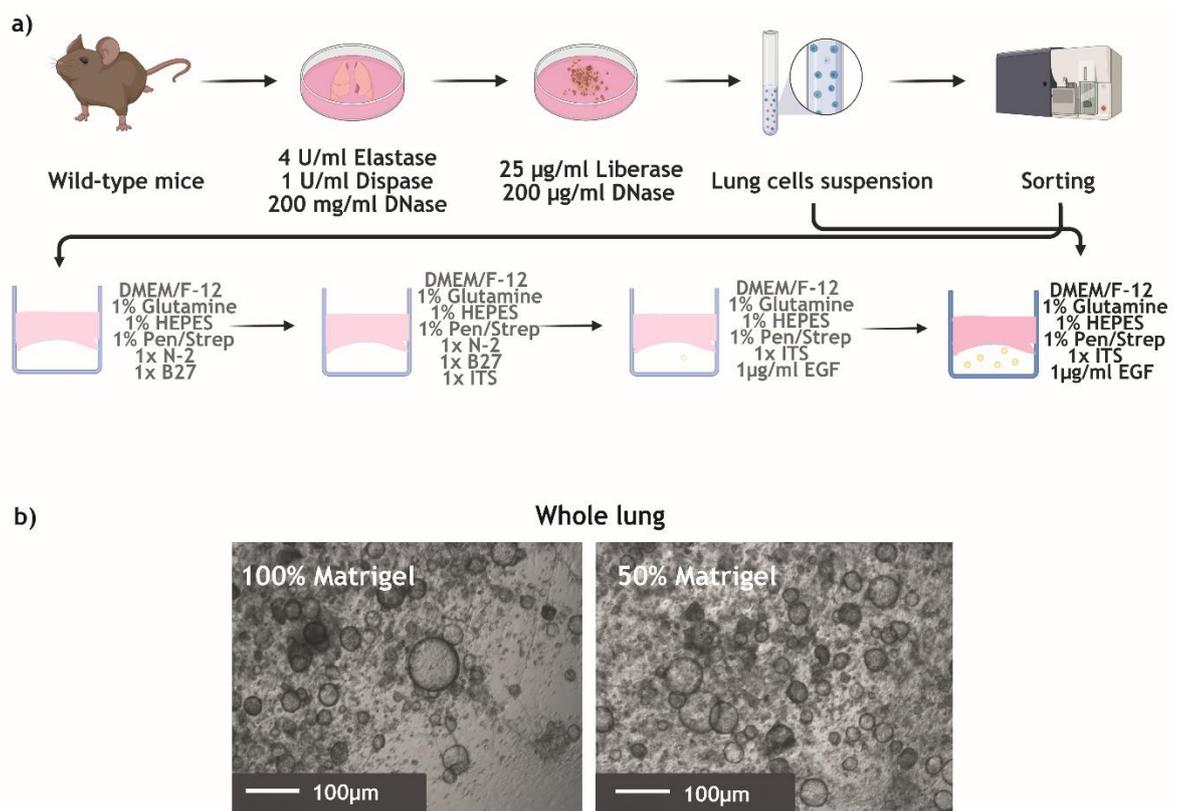
We tried to improve this protocol by adding 5 µg/ml of ITS as indicated in the bibliography (Barkauskas et al., 2017, Choi et al., 2016). Different Matrigel concentration as detailed above were tried when seeding the AT2 sorted cells. No organoids were seen in growing in vitro as seen in figure 3.8c.

After some literature research N2 and B27 were substituted for 1µg/ml of EGF and noggin was replaced by 10 µM of ROCK/RHO inhibitor to maximise the survival and the proliferation of epithelial cells. As a result, no more than five organoids with an irregular shape were seen growing in the petri dish (Figure 3.8d). There were no changes in the number of organoids or the shape of them over the next month.



**Figure 3.8. AT2 sorted cells 3D seeding strategies are not able to sustain the organoids growth and proliferation in vitro.** (a) Schematic representation of the protocol that has been followed to isolate and sort AT2 cells and the different seeding strategies followed, Created with BioRender.com (b) representative pictures of the 3D cultures established with the first culturing protocol with two different Matrigel stiffness, (c) representative pictures of 3D cultures established with the second detailed culturing protocol with two different Matrigel stiffness, (d) representative pictures of 3D cultures established with the third culturing protocol with two different Matrigel stiffness. The incorporation of EGF in the culture allow some AT2 cells to grow in 3D even though the number of organoids seen in culture was low and their survival very limited. Created by BioRender.com

At this point, we suspected we were missing some elements in this protocol. To test this theory, wild-type murine lungs were dissected following the protocol B detailed above. These cells were not sorted, instead 500,000 of total lung cells were resuspended in 100% or 50% Matrigel and seeded on a 24-well plate well. The plate was placed at 37°C upside down for 10 minutes for the Matrigel to solidify. Finally, 500 µl of DMEM F/12 with 1% glutamine, 1% Penicillin/Streptomycin, 1% HEPES, 5 µg/ml of ITS and 1 µg/ml of EGF were added to the well (Figure 3.9a). Within four weeks, several spherical organoids grew in culture, as seen in figure 3.9b. We assume that the epithelial cells are forming the organoids and some other cell types found in the lung like endothelial, immune or mesenchymal cells are also present in this culture as the black spots and shadows in the background. As a result, a confluent and proliferating organoid culture was seen in vitro for the first time.



**Figure 3.9. Seeded whole lung single cell suspension show a successful organoid culture with different Matrigel stiffness.** (a) Schematical representation of the protocol that has been followed to seed lung cells, prior to cell sorting and the medium used created with BioRender.com (b) representative pictures of the 3D cultures established with the first culturing protocol with two different Matrigel stiffness. The seeding of all the cell types obtained after lung disassociation results in an organoids confluent culture. Created with BioRender.com.

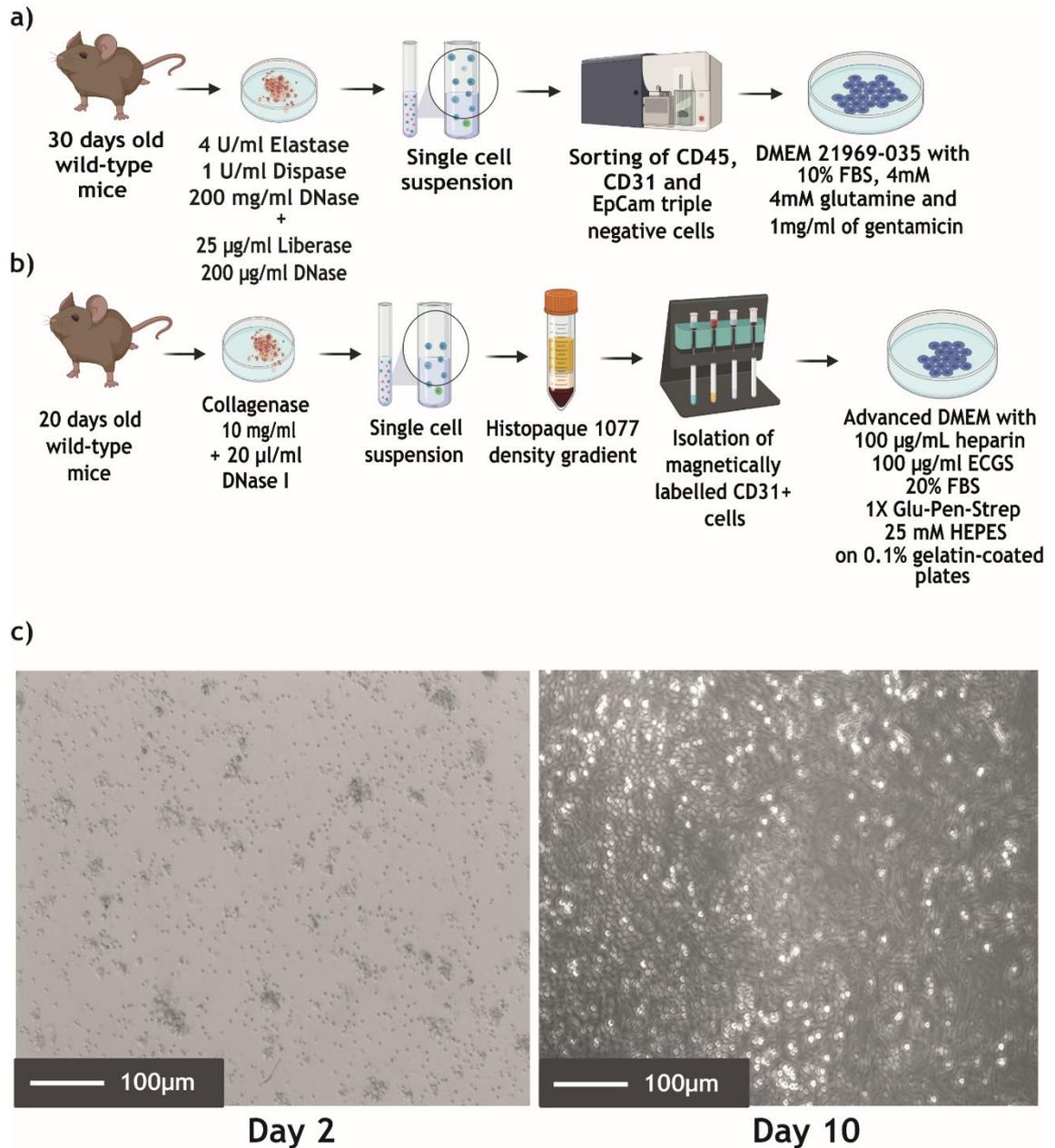
### 3.2.5 Isolation of support cells

Given the successful growth observed after the seeding and co-culture of several cell types in vitro, we contacted with Dr Christine Brainson from the University of Kentucky and Dr Joo-Hyeon Lee from Cambridge Stem Cell Institute as they regularly use this cell culture model. They suggested the need of a support cell system for the organoid growth of a single lineage cells in culture.

The point of this culture is to isolate AT2, club, basal and BASCs using cell sorting and culture them separately to study their metabolic differences in a controlled environment. Therefore, we need to apply sorting to isolate these cell types. In order to have the cleanest and simplest culture system as possible, support cells needed to be identified and isolated in the first place.

We were first advised to isolate murine lung mesenchymal cells from 4 week old WT mice by sorting negative cells for CD31, CD45 and EpCAM (Figure 3.10a). By doing this we discarded endothelial, hematopoietic and epithelial cells, respectively. The remaining cells are called mesenchymal cells by Dr Joo-Hyeon Lee, and they are part of the lung connective tissue (Leeman et al., 2019). The number of these cells obtained after sorting was relatively small. On average we obtained between 3,000 and 10,000 cells when sorting five pairs of murine lungs. They were seeded as detailed in the summary below (Figure 3.10a) and we observed that most of them did not survive in culture and the alive remaining cells were not proliferating. Our aim was to seed these cells so they could proliferate in vitro and create enough stocks to sustain the organoid culture in the research group. Consequently, we stopped working with this support cell model as we did not obtain enough cells to support the organoids culture.

We then decided to use endothelial cells as support cells (Lee et al., 2014). We isolated them from 20 day old WT mice following the procedure described in figure 3.10b. We obtained around 500,000 cells in average when sorting five pairs of murine lungs. The culture was 100% confluent after 10 days in culture (Figure 3.10c) These cells were growing following freezing and thawing.

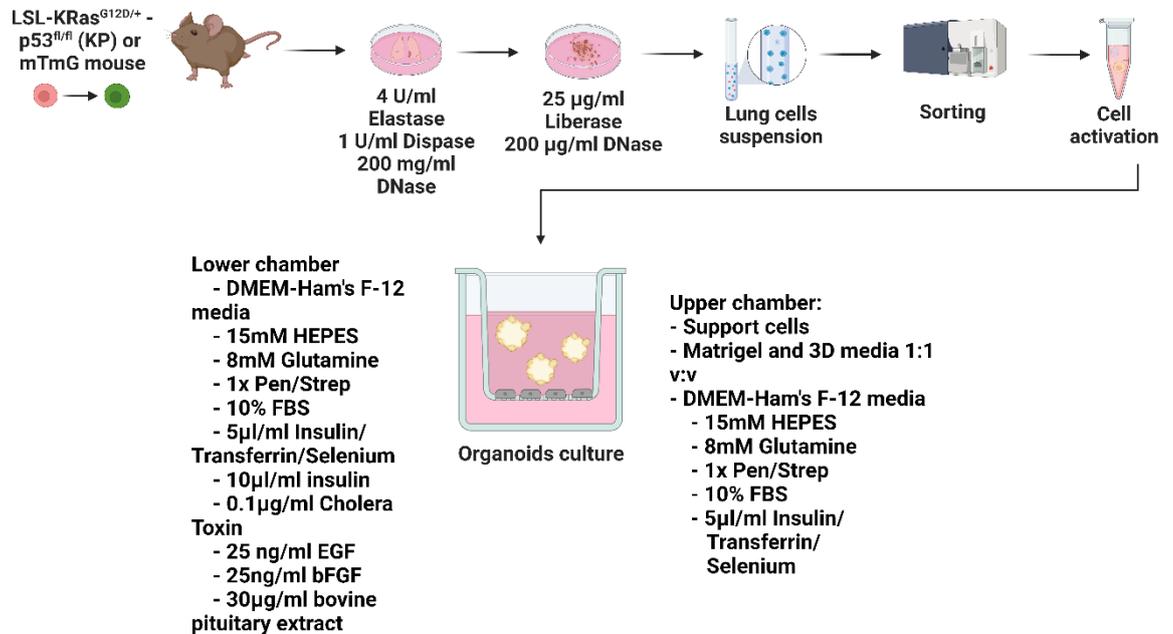


**Figure 3.10.** Endothelial cells proliferate and grow faster in culture than mesenchymal cells and therefore they are used as the most optimal cell support system for 3D lung culture. (a) schematic representation of the process followed for mesenchymal cell isolation and culture, (b) schematic representation of the protocol used to isolate and culture of murine primary endothelial cells, (c) representative picture of an endothelial cell culture 2 days post-seeding and 10 days post-seeding. Cells can grow from 15% confluency to a 100% confluent cell culture. Created by BioRender.com

### 3.2.6 AT2 cells organoid culture

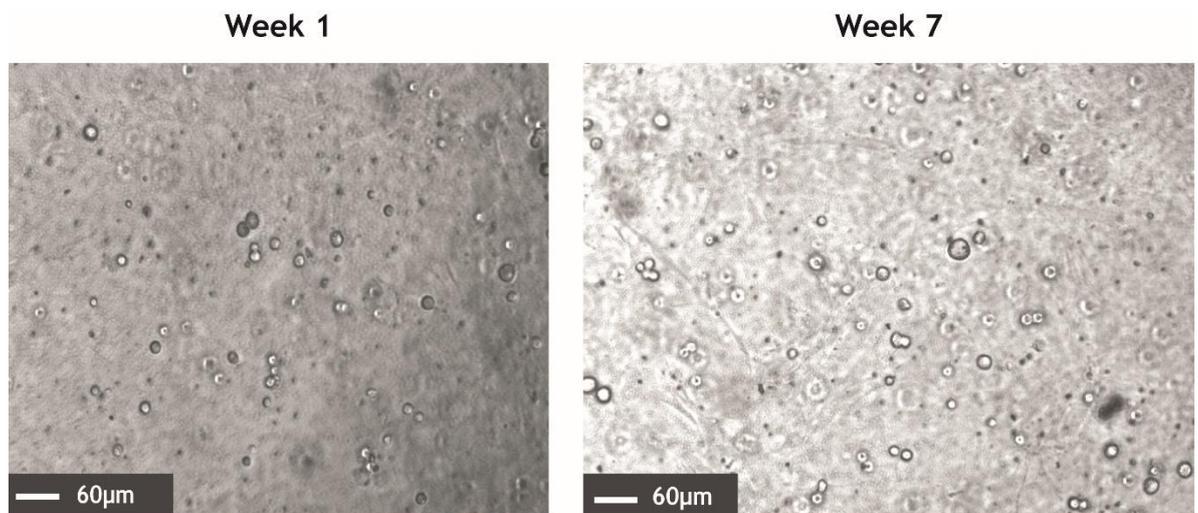
Given our lack of experience with this culture system, Dr Christine Brainson and Dr Joo-Hyeon Lee shared with us an organoid culture protocol, which have been modified and it is detailed in section 2.2.8 and summarised in Figure 3.11.

Originally between 10,000 to 50,000 of lung endothelial cells are resuspended together with 50,000 to 100,000 endothelial support cells in 100  $\mu$ l of Distal 3D media and growth factor reduced Matrigel 1:1 v:v. As a result, we observed AT2 organoids in vitro although they did not increase their size with time (Figure 3.12)



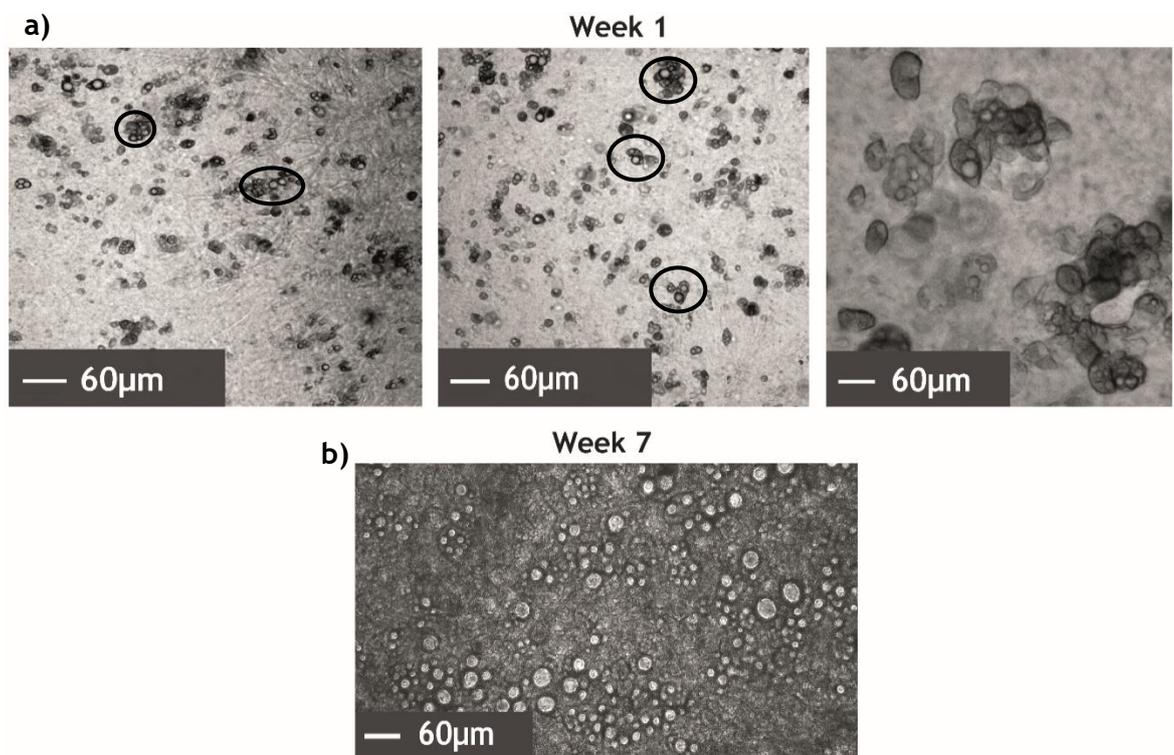
**Figure 3.11. Schematic representation of the protocol followed to isolate, transform and seed AT2 cells in 3D cultures.** Cells are dissected following protocol B described above, stained and sorted based on the expression of extracellular makers. The cell activation is performed in vitro with a virus vector following by the seeding of the cells in the environment described.

Created by BioRender.com



**Figure 3.12.** Organoids are growing in culture as a result of complementing the media with FGF, cholera toxin and insulin and adding support cells. Cells were seeded in a concentration of 50% Matrigel diluted in 50% medium. As a results, organoids survive in culture after activation but neither the number of organoids or their size grew over the weeks.

Matrigel stiffness can affect organoid growth as it may prevent them to increase their size if the space is limited. For that reason, we seeded AT2 cells in lower Matrigel stiffness of 1:3 v:v with Distal 3D media. 1:3 v:v ratio of Distal 3D volume and Matrigel respectively, seems to be the most appropriate to guarantee the organoid survival and the size increase after 8 weeks in culture (Figure 3.13).



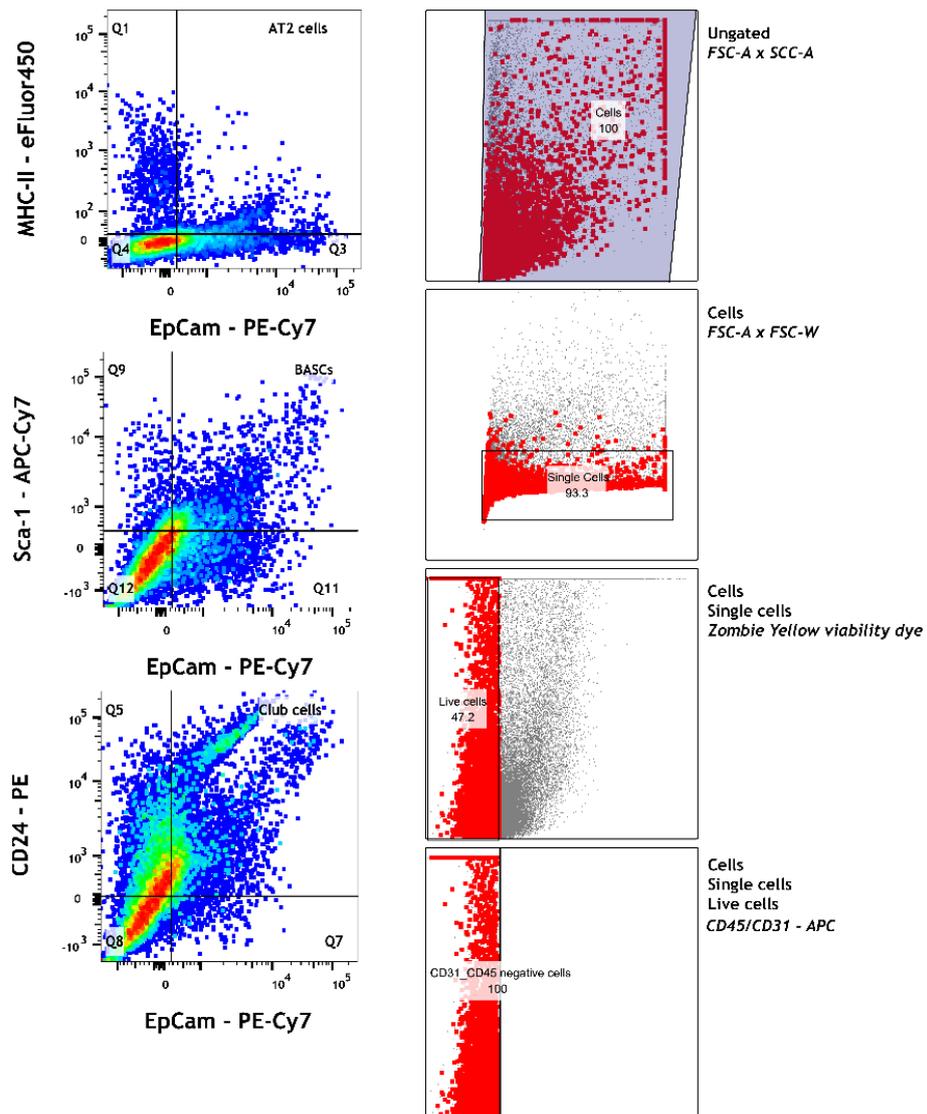
**Figure 3.13.** AT2 cells organoids are growing in culture as a result of complementing the media with FGF, cholera toxin and insulin and adding support cells and they expand themselves in vitro. Cells were seeded in a concentration of 25% Matrigel diluted in 75% medium. As a results, organoids survive in culture after activation (a) and proliferate until the culture becomes confluent around week 7 (b). Circles represent small organoids structures.

### 3.2.7 Protocol checkpoints

Given the modification introduced in Dr Christine Brainson's and Dr Joo-Hyeon Lee's protocol, we aim to introduce some checkpoints. The one described in this section aims to understand if these tumour-initiating cells undergo spontaneous

and quick differentiation after sorting isolation which would not allow them to be transformed by cell-specific viruses. The level of differentiation of these cells-of-origin was checked after the disassociation protocol B which differs from Dr Christine Brainson and Dr Joo-Hyeon Lee dissociation protocol. Lungs from 8-15 weeks old mice were dissected, digested and AT2, club and BASCs cells sorted. They were seeded in a 24 well plate previously coated with 0.1% gelatine. They were culture in MTEC 3D medium with 10  $\mu$ M of ROCK/RHO inhibitor. After 10 days in vitro cells were stained and analysed using flow cytometry.

As observed in figure 3.14 and table 3.3, AT2 cells are predicted to be the ones with a higher differentiation as only 5.9% of stained cells in average retained the AT2 cells status (EpCAM and MHCII positive). BASCs and club cells, on the contrary, seem to retain the undifferentiated state: 21.6% of stained cell still retain the BASCs phenotype and 33.8% of stained cells were positive for club cells phenotype.



**Figure 3.14. Gating strategy for lung epithelial cells panel used to identify AT2, BASCs and club cells, respectively.** Alive cells were gated due to their negative staining for zombie yellow viability dye. Then, blood and immune cells were distinguished from epithelial cells due to their expression of CD45 and CD31 (APC staining). Finally, both basal cells were gated based on the expression of EpCam (PE-Cy7) an epithelial cell marker and the cell surface marker expression of MHC-II (eFluor450), Sca-1 (APC-Cy7) and CD24 (PE). EpCAM epithelial cell adhesion molecules. Grey dots represent ungated cells. MHC major histocompatibility complex. Red dots represent gated cells.

**Table 3.3.** Summary of the percentage of each cell type identified in every experiment after 10 days in culture.

Cell type	AT2	BASCs	Club cells
<b>E1: % of total cells</b>	6.1	14.1	27.3
<b>E2: % of total cells</b>	3.9	27.7	38.3
<b>E3: % of total cells</b>	7.6	23.1	35.8
<b>Mean</b>	5.9	21.6	33.8

### 3.3 Discussion

#### 3.3.1 Epithelial lung cell isolation

Two lung cell isolation methods were tested. Protocol A was the protocol established in our research group for tumour cell disassociation but when used for normal lung tissue disassociation the number of alive cells obtained was low. This protocol contained a 3-hour incubation, although results were similar when shorter incubations were tested. This may indicate that the enzyme combination is relevant to obtain a significant number of alive cells. However, this effect was not explored further.

#### 3.3.2 Virus vectors specificity

Tumour-initiating cells were detected using flow cytometry. The gating strategy used first detected single cells. Then zombie yellow viability dye allows us to distinguish between alive and dead cells, as this dye can penetrate only into the cells with a compromised membrane. Therefore, negative cells for zombie yellow dye will be alive cells. Finally, CD31 and CD45 are markers for endothelial cells and blood cells respectively. Negative cells for this staining are then selected. Gating has been performed using fluorescence minus one (FMO) controls.

Ad5-mSPC-Cre vector is specific for AT2 cells and BASCs, although there is a significant percentage of EGFP positive cells which do not fall into AT2 or BASCs

cells classification. This effect is not likely to be explained by the cell accessibility to the virus. SPC cells are located in the BADJ and the alveolar space, which is the most abundant epithelium in the lung (Rock and Hogan, 2011), and there is no other identified cell type which expresses SPC. Therefore, this could be explained by perhaps a lower specificity from the Ad5-mSPC-Cre virus.

Ad5-CC10-Cre and Ad5-KT5-Cre have the highest rate of specificity. Ad5-CC10-Cre targets club and BASCs turning on *Egfp* gene but does not hit 100% of CC10 positive cells. One possible explanation for this discrepancy could be the existence of different subtypes of club cells, all expressing CC10, making them likely targets for the Cre recombinase expression upon virus vector activation, and thus leading to EGFP expression. However, this population of cells may not be defined by EpCAM and CD24 membrane expression, resulting in the observed variations in the EGFP positive cells (Barkauskas et al., 2017). Moreover, it can also be explained by the anatomy and cellular distribution of the lung. CC10 expressing cells are also distributed all along the airways. CC10 expressing cells are located in the trachea, the bronchi and in the alveolar space whereas SPC positive cells, both BASCs and AT2 cells, are only located in the alveolar space and BADJ (Rock and Hogan, 2011). All this together may contribute to explain why most of EGFP positive cells after activation are not EpCAM and CD24. Moreover, Ad5-CC10-Cre vectors seems to target better BASCs than club cells. This can be explained by the fact that, alveolar and BADJ epithelium is the most accessible and abundant than bronchi and bronchioles epithelium which may facilitate the access of the virus vector (Rock and Hogan, 2011).

Flow cytometry data revealed that Ad5-mSPC-Cre and Ad5-CC10-Cre vectors shared an overlap in their ability to express the Cre recombinase in BASCs cells, as both these cell types express both SPC and CC10 genes. While this overlap may not pose an issue for the organoids, as BASCs can be specifically isolated from club and AT2 cells through EpCAM and Sca-1 expression, it is essential to consider that in vivo models detailed in chapter four using Ad5-mSPC-Cre and Ad5-CC10-Cre may give rise to tumours from AT2 and BASCs cells or club and BASCs cells, respectively.

Moreover, the Ad5-CGRP-Cre vector demonstrated limited specificity as only 0.02% of EGFP positive cells were positive for CD56, which accounts for less than 1% of the lung epithelium (Garg et al., 2019). This raises questions about whether the low titre of  $1 \times 10^8$  pfu/ml is insufficient for targeting such a small cell population, if the virus itself lacks specificity or if the isolation method used was not efficient enough to isolate this population and perhaps it needs to be adapted. Therefore, it was decided not to proceed with neuroendocrine tumour organoids as we faced difficulty in specifically isolating neuroendocrine cells.

The virus vectors demonstrated specific targeting of these cells. However, the number of other tumour-initiating cells targeted by these viruses has not been explored. Therefore, we could proceed with the development of lineage-restricted LTOs in vitro. However, only AT2 organoids were generated successfully because they are the most abundant cell type which increases the number of cells sorted and because a significant lack of time during this project.

### **3.3.3 Organoids protocol development**

The development of the protocol was met with several challenges. Firstly, the lack of experience in our research group with 3D cultures led to a series of unsuccessful attempts to generate proliferating organoid cultures. We experimented with various protocols from the literature, modifying them to create the simplest and most effective one possible. However, the lack of standardization in 3D culture methods resulted in the need to explore different resources and protocol variations, such as the use of EGF instead of N2 and B27, and the selection of the appropriate support cell system.

Secondly, we also faced technical difficulties due to the COVID-19 pandemic. Due to the restrictions in place during 2020 and 2021, flow cytometry and sorting training were significantly reduced and the number of mice in colonies was decreased for safety reasons during short-staffed periods. Therefore, breeding efforts were essential to restore the required number of mice for the mTmG and MTMGKP mouse models used for cell-of-origin isolation and wild-type mice used for endothelial cell isolation. This led to a delay in the development of the protocol, and it was crucial for club, BASCs and basal cells as their numbers in

lung epithelium is reduced therefore between 4-6 pair of lungs were sorted at a time to obtain sufficient cells for 3D culture.

We first adapted Owen Sansom's research group murine colon crypt organoids culture protocol. This medium includes BSA, which benefits the growth and survival of cells when cultured in an FBS-free media due to its antioxidant properties (Francis, 2010); HEPES, for additional buffering capacity (J.D. WILLIAMSON, 1968); N2 and B27, which are neural growth factors efficient to keep primary neurons and colon crypts alive in culture (Zhang et al., 2020); and noggin to prevent spontaneous differentiation of stem cells in culture (Chaturvedi et al., 2009). Although murine R-Spondin was used for crypt organoid culture to induce crypt proliferation, there was no evidence in the literature regarding its effect on lung cells; thus, it wasn't used in this culture. To improve cell survival, ITS was later added to the media as it has been proved effective in other lung organoids cultures (Choi et al., 2016). Furthermore, we experimented with seeding cells into Matrigel of different stiffnesses by diluting it in media to provide LTOs with sufficient space in the matrix to grow. Unfortunately, LTOs did not grow in any of these cases.

Several factors may have contributed to the failure of the initial protocol. For instance, N2 and B27, while efficient for specifically enabling colon crypt survival and proliferation (Miller et al., 2019) there is no evidence of their effectiveness in murine primary cell culture lung organoids. Similarly, noggin, though effective in preventing stem cell differentiation in vitro for human pluripotent stem cell-derived lung organoids, lacked evidence for LTOs (Miller et al., 2019). Consequently, it appears that AT2 cells in our culture lacked essential growth factors and differentiation inhibitors, potentially resulting in cell death and/or unwanted cell differentiation, leading to the absence of organoid structures. However, upon replacing N2 and B27 with EGF, which specifically enhances epithelial cell proliferation, and replacing noggin with 10 $\mu$ M of ROCK/RHO inhibitor, widely used in our research group for lung cell culture, we observed a low number of organoids in culture.

While these changes and the observed organoid structures indicated progress in the right direction, the limited growth and number of organoid structures raised concerns about the possible lack of essential growth factors, nutrients, or other

critical molecules in the culture. Notably, when the whole lung single-cell suspension was seeded, organoids proliferated and grew, indicating a potential role of support cells or factors secreted by these cells in organoid growth. Upon consulting with Dr. Christine Brainson and Dr. Joo-Hyeon Lee, we learned that support cells, particularly endothelial or mesenchymal lung cells, may be important for organoid survival and growth.

Mesenchymal cells proved challenging to culture *in vitro* for an extended period, potentially due to their low proliferation rate or the lack of specific growth factors in the media. This was not explored any further as endothelial cells were easy to isolate and culture although this change delayed the protocol development. Endothelial cell culturing protocol was provided by Dr Christine Brainson and Dr Joo-Hyeon Lee and it involves the isolation of CD31 positive cells from 20-day-old WT mice and their culture in Advanced DMEM containing heparin, which is used for cell expansion, ECGS, essential for endothelial cell survival and proliferation *in vitro*. It is not clear what it contains but it is suspected to be a mix of growth factors and serum components, penicillin, streptomycin, glutamine, to improve cell survival, and FBS. Additionally, plates were previously coated with 0.1% gelatin to enhance cell adhesion and differentiation.

As for AT2, club and BASCs organoids, it became evident that we were missing critical factors. Several new culture factors were added to the medium at this stage. For instance, insulin was added, which promotes glucose uptake in endothelial cells, promotes cell growth and reduces apoptosis in stem cells (Shahbazi et al., 2019); along with cholera toxin (Yu et al., 2016), which promotes colony and organoid formation. Furthermore, bFGF, known for its multiple biological functions such as cell proliferation and cell renewal (Onuma et al., 2015), was also introduced, along with bovine pituitary extract, containing hormones and growth factors essential for cell culture under reduced FBS conditions (Xu et al., 2013).

Our lung tissue dissociation protocol, as proven in section 3.2.1, was found to be the most efficient for obtaining viable cells from murine lung epithelium. This protocol differs from the one received from Dr Christine Brainson. Additionally, we used endothelial cells as support cells and we reduced the stiffness of

Matrigel from 1:1 v:v to 1:3 v:v, as no increase in organoid size was observed with the previous ratio. According to Dr. Christine Branson and Dr. Joo-Hyeon Lee's protocol, organoids should become confluent and ready for splitting within three weeks after seeding. However, it took us up to 8 weeks to reach this point.

### 3.3.4 Organoids protocol checkpoints

To address these challenges, we decided to implement certain checkpoints. The differentiation state of the cells after the dissociation protocol and sorting was assessed, as it is crucial to preserve the stem cell-like properties of the tumour cells-of-origin in the lung epithelium. As explained in the introduction, these cells have stem cell-like properties and may differentiate into other lung cell types under stressful conditions or during tissue insults. Most importantly, Ad5-mSPC-Cre and Ad5-CC10-Cre vectors that we use for cell activation control CRE recombinase with mSPC and CC10 promoters, which are expressed in AT2 and club cells respectively and BASCs cells, but they are not expressed in the cells they differentiate to (Chen et al., 2014). Consequently, if CRE recombinase is not expressed there will not be any *KRas* activation or *P53* deletion.

Flow cytometry analysis was not performed right after the sorting because antibodies used for sorting might still be attached to their antigens for some time, making it difficult to interpret the results. Thus, cells were cultured for a few days before restaining to avoid this issue. AT2 cells showed a higher differentiation rate after 10 days than club cells and BASCs. However, the presence of undifferentiated AT2 cells in culture was confirmed and given that the activation in vitro is performed immediately after the sorting if the protocol is followed, it is assumed that the state of AT2 cells is still adequate for a proper and efficient activation right after the sorting.

## 3.4 Conclusion

We successfully identified and isolated several lung cancer cells-of-origin and the virus vectors used for cell activation demonstrated specific targeting of these cells. Only AT2 cell organoids protocol has been developed but club, basal and BASCs cells organoid protocol looks promising, even though our lack of experience and COVID-19 related issues.

## Chapter 4 Mouse model characterisation

### 4.1 Introduction

GEMMs have significantly contributed to our understanding of molecular mechanism of cancer pathogenesis (Cheon and Orsulic, 2011). GEMMs were first developed in 1980s for mammary adenocarcinomas carrying *MTC/MYC* fusion genes (Timothy A. Stewart, 1984) and nowadays they have been developed for almost all cancer types and they are widely used in laboratory research around the world (Cheon and Orsulic, 2011).

There are several reasons behind the expansion of GEMMs. Among the advantages of working with GEMMs, it is important to consider that mice are small in size, they are not expensive to maintain compared to other bigger mammals, they reproduce quickly and have big litters and they can be genetically manipulated as their genome is well established (Cheon and Orsulic, 2011, Mukherjee et al., 2022). Moreover, most of them mimic the clonal origin of human tumours, tumour histopathology and they recapitulate the processes of tumorigenesis, including early-stages processes, metastatic spread and recurrence after treatment (Balmain, 2001). Consequently, mouse models research has provided several insights into tumour biology.

GEMMs have helped researchers to understand the physiological role of cancer genes. The altered expression of an oncogene in mice results in disruption of the embryonic process or several tissue phenotype alterations, thus the understanding of a gene's physiological role is crucial to understanding its role in cancer (Lori C. Gowen, 1996, R Hakem, 1996). Moreover, it has been shown using GEMMs that only certain combinations of genetic alterations are able to induce tumours and that these combinations are tissue specific. For example, only *c-Myc* cooperates with *Tp53* in the development of mouse ovarian epithelial cells (Xing and Orsulic, 2006) or *Lkb1* loss together with *KRas* mutation is necessary to drive lung squamous carcinoma in mice (Zhang et al., 2017). In addition, GEMMs have proved that tumours activate hypoxic, angiogenic, metabolic and survival pathways to maintain their growth kinetics and to adapt themselves to different treatments (Cheon and Orsulic, 2011). GEMMs have also been used to study the tumour microenvironment and to understand the role of it in tumour

development (Cheon and Orsulic, 2011) and to identify novel oncogene or therapy gene targets as most important regulatory pathways in cancer are conserved among the species (Cheon and Orsulic, 2011).

Nevertheless, these models present certain limitations and challenges. One of the most significant limitations is the species-specific differences between mice and humans. Humans and mouse not only differ in terms of their size and lifespan, but also in the genomic instability of human cancer, due to telomerase activity which remains active in adult mice, but it is inactive in adult human cells. Consequently, mice require less genetic alteration for tumour progression than human tumours do (Weinberg, 2003). These species-specific differences between mice and human also result in mouse tumours with a different histology even though they share the same genetic alterations as human tumours. For instance, retinoblastoma heterozygous mice develop pituitary adenocarcinoma while humans develop retinoblastoma (Tyler Jacks, 1992). Finally, mouse cancer models develop metastasis in a lower frequency than human tumours or metastasis with different tissue specificities and differences in metabolic and signalling pathways which might result in differences in the anticancer drugs response (Cheon and Orsulic, 2011).

This deficiency in effective drug testing is still uncertain and it is another major challenge of GEMMs. The differences between species described above indicate that the response to anticancer drugs may vary from mouse to human. Additionally, xenografts models still present certain advantages for drug testing compared to GEMMs because they are cheaper and easier to generate than mouse models. GEMMs may delay a drug study because tumours develop with a long latency which may take months to reach maturity. The use of GEMMs can be restricted by intellectual property rights and patents (Cheon and Orsulic, 2011).

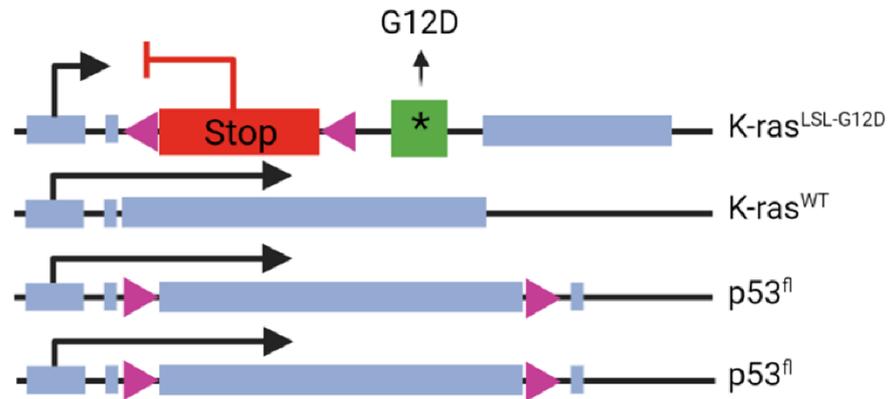
The last major limitation of mouse models is the recapitulation of clinically relevant tumour mutations. There are several techniques to create a GEMMs, including alterations in germ line or in a large proportion of somatic cells, in contrast to human tumour mutations, which are stochastic. Additionally, cancer mouse models are based on genetic alterations, and they omit the effect of the environment on tumour development such as the diet, the lack of physical exercise, epigenetic alterations and the importance of tumour

microenvironment. As a result, GEMMs tumours tend to be more homogeneous than human tumours. This hinders the study of human tumour heterogeneity and the anticancer drug development (Cheon and Orsulic, 2011).

Currently, GEMMs are the closest model to human cancer despite their limitations and that is why there were used in this project. In particular, we used the KP mouse model. This GEMM is a conditional system, which is based on the excision or inversion of DNA segments. There are several types including: the *Cre-loxP* system, the *flippase recognition target*, the inducible CreER system, and the tetracycline-inducible system. We used the KP mouse model with the *Cre-loxP* system (Cheon and Orsulic, 2011). The *Cre-loxP* system was first used for mouse genetic modification in 1998 when the transgenesis allowed the addition of functional DNA fragments into the mouse embryonic genome, thus the induction of somatic mutations in a tissue specific and time-controlled manner (McMahon, 1999).

The *Cre-loxP* system is based on the CRE recombinase of the P1 bacteriophage which catalyses the recombination between two of its recognitions sites *loxP*, which is a 13 pair of bases flanking sequences (Daniel L. Hamilton, 1984). A CRE recombinase enzyme binds each flanking *loxP* sequence and forms a tetramer by bringing the two *loxP* sequences together. CRE recombinase excises in the core spacer area of this sequence formed by 8 pairs of bases (Nagy, 2000). The resultant genetic modification depends on the location and orientation of the *loxP* sequences. They can be located cis or trans from the target gene. If they are located cis, their orientation can be the same or the opposite. In the first case, the flanked gene will be excised. If *loxP* sequences orientation is the opposite, the gene will be inverted. On the other hand, if the position of these flanking sequences is trans, a piece of DNA can be inserted into another or there could be a translocation between two molecules (Nagy, 2000). The potential of this system is wide, as it allows researchers to introduce oncogene mutations, delete tumour-suppressor genes or add an additional copy of an oncogene. For example, conditional activation of *KRas* in LSL-*Kras*<sup>G12D</sup> mice was achieved by targeting *loxP-STOP-loxP-KRas*<sup>G12D</sup> into the endogenous *KRAS* locus. The mutated version of *KRas* can only be expressed after CRE-mediated removal of the *STOP* sequence (Jackson et al., 2001). Another example is the deletion of both alleles of *P53* gene. In this case, *P53*<sup>fl/fl</sup>, 2 to 10 exons are floxed by *loxP* sequences.

After the CRE-mediated excision of these two *loxP* sequences, *P53* becomes inactive (Dodd et al., 2015).



**Figure 4.1. Schematic representation of KP reporter gene set in mice.** KRAS G12D mutated gene is preceded by a STOP codon which is surrounded by *LoxP* sequences. 2 to 10 exons of *P53* are surrounded by these *LoxP* sequences too. After the expression of CRE recombinase, G12D RAS mutation is introduced and *P53* deleted. Created with BioRender.com.

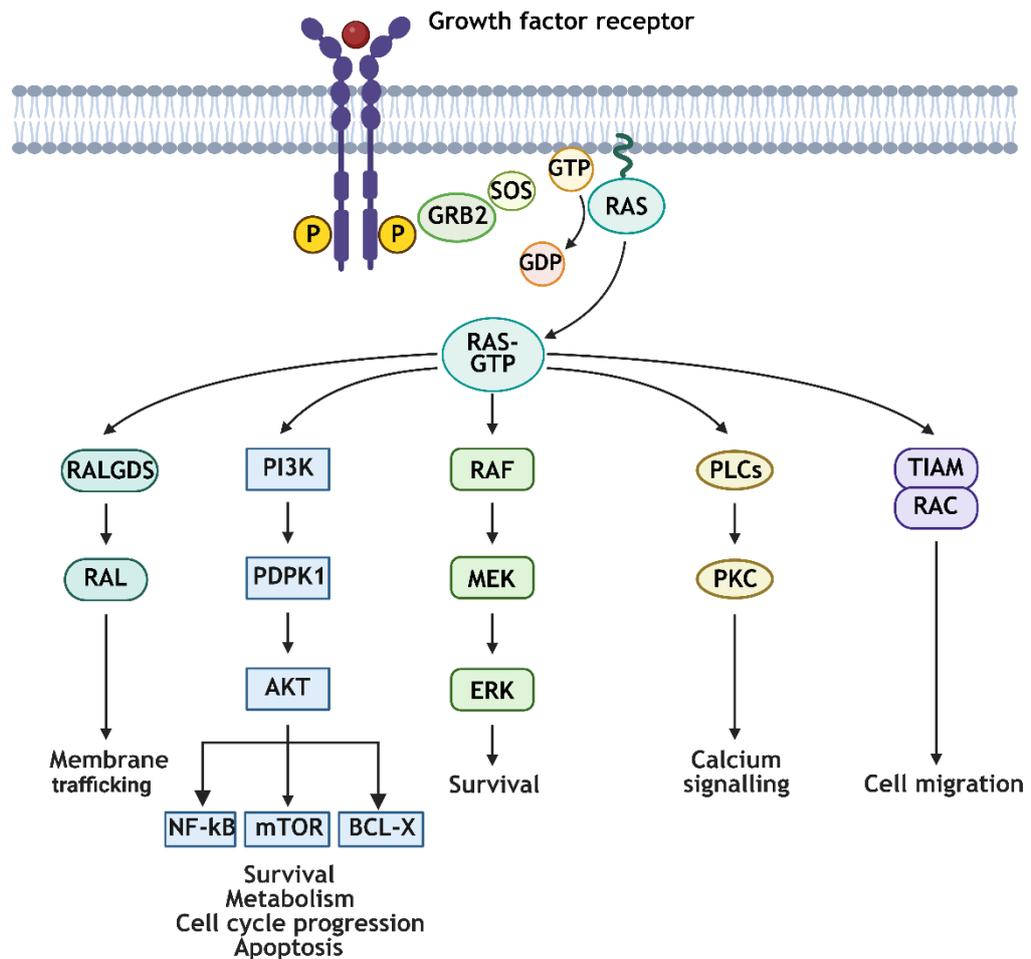
The advantage of this model is that the gene function can be studied in a specific tissue or cell type at specific time points. The major disadvantage of this conditional model is that it is irreversible, especially gene deletion. Once the target gene is excised, all this mouse descendants will carry this deletion, even though if the *Cre* is not expressed again (Cheon and Orsulic, 2011).

The only requirement for this model is the expression of CRE in the target cell types or tissue. CRE is a phage enzyme therefore its expression is exogenous of the mouse model. There are several ways to achieve this in vivo, for instance, a cross between a *loxP* provider and a CRE expresser transgenic mouse line or by a pronuclear injection of a *CRE* expression vector (Nagy, 2000). The most widely used method is to use an established transgenic line expressing CRE under the control of a generic or tissue/cell type specific promoter in an adenovirus or lentivirus vector. There are, of course, several limitations to this method. For instance, depending on the target tissue, surgery may be required to deliver the virus, or its delivery may be restricted; leakage to neighbour tissue or the possibility to host an immune response (depending on the dose or the frequency of induction) or cytotoxic activity can make the analysis complicated (Cheon and Orsulic, 2011).

As for lung cancer, intranasal inhalation or intratracheal intubation of a virus vector, typically an adenovirus, containing the *Cre* recombinase gene under the control of typically a generic promoter is the most used method (DuPage et al., 2009). By using these induction techniques, several mouse models for lung cancer research have been developed the recent years which recapitulate lung adenocarcinoma, lung squamous carcinoma (both non-small cell lung cancer) and small cell lung cancer. For example, most of the studies in NSCLC have been performed in *loxP-Stop-loxP* conditional *KRas*<sup>G12D</sup> mutation (Jackson et al., 2001), which is sufficient to drive tumours lesions by itself although it can be found together with *P53* deletion (Leisa Johnson, 2001). Less commonly, *loxP-Stop-loxP* conditional *KRas*<sup>G12D</sup> mutation can be found in combination with *PTEN*, *JNK1*, *GATA2* or *LKB1* deletion. *EGFR* and *ALK* fusion gene have also been used as driving mutations for lung adenocarcinoma. Moreover, overexpression of *SOX2* and loss of *LKB1* in mouse model seem to arise SCC. Finally, retinoblastoma and *P53* deletion are frequently derived in SCLC tumours (Kwon and Berns, 2013).

As mentioned above, KP mouse model was used in this project, which consists in *KRas*<sup>G12D</sup> mutation and deletion of both alleles of *P53* induced by intranasal inhalation. Both *KRas* activating mutations or *P53* deletions are tumour driving alterations due to all the cellular functions that derived from them.

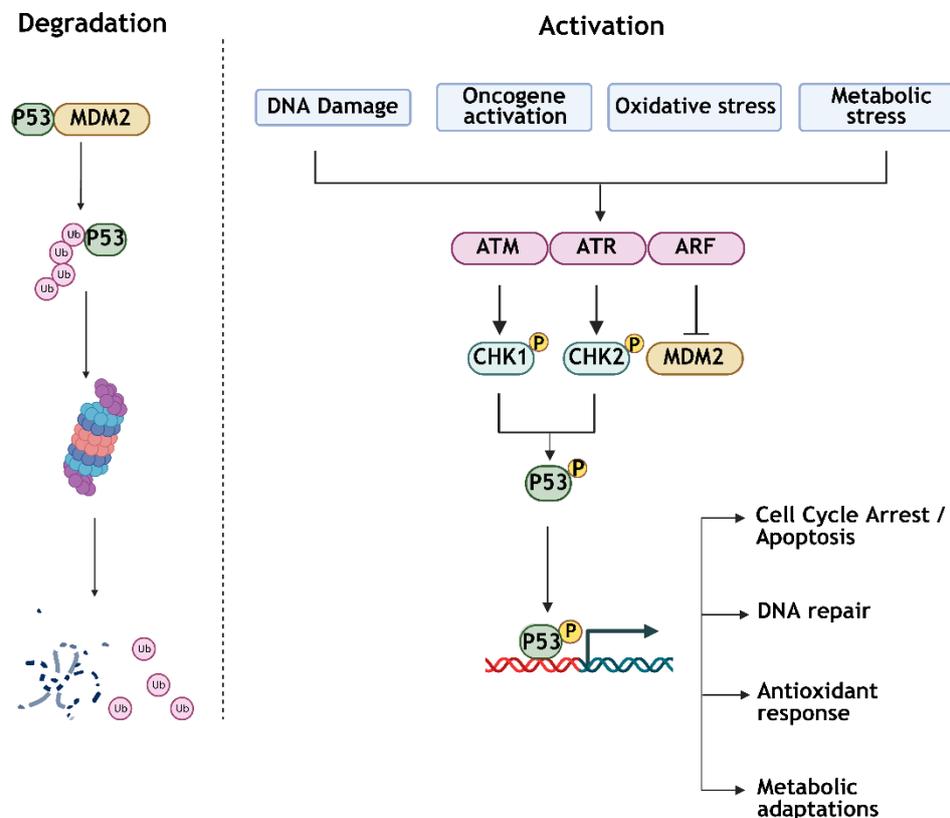
*KRas* is a GTPase, which is activated by cell growth and proliferation receptors such as EGFR and it is highly involved in malignant transformation by regulating cell proliferation and survival through RAF-MEK-ERK signalling, actin reorganising and cell migration by RHO-TIAM-RAC activation, increasing intracellular vesicular transport by RAL activation and apoptosis inhibition and changes in metabolism through PI3K-AKT-mTORC1 activation (Kim et al., 2021).



**Figure 4.2. Summary of the KRAS pathway and its effects in cell culture.** KRAS is a GTPase which is activated by several growth factors. GRB2 and SOS form a protein complex attached to the phosphorylated receptors. This complex recruits a GTP molecule which activates KRAS. Through several intermediate proteins, KRAS can regulate several cellular processes related to tumour progression. RALGDS/RAL regulated membrane trafficking, PI3K/AKT regulates survival, metabolism, and cell cycle progression, Raf/MEK/ERK regulate cell survival, PLC/PKM control calcium signalling, and TIAM/RAC regulate cell migration. RALGS ral guanine nucleotide dissociation stimulator, RAL Ras-related protein ral-A, PI3K phosphatidylinositol-4,5-bisphosphate 3-kinase, PDPK1 3-phosphoinositide dependent protein kinase 1, AKT serine/threonine kinase 1, NF-kB nuclear factor kappa B subunit 1, mTOR mechanistic target of rapamycin kinase, BCL-X apoptosis regulator, RAF proto-oncogenic serine/threonine kinase, MEK/ERK mitogen activated protein kinase 1, PLC phospholipase C, PKC protein kinase C, TIAM T-lymphoma invasion and metastasis-inducing protein 1, RAC Rac family small GTPase. Created by BioRender.com

*P53*, on the contrary, is a tumour-suppressor gene which in non-stress conditions is degraded by *MDM2* in the proteasome. After a cellular stress such as DNA damage, oncogene activation, nutrient starvation or oxidative stress activate *P53* pathway (Mantovani et al., 2019). *ATM*, *ATR* or *ARF* surveillance proteins inhibit the degradation of *P53* by phosphorylating *CHK1* and 2 which

phosphorylates P53. P53 is a transcription factor that will then go to the nucleus and increase the expression of genes related to cell cycle arrest, apoptosis, DNA repair, antioxidant response or metabolic adaptations (Labuschagne et al., 2018).



**Figure 4.3. Schematic representation of P53 pathway and its activation effects in cellular processes.** In cellular stress-free conditions, P53 is ubiquitinated by MDM2 and degraded in the proteasome. After DNA damage, oncogene activation, oxidative stress or metabolic stress P53 is phosphorylated and translocated to the nuclei to increase genes related to cell cycle arrest or apoptosis, DNA repair, antioxidant response, and metabolic adaptations. MDM2 E3 ubiquitin protein ligase, Ub ubiquitin, ATM ATM serine/threonine kinase, ATR ATR serine/threonine kinase, ARF ADP-ribosylation factor 1, CHK checkpoint kinase. Created by BioRender.com.

To retain control over the cell-of-origin of the potential tumour, Adenocarcinoma viruses were used to target these cells. These vectors contain a construct with *Cre* recombinase gene under the control of a cell type-specific promoter. AT2 and BASCs specifically expressed SPC, thus, Ad5-mSPC-Cre vector was used to target these two cell types. At the same time, club cells and BASCs specifically express CC10, therefore the Ad5-CC10-Cre vector used to target these two cell types. Both these viruses have an overlap and BASCs express both CC10 and SPC genes. Finally, basal cells in the murine trachea have been suggested to be the cell-of-origin of squamous cell carcinoma and basal cells specifically express the

KT5 gene. Therefore, Ad5-KT5-Cre vector was used to target those cells and obtained tumours arising only from basal cells (Ferone et al., 2020).

### 4.1.1 Hypothesis and Aims

The primary aim of this chapter is to normalise tumour burden among the Ad5-mSPC-Cre, Ad5-CC10-Cre, and Ad5-KT5-Cre mouse models and to characterise the resulting tumour growth. To accomplish this, we will modify the inducing virus titre in each model consistently with the number of target cells present in the lung epithelium. By doing so, we can directly compare the metabolic profiles of the tumours, independently of their sizes or growth rates. This approach ensures that the differences observed in tumour metabolism are due to the specific cell types targeted by the virus vectors, eliminating any confounding factors related to tumour volume or kinetics.

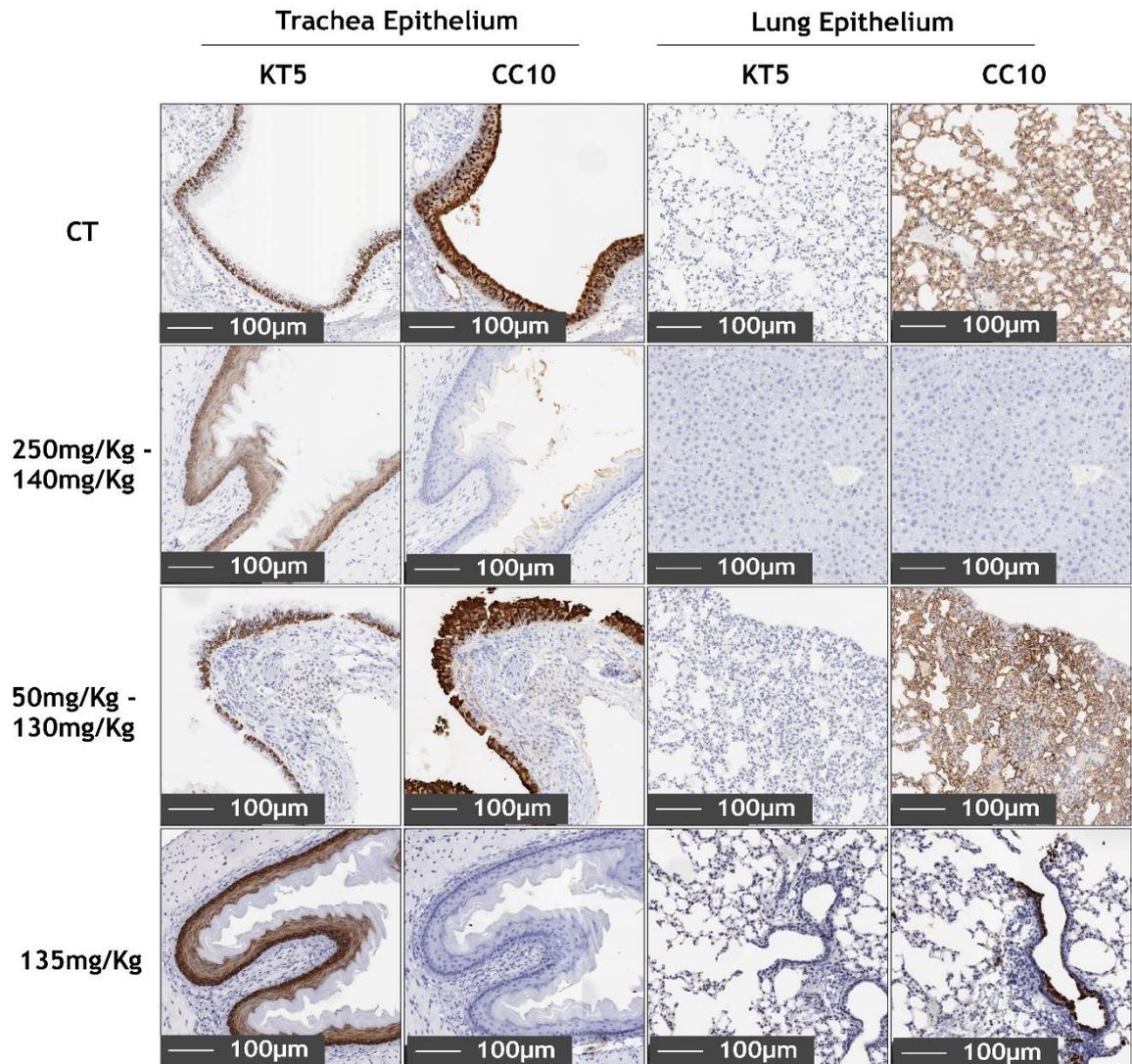
## 4.2 Results

### 4.2.1 Generation of KT5 mouse model

KP mice were induced with Ad5-mSPC-Cre, Ad5-CC10-Cre and Ad5-KT5-Cre. Ad5-KT5-Cre specifically target basal cells which are located in the mouse trachea, which express *Kt5* specifically (Kyung U. Hong, 2003). Club cells are more abundant in the bronchi and bronchioles although they can also be found in the murine trachea (Rock and Hogan, 2011). These club cells create a physical barrier in the trachea, covering basal cells (Figure 4.4 CT) (Ferone et al., 2016). This makes the murine tracheobronchial epithelium relatively refractory to adenoviral infection. However, naphthalene specifically depletes club cells temporarily (Ferone et al., 2016), facilitating the virus vector access to tracheobronchial basal cells.

Therefore, KP mice were treated with several naphthalene concentrations, from 250 mg/kg to 50 mg/kg to optimise the most effective naphthalene concentration (West et al., 2001). Mice treated with concentrations over 135 mg/kg were culled within 24 hours after the IP injection as the animals looked sick and they exhibited a massive club cell depletion not only in the tracheobronchial epithelium but also in lower airways compared to the control (Figure 4.4). Trachea club cells in mice treated with concentrations below

135 mg/kg were not removed completely from the epithelium (Figure 4.4). However, animals treated with 135 mg/kg of naphthalene showed a depletion of club cells in the trachea but not a complete CC10 positive cells in the airways (Figure 4.4), which significantly reduced their disease signs. This allows the virus vector to reach and infect basal cells in the trachea.

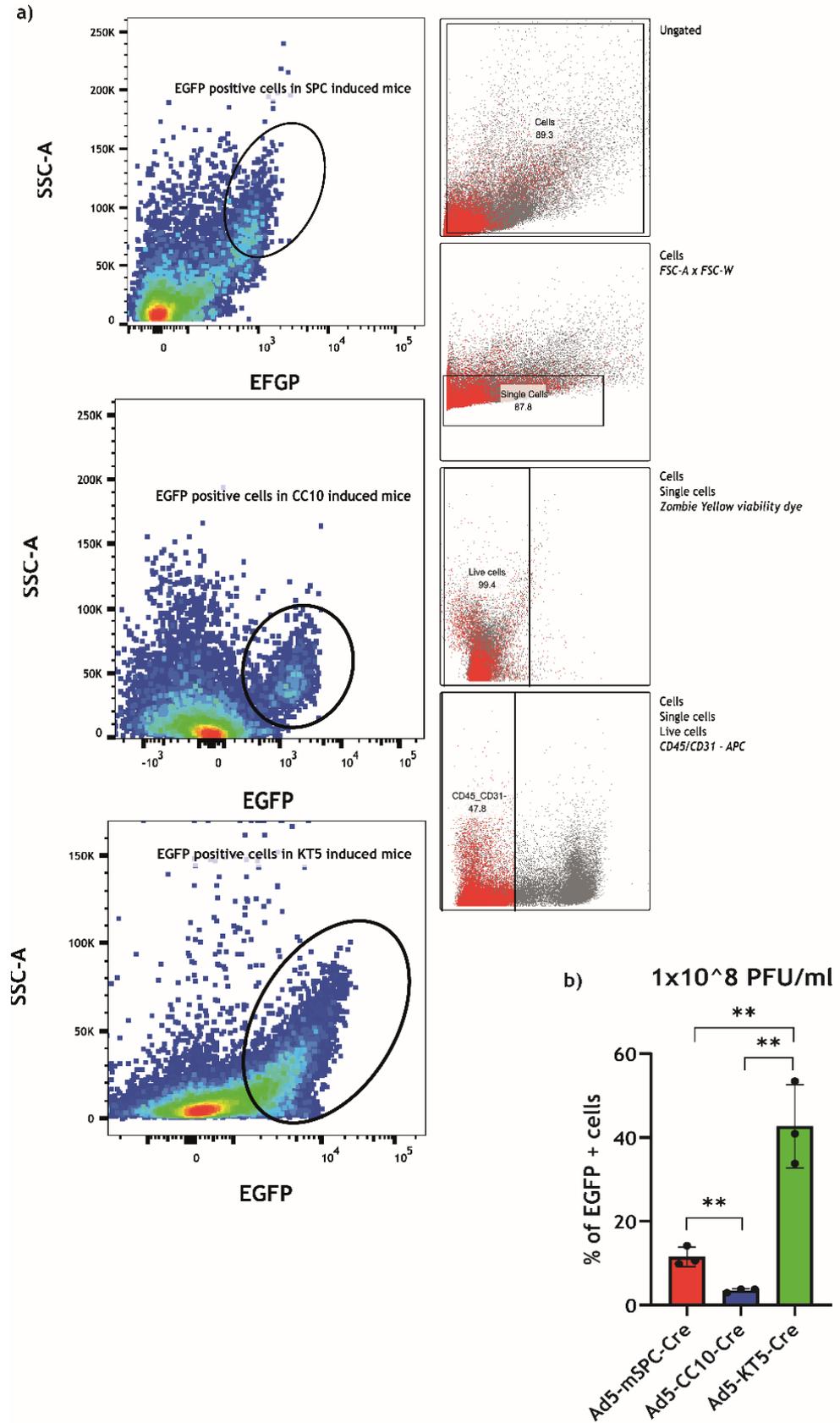


**Figure 4.4.** Naphthalene treatment of 135 mg/Kg reduces CC10 staining in the trachea epithelium but maintains CC10 staining in the lung epithelium which will allow the regeneration of CC10 and the mouse survival. Representative CC10 and KT5 staining sections showing the expression of these two proteins in the lung and trachea epithelium in control and several doses treated mice.

### 4.2.2 Effect of the virus vector titre on target cells

Virus titres have been modified in each mouse model consistently with the target cells in the lung epithelium to ensure that they are both targeting a similar number of cells and reduce variability within these models. mTmG mouse model described in section 3.1 was induced with two different virus titres. Firstly,  $1 \times 10^8$  pfu/ml was used to induce mTmG with Ad5-mSPC-Cre, Ad5-CC10-Cre and Ad5-KT5-Cre vectors. Afterwards,  $5 \times 10^6$  pfu/ml of vector was used to induced mTmG mice with Ad5-mSPC-Cre, and  $1 \times 10^8$  pfu/ml was maintained to induce the same mouse model with Ad5-CC10-Cre. The aim was to quantify EGFP positive cells to ensure that which titre combination was targeting a similar number of cells in all mouse models. Results were analysed using FlowJo and paired t-test and one-way ANOVA was used to compare the results.

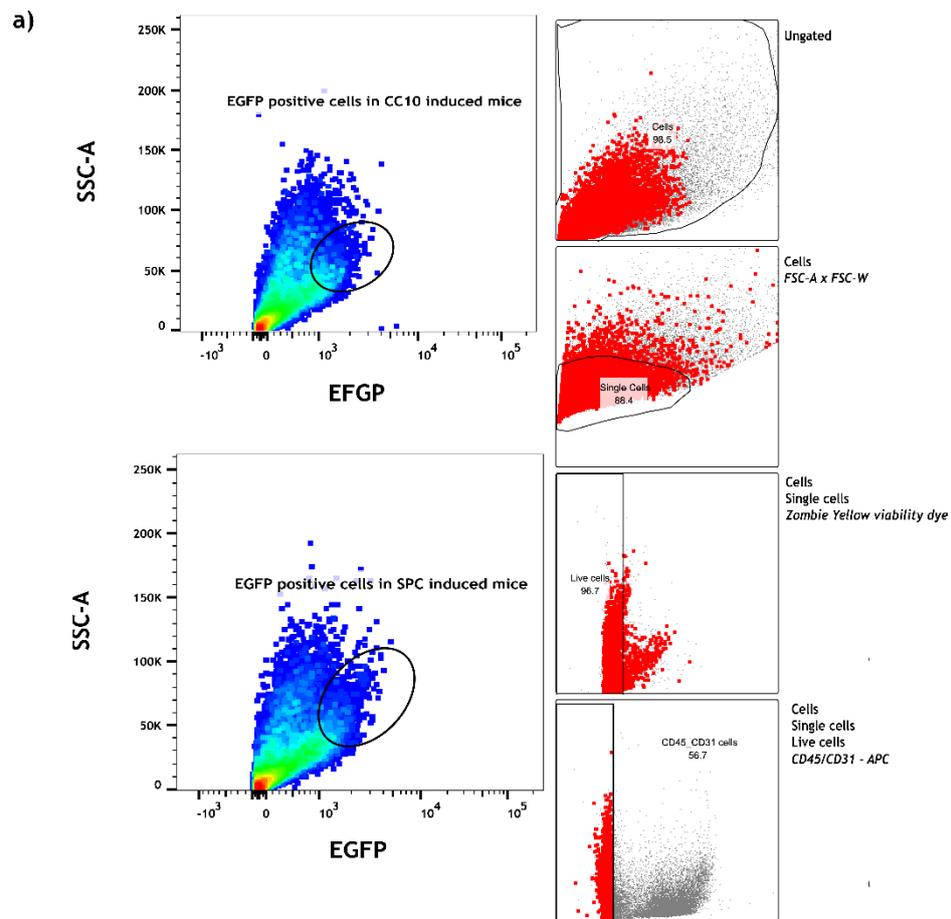
As seen in figure 4.5b, when mTmG mice were induced with  $1 \times 10^8$  pfu/ml of Ad5-mSPC-Cre, Ad5-CC10-Cre and Ad5-KT5-Cre, there is a statistically significant difference between the percentage of EGFP positive cells detected between these three virus vectors, suggesting a difference in the number of activated cells.



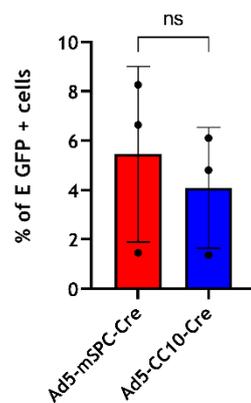
**Figure 4.5.** Gating strategy for lung epithelial cells panel used to identify SPC, CC10 and KT5 EGFP positive cells. Live cells were gated due to their negative staining for zombie yellow viability dye. Then, blood and immune cells were distinguished from epithelial cells due to their expression of CD45 and CD31 (APC staining). EGFP positive cells were identified based on their

slightly increase in green fluorescence. EGFP enhanced green fluorescence protein, dots represent ungated cells. Red dots represent gated cells. (b) Each data point in figure represents the result of an experimental trial (n=3 for each vector). Error bars in (b) are generated from mean  $\pm$ SD. Statistical analysis using ordinary one-way ANOVA in GraphPad Prism was used to determine a statistically significant difference across the three cohorts for each of the graphs presented. EGFP enhanced green fluorescent protein.

Additionally, when SPC virus target was modified to  $5 \times 10^6$  pfu/ml for Ad5-mSPC-Cre, and Ad5-CC10-Cre titre was maintained to  $1 \times 10^8$  pfu/ml, there is no significant difference in the percentage of EGFP positive cells for both mouse models, indicating a similar number of activated cells after the induction (Figure 4.6b).



b)  $5 \times 10^6$  PFU/ml &  $1 \times 10^8$  PFU/ml

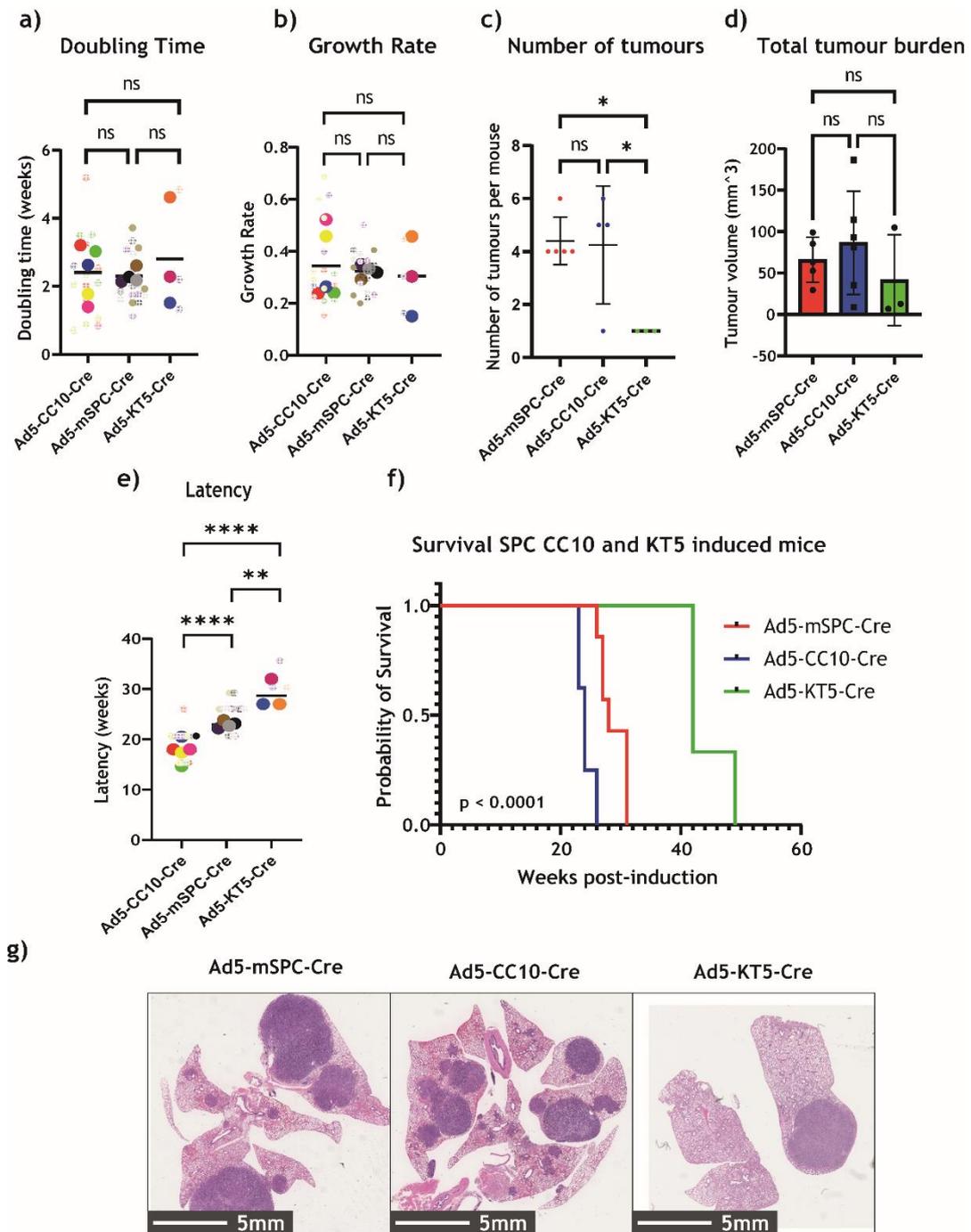


**Figure 4.6. Gating strategy for lung epithelial cells panel used to identify SPC and CC10 EGFP positive cells.** (a) Alive cells were gated due to their negative staining for zombie yellow viability dye. Then, blood and immune cells were distinguished from epithelial cells due to their expression of CD45 and CD31 (APC staining). EGFP positive cells were identified based on their slightly increase in green fluorescence. (b) Each data point in figure represents the result of an experimental trial ( $n=3$  for each vector). Error bars in (b) are generated from mean  $\pm$ SD. Statistical analysis using unpaired two-tailed t-test performed in GraphPad Prism was used to determine a statistically significant difference between the two vectors. ns  $p>0.05$ . EGFP enhanced green fluorescence protein, dots represent ungated cells. Red dots represent gated cells (b).

### 4.2.3 Mouse models characterisation

Our goal was to normalise tumour burden between SPC, CC10 and KT5 mouse models. Therefore, these three models were characterised. Tumour volume was calculated using VivoQuant in three mouse models. Five CC10, four SPC, and three KT5 mice were induced using  $1 \times 10^8$  pfu/ml,  $5 \times 10^6$  pfu/ml and  $1 \times 10^8$  pfu/ml respectively, and mice were scanned at 13-, 18-, 23- and 28-weeks post-induction. An exponential Malthusian growth equation and ordinary one-way ANOVA were applied to compare the growth rate, the doubling time, the latency, the number of tumours per mouse, and the total tumour burden between all three models. The survival of these mouse models was also analysed using a Mantel-Cox test.

Between SPC and CC10 tumours, there is not a significant difference in the number of total tumours per mouse, the total tumour burden, the growth rate, or the doubling time between both mouse models (Figure 4.7a, b, c, d), suggesting that tumour volumes increase at the same rate and that the time needed for the tumours to double is the same in both tumour types. However, the coefficient of variation is higher for the Ad5-CC10-Cre tumours (Table 4.1). We have found a significant difference in the tumour latency between the SPC, KT5 and the CC10 tumours. SPC tumours took an average of 22.9 weeks to be recognisable in the MRI scans, CC10 tumours took an average of 17.6 weeks while KT5 tumours take an average of 29 weeks (Figure 4.7e, f). This longer latency in Ad5-KT5-Cre mice is causing a difference in the survival curve, as the mice induced with the KT5 vector have a significantly longer life span when compared to CC10 and SPC induced mice.



**Figure 4.7. Growth dynamic and location data of CC10, SPC and KT5 tumours.** (a), (c) and (e) Graphs representing doubling time of tumours in weeks, growth rate and latency in weeks. Data is represented as superplots. Big data points represent the mean for all tumours within one mouse. Small data points represent the measure of every tumour. Horizontal bar represents the general mean of each tumour subtype. CC10 n=5, SPC n=4 and KT5 n=3. Statistical analysis using ordinary one-way ANOVA in GraphPad Prism was used to determine a statistically significant difference across the three vectors. No statistically significant difference was found in growth rate and the doubling time date. Latency is significantly different between CC10, SPC and Cre models. (c) and (d) graphs represent the total number of tumours and the tumour burden by the endpoint of the mouse. Each point represents the measurement of a mouse. Error bars are generated from mean  $\pm$ SD. Statistical analysis using ordinary one-way ANOVA in GraphPad Prism

was used to determine a statistically significant difference across the three vectors. No significant differences were found between SPC and CC10 tumours although the total number of KT5 tumours is statistically different against SPC and CC10 tumours. NS  $p > 0.05$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\*\* $p < 0.0001$ . Kaplan Meier curves comparing overall survival of Ad5-mSPC-Cre (n=6), Ad5-CC10-Cre (n=8) and Ad5-KT5-Cre (n=3). GraphPad prism was used to perform a Log-rank (Mantel-Cox) test for the Kaplan Meier curves. (g) Representative H&E images of clinical endpoint SPC, CC10 and KT5 mice revealing the location of the tumours.

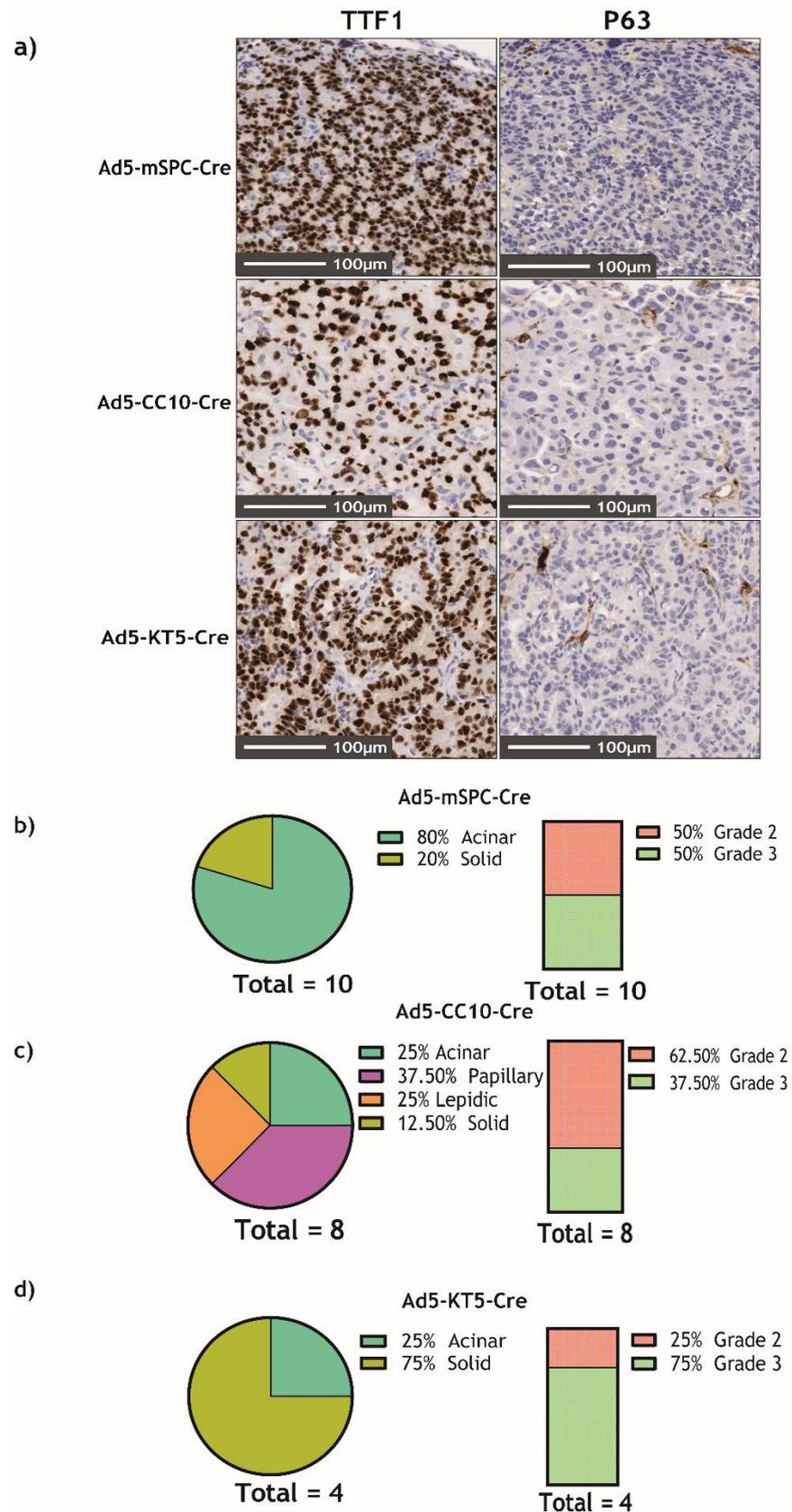
**Table 4.1. Summary of the tumours K rate and doubling time coefficients of variation from each one of the virus vectors used.** Coefficients of variation were calculated as standard deviation per mouse model/mean of mouse model x 100.

Virus Vector	K Rate			Doubling time		
	Ad5-CC10-Cre	Ad5-mSPC-Cre	Ad5-KT5-Cre	Ad5-CC10-Cre	Ad5-mSPC-Cre	Ad5-KT5-Cre
Coefficient of variation	0.567107	0.266463	0.506156	0.490477	0.284285	0.574984

KT5 tumours growth rate, doubling time, and total tumour burden are not significantly different to SPC and CC10 tumours (Figure 4.7a, b, d). However, it was observed a significant decrease in the number of tumours, as they only showed one tumour per mouse (Figure 4.7c). Additionally, latency of KT5 tumours is longer than SPC tumours and life span is significantly longer than SPC and CC10 mouse models (Figure 4.7e, f). They also present the largest coefficient of variation (Table 4.1).

Finally, we examined the tumours position within the lung, as CC10, SPC and KT5 positive cells have specific locations within the lung epithelium. KT5 tumours grew surrounding the main bronchi, consistent with the location of basal cells in the murine trachea and bronchi (Figure 4.7g) (Kyung U. Hong, 2003, Ferone et al., 2016). Based on bibliography, CC10 tumours appear in the bronchi and bronchioles, bronchoalveolar duct junctions and the alveolar space (Rock and Hogan, 2011). Finally, SPC tumours are located in the alveolar space where SPC positive cells are located (Figure 4.7g) (Rock and Hogan, 2011). Specific location markers staining are needed to fully understand tumour location.

In collaboration with John Le Quesne and Kai Roakovic from the histopathology department SPC, KT5 and CC10 tumours were histologically classified. Given that the tumours lesions were heterogeneous, presenting different patterns in the same tumour, the characterisation was performed on a per mouse basis where the presence or the absence of a tumour pattern was noted.



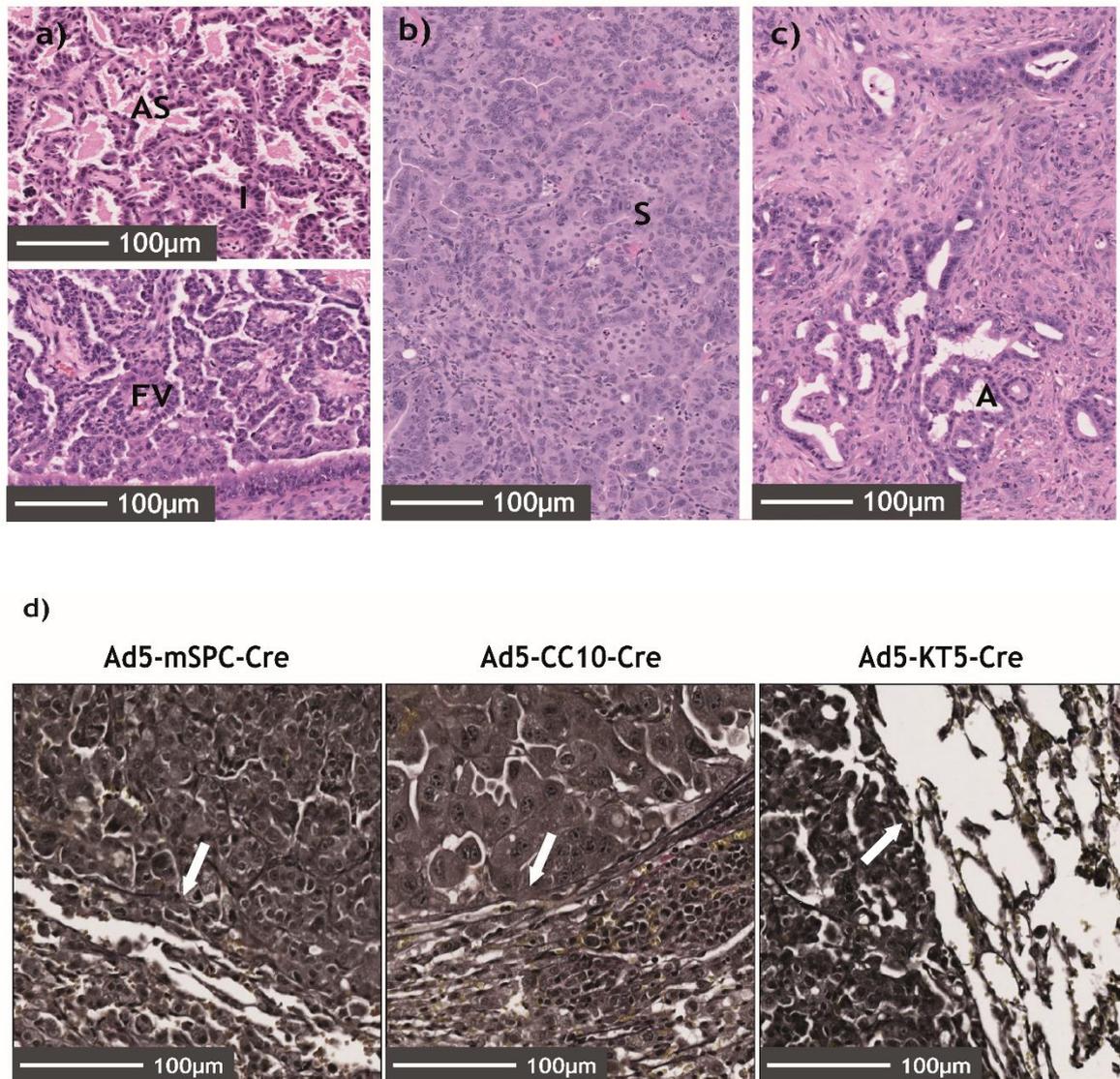
**Figure 4.8.** Histopathological analysis of TTF1 and P63 in SPC, CC10 and KT5 tumours. (a) Murine lungs were dissected at clinical endpoint and stained for TTF1 and P63. CK18. Images are representative of SPC n=11, CC10 n=8 and KT5 n=3. (b), (c) and (d) show proportions of tumours

which predominant pattern is one of the stated. The proportion of tumours within grade 2 and 3 are also represented.

All three models developed lung adenocarcinomas, based on the expression of TTF1 (Rekhtman et al., 2011) and the lack of expression of p63 which is a squamous cell carcinoma marker (Conde et al., 2010) (Figure 4.8a). There are different adenocarcinoma histological patterns. For instance, as seen in figure 4.8b, solid adenocarcinoma tumours are predominant in 20% of the SPC induced mice and acinar tumours are the predominant histological pattern in the 80% of SPC induced mice. Acinar pattern is defined by irregularly shaped gland formation invading the stroma, in this case it is non-mucinous (Figure 4.9c) (William D. Travis, 2004). As explained before, the combination of these patterns provides an idea of the grade of these tumours which would lead to a prognostic idea in the clinic. SPC tumours are equally grade 2 meaning that they are moderately differentiated and grade 3 which are highly differentiated tumours (Travis et al., 2015, Nicholson et al., 2022) (Figure 4.8b).

CC10 tumours are highly heterogeneous. Solid tumours are the main tumour pattern in a 12.5% of CC10 mice, and acinar and lipidic tumour areas are predominant in 25% of CC10 mice each. Finally, papillary tumours are the predominant tumour subtype in the 37.5% of CC10 induced mice (Figure 4.8c). This last tumour pattern is characterised by fibrovascular cores with cuboidal/columnar cells (Figure 4.9a) (William D. Travis, 2004). They appear to be mainly grade 2 tumours (Figure 4.8c). Additionally, to papillary patterns, lipidic areas are only found in CC10 tumours.

Moreover, solid adenocarcinoma tumours are the main histological pattern present in 75% of KT5 induced mice (Figure 4.8d), which consists of sheets of neoplastic cells lacking glandular formation (Figure 4.9b) (William D. Travis, 2004). They present characteristics from both grade two and grade three.



**Figure 4.9.** H&E patterns and EVG staining for SPC, CC10, KT5 tumours. Exemplar histological images showing common growth patterns in the three tumour models. Tumours arising from CC10 (a) frequently exhibit lepidic (top) and papillary (bottom) morphology. Lepidic growth pattern is characterised by preservation of the overall architecture of the alveoli but with variable interstitial thickening and replacement of alveolar epithelium by the malignant population. Papillary growth pattern is defined by the presence of finger-like projections from the alveolar wall into the airspace, which are covered by the malignant population and contain a central fibrovascular core. Those arising from K5 (b) often have solid morphology, where the malignant population is arranged in expansile nests full of crowded, highly proliferative malignant cells. Finally, SPC tumours (c) often display acinar pattern, in which invasive malignant cells continue to make well-formed glands. (d) EVG staining shows broken elastin fibres in all tumour subtypes. White arrows indicate these areas. Images are representative of SPC n=11, CC10 n=8 and KT5 n=3. AS airspace, FV fibrovascular core, S solid nest EVG elastin van Gieson's stain and A acinus.

Finally, invasion pattern was described using elastin van Gieson staining. This staining highlights the elastic fibres in the connective tissue. This stain is useful

in demonstrating changes in these fibres such as breaks or splitting that in cancer, may indicate an invasive phenotype (Kalof et al., 2022). As observed in figure 4.9d, all three mouse types present broken fibres of elastin which may suggest an invasive phenotype of these tumours.

## **4.3 Discussion**

### **4.3.1 KT5 mouse model**

As explained above, club cells in the murine trachea epithelium physically cover basal cells, but they can be specifically depleted by naphthalene treatment. As this treatment had never been performed in the research group, certain optimisation was needed. Firstly, the administration route was considered. Naphthalene can be administered via inhalation or intraperitoneal. Given that the intraperitoneal route is simple and quicker to administer, and it allows a similar effect on the lung epithelium with lower doses (West et al., 2001), it was decided to use it as the administration route. Dose was also optimised as it has been described a wide of different concentrations in mice. In order to achieve a club cell depletion in the trachea epithelium only, 250 mg/kg was first injected to the mice as it has been described in the literature (Ferone et al., 2016). However, mice injected with such concentration during this project experienced decreased body temperature, a prolonged hunched appearance and reluctance to move due to a massive depletion of club cells not only in the trachea, but also in the distal airways. This may have led to difficulties in the respiration which may have increased the symptoms and resulted in the death of these mice within 24 hours. On the other hand, naphthalene concentrations lower than 135 mg/kg do not completely deplete CC10 positive cells from the trachea epithelium.

Naphthalene 135 mg/mg injected intraperitoneally seems to be the most accurate concentrations to ensure both the depletion of club cells of the trachea epithelium and the survival of the mice. As mentioned before, this depletion is temporary (Hsu et al., 2014). It has been demonstrated that from day 2 to 5 post-injection, basal cells start to proliferate and by day 10 the club cell population has been regenerated (Hsu et al., 2014). Therefore, mouse induction

was performed 24 hours post-injection when club cells were ablated but basal cells have not started proliferating.

Consequently, KP mice were treated with 135 mg/kg 24 hours before induction to ensure club cell depletion in the trachea, allowing the virus vector to access and target basal cells.

The effects of naphthalene in tumorigenesis are not well understood. It is known that regenerated club cells after naphthalene treatment arise from basal cells which start proliferating (Hsu et al., 2014). This might mean that Cre will act in basal cells which will proliferate and differentiate into club cells. Therefore, resultant tumours will be similar to CC10 tumours.

### **4.3.2 Targeted cells by cell activation**

AT2, club, BASCs and basal cells are distributed along the airways and their frequency in the lung epithelium differs. The percentage of EGFP cells activated by the virus vectors is similar in both SPC and CC10 mouse models when  $5 \times 10^6$  pfu/ml and  $10^8$  pfu/ml titres are used respectively. The reason behind this is that AT2, together with AT1 cells, account for a small proportion of the total lung cell population but they comprise around 95% of the lung surface area, because they are located in the alveolar space, the most abundant structure in the lung (Rock and Hogan, 2011, Williams, 2003). As a consequence, Ad5-mSPC-Cre virus particles have a higher chance of finding and activating a SPC positive cell than a A5-CC10-Cre virus particles of finding a CC10 positive cell. Therefore, by decreasing Ad5-mSPC-Cre to  $5 \times 10^6$  pfu/ml, the number of activated cells is potentially reduced, and it becomes like the number of CC10 activated cells, as the EGFP data reflect.

Basal cells are present only in the murine trachea and the bronchi and they account for around 30% of the lung pseudostratified epithelium which includes basal, club and ciliated cells (Rock and Hogan, 2011, Jason R. Rock, 2009). Basal cells are physically covered by club cells in the trachea epithelium, and a naphthalene treatment is necessary to deplete club cells and allow the virus vector to access basal cells and activate them. Given that this depletion is only temporary (Hsu et al., 2014) and that basal cells themselves are thought to

proliferate and differentiate to club and ciliated cells in the epithelium (Hsu et al., 2014), we decided to go for a relatively high titre of the Ad5-KT5-Cre vector ( $10^8$  pfu/ml) to make sure enough cells were targeted as basal cells epithelia initially seemed refractory to the virus. However, a significantly higher percentage of cells seem to be activated compared to SPC or CC10 tumours when mice are induced with the same titre. This might be explained by basal cell proliferation after naphthalene treatment (Hsu et al., 2014) which may result in a higher number of positive EGFP cells being detected. Consequently, a modification in this virus titre is being explored.

### **4.3.3 Mouse model characterisation and immunohistochemistry subtype**

As explained in section 4.3.1, titre for Ad5-mSPC-Cre, Ad5-CC10-Cre and Ad5-KT5-Cre virus vectors was adapted considering the frequency of the target cell types of the virus and how accessible these cell types are in order to obtain similar levels of cell activation. The aim was to obtain models where a similar number of cells were targeted in order to normalise the resultant tumour burden and growth kinetics of the CC10, SPC and KT5 tumours.

There were similar parameters analysed in order to understand these growth dynamics. Firstly, we examined the growth rate of these tumours. To do this, exponential Malthusian growth was the growth model preferred to analyse tumour growth. Malthusian growth curve is an exponential growth model which assumes that the cell population grows indefinitely with a constant doubling time. Cancer cells can keep growing in time as long as there are enough nutrients and resources to do so. Cancer cell growth is also exponential as from one cell two cells are produced after mitosis. Therefore, it was considered the Malthusian growth curve to be the most adequate model to study tumour growth.

Both parameters are not significantly different in all three models. As for CC10 and SPC tumours, the number of tumours by an endpoint and the total tumour burden were also similar. Nevertheless, the latency is significantly longer for SPC tumours. This might be explained by the upregulation of immune pathways in these tumours, which may indicate that SPC lesions may need more time to

overcome these immun checkpoints. As a result, there's a significant difference in the life span of these two models being SPC mice living longer than CC10 mice. Regardless of this last parameter, it was decided that these titres were good enough to proceed with the metabolic characterisation of these tumours.

CC10 growth rate and doubling time have a higher coefficient of variation compared to SPC tumours. This could be explained by the existence of different subtypes of club cells, all of which express CC10, making them more likely to be targeted by the virus vector (Nicholson et al., 2022). CC10 expressing cells are also distributed all along the airways. Club cells are in the trachea, the bronchi and the bronchioles and BASCs are in the alveolar space whereas SPC positive cells, both BASCs and AT2 cells, are only located in the alveolar space (Rock and Hogan, 2011). All this heterogeneity may contribute to the creation of different tumour subtypes which may explain this variation in the CC10 tumours.

As for KT5 tumours, they significantly differ in the number of tumours but not in the total tumour burden. Indicating that by an endpoint this single tumour can grow to a considerable size or several small tumour are merged together to form a bigger tumour, big enough to be like the burden of multiple tumours combined for SPC and CC10 models. They also have an even longer latency than SPC tumours and as a result a longer lifespan. As explained in section 4.3.2, it looks like the current virus titre ( $1 \times 10^8$  pfu/ml) seems to be targeting a higher number of cells than the other two vectors, which may suggest a reduction of the titre. Nonetheless, the low number of tumours per mice, the long latency and the higher coefficient of variation for KT5 mice would suggest an increase of the current titre. As these results seemed inconclusive, we decided not to pursue the metabolic characterisation of the KT5 mouse model.

#### **4.3.4 Clinical subtypes classification**

All three tumour types (CC10, SPC and KT5) are adenocarcinomas due to the positive TTF1 staining and the lack of p63 staining. As explained above, AT2, club, and BASCs cells are reported to be cells-of-origin of lung adenocarcinomas. Basal cells have been described as the cell of origin of squamous cell carcinoma. However, as it has been explained, GEMMs which produced squamous cell carcinomas are normally mutated in *KRas* and *Lkb1* for whereas, lung

adenocarcinoma lesions have been described to arise from alterations in *KRas*, *Egfr* or *P53*. Therefore, it is likely that, in this case, it is genetic mutations and not the cell-of-origin which drives the lung phenotype. Because it was expected to obtain SCC tumours from KT5 mice and not ADC as it was seen, most likely due to the genetic mutations introduced in these mice, it was found another reason to not to pursue the metabolic characterisation of KT5 tumours.

SPC, CC10 and KT5 tumours are either grade 2 or 3 but there are no low-grade tumours. This is a common characteristic of tumours induced in KP mice (Boumelha, 2022). SPC induced mice showed a higher percentage of high-grade histological patterns. CC10 tumours are more heterogenous and present more than half of the tumours as grade 2. SPC tumours presented a higher number of mice with grade 3 tumours. This is important as it has been shown that high grade KP lung tumours present metabolic alterations. Indeed, high grade KP (grade 2 and 3) cells showed increased expression of glycolytic genes and altered glucose metabolism (Kerr et al., 2016, Kerr and Martins, 2018). Nevertheless, SPC induced mice have a longer lifespan than CC10 mice. This might be explained by the location of the tumours for these mice. SPC positive cells are only present in the alveolar space but CC10 positive cells are distributed all along the murine airways, from the main bronchi to the alveolar space. This may result in tumour growing in respiratory tract where the air is circulating and might result in an obstruction of the airways, compromising free air circulation and mouse health. However, alveolar epithelium is significantly more abundant in the lung and easier to compensate so it may require more time, bigger or more aggressive tumours to compromise the air exchange. The MRI imaging does not have resolution enough to locate CC10 and SPC tumours with that precision. This would explain that even though, SPC tumours have worse prognosis, if CC10 tumours cover more healthy lung area this may influence negatively in the outcome of the mice.

As for KT5 tumours, 75% of induced mice present predominantly high-risk solid patterns and grade 3 tumours. However, these mice have the longer life span than CC10 and SPC tumours, probably explained by the long latency of these tumours. In other words, these tumours seem more aggressive, but they require more time to arise. But again, it is not likely that such big and aggressive tumour do not have an early impact on the mouse health. Due to these contradictions

explained KT5 mice were not further characterised. A possible explanation for this incongruencies is the naphthalene treatment. This CC10 cells depletion forces KT5 to proliferate and differentiate. It is possible that during the first 24 hours after treatment they undergo through some cell programs that change their phenotype, altering the phenotype of the resultant tumours.

Finally, EVG staining showed that these three tumour subtypes may have developed an invasive phenotype. No metastasis was detected in these mouse models, indicating an M0 phase in the TNM system. However, the presence of cancer cells in the neck lymph nodes should be analyse to confirm their invasiveness, and further characterised their stage (N0 or N1) and the implication that this process would have in the tumour metabolism.

#### **4.4 Conclusion**

Even though, there are some differences between SPC and CC10 tumour types, we considered that the current virus titres for CC10 and SPC were accurate enough to produce tumours with similar growth kinetics. Therefore, we decided to carry on with the metabolic characterisation. As for KT5 tumours, the data obtained in the tumour characterisation seemed a bit contradictive. Firstly, we were expecting to obtain squamous cell carcinomas but instead, Ad5-KT5-Cre virus was producing adenocarcinomas. Finally, flow cytometry data, suggests a decrease in the virus titres but the total number of tumours, the tumour burden and the latency of the tumours suggests an increase in the titre. Therefore, it was decided not to proceed with the metabolic analysis of these mouse model.

## Chapter 5 Metabolic, Transcriptomic and Proteomic Tumour Characterisation

### 5.1 Introduction

As stated in section 4.1, KP mice have been used for this project and they have been induced with Ad5-mSPC-Cre and Ad5-CC10-Cre virus vectors. These vectors target AT2 cells and club cells respectively, and they both target BASCs as these last cells express both SPC and CC10 genes. In this analysis five KP mice were also induced with Ad5-CMV-Cre. This virus vector is widely used in our research group with several GEMMs. In this case, the *Cre* recombinase is under control of the CMV promoter. The mouse cytomegalovirus (CMV) immediate-early promoter is normally used to produce high-level recombinant protein (Rita Costa et al., 2010), in this case the CRE recombinase, in a murine cell. It has been shown that this CMV promoter is active in most cell types (Schlabach et al., 2010), including all types of lung cells (Schlabach et al., 2010). Therefore, it can be assumed that this vector targets a wide range of cells-of-origin. However, it was interesting to analyse these tumours to incorporate the transcriptomic analysis into the CMV tumour research in our research group, secondly, to bring these results closer to the literature where CMV virus is commonly used and thirdly, to use it as a positive control of cell transformation.

To get a full understanding of the differences and similarities of SPC, CC10 and CMV tumours, several techniques have been used. One of them is transcriptomics. Transcriptomics is the study of the “transcriptome”, in other words, the complete set of transcripts in a cell, including mRNAs, non-coding RNA and diverse small RNAs, and their quantities in order to understand the tissue physiological stage (Zhong Wang, 2009). There are several sequencing methods but the most used one nowadays is the RNA sequencing (RNAseq), a high-throughput DNA sequencing method (Zhong Wang, 2009).

Overall, RNAseq converts the tissue isolated RNA into a cDNA library with certain sequencing adaptors attached to their ends. A high-throughput DNA sequencing technology is then used to sequence the cDNA fragments and obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing). This method has several advantages, for instance, this technique

has very low background signal, it is highly accurate to quantify genes expression levels, it has high levels of reproducibility both technical and biological, it requires less RNA levels than other sequencing techniques and generally at a lower cost (Zhong Wang, 2009).

The first step of the data analysis is quality control of the raw data, which removes low quality reads from the analysis (Hong et al., 2020). Afterwards, the resulting fragments are then aligned to reference longer genome sequences corresponding to the RNA in the original sample in order to generate a transcription map which described the transcripts (Zhong Wang, 2009, Skerrett-Byrne Anthony et al., 2023). Then, expression can be quantified and normalised with several bioinformatic tools to produce an expression matrix, which can be used to analyse the differential expression analysis in multiple groups with several statistics and bioinformatic tools (Hong et al., 2020).

Transcriptomics have several applications in lung cancer. For instance, differential expression analysis by RNA sequencing between tumour and normal tissue or between tumour types has allowed the researchers to identify biomarkers for tumour diagnosis, prognosis and therapeutic planning. Moreover, RNAseq differential expression analysis in tumour overtime or before and after the treatment has been useful to identify the heterogeneity within the tumour, the tumour evolution and the drug resistance strategies. Additionally, transcriptomics can be used to identify resistance effectors or targetable antigens in immunotherapy (Hong et al., 2020). In the field of lung cancer, transcriptomics has been recently used to classify lung tumours into different transcriptional subtypes. Analysis of transcriptomic data from over 800 KRAS active early-stage and advanced-stage tumours have identified three robust subtypes dubbed mucinous (also known as subtype 1 (S1) in KP mouse model), proliferative (also known as subtype 2 (S2) in KP mouse model), and mesenchymal with respective pathway phenotypes. Proliferative tumours showed low expression of inflammatory and immune response genes and less active KRAS signalling. Proliferative tumours also displayed enhanced oxidative phosphorylation, high expression of E2F targets, genes involved in the G2-M-phase checkpoint, and MYC targets. Expression of cell cycles genes is low in the mesenchymal subtype and even lower in the mucinous subtype. These transcriptional subtypes were confirmed in cell lines and patient-derived

xenografts. KP GEMMs showed only mucinous and proliferation subtypes (Daemen et al., 2021).

Pathway analysis for RNAseq data is normally carried out with GSEA. GSEA is a computational method that determines if an a priori defined set of genes shows a statistically significant difference between two biological states. This analysis considers two groups of samples and the genes expressed are ranked based on the correlation between their expression and the class distinction. Given a previously defined set of genes  $S$ , the goal of this analysis is to determine whether the members of  $S$  are randomly distributed throughout sample group gene list or primarily found at the top or bottom (Aravind Subramanian, 2005). As a result, this analysis returns several parameters. The enrichment score (ES) which reflects the degree to which a gene set  $S$  is overrepresented at the extremes, top (positive ES) or bottom (negative ES), of the entire ranked list  $L$ . Additionally, it provides an ES nominal P value which adds a statistically significant value of the differences seen by the ES. Adjustment for multiple hypothesis testing is also provided, where the statistical significance is adjusted to account for multiple hypothesis testing. First the ES is normalised by the size of the gene set  $S$ , resulting in the normalised enrichment score (NES). Finally, the false discovery rate (FDR), which is defined as the estimated probability that a set with a given NES represents a false positive finding, is set up for every NES (Aravind Subramanian, 2005).

Three different databases containing multiple gene sets have been used in this project. MSigDB database was originally developed for use with GSEA and it contained 1325 gene sets when it was originally released in 2005 (Liberzon et al., 2011), being the database with the largest number of gene sets overall and it represents specific well-defined biological states or processes, and they contain genes that display coordinate expression on that biological state or process (Liberzon et al., 2015). GO BP is the second database used. GO develops and uses ontologies to cover several domains of molecular and cellular biology. GO BP database brings together the larger processes in the cells accomplished by multiple molecular activities (Harris et al., 2004). Finally, KEGG pathway database, which is a collection of pathway maps representing molecular interaction and reaction for a wide range of pathway such as metabolism,

genetic information processing, cellular processes, drug development, human diseases, among others (Kanehisa et al., 2016).

Proteomic analysis has also been used to characterise these tumours. The “proteome” can be defined as the overall protein content of a cell, regardless of its localization, interactions or post-translational modifications. Most of the functional information of the genes is usually represented by the proteome, thus its study and comprehension are essential for diagnosing, prognosis and cancer development (Aslam et al., 2017). Proteomics is, therefore, the study of the interactions, function, composition, and structures of proteins and their cellular activities (Al-Amrani et al., 2021).

The workflow is normally based on the sample collection, the protein digestions with several enzymes and the peptide labelling (Al-Amrani et al., 2021). There are several labelling methods available for researchers but in this research project isobaric labelling using tandem mass tags (TMTs) has been used, which quantifies relative protein abundances in proteomic studies (Andrew Thompson, 2003). These TMT groups are peptides comprising one “tag” amino acid linked to a sensitization group. These tags can be used to label primary amines in the digested peptides (Zecha et al., 2019). Then, peptides can be separated in liquid chromatography.

Chromatography describes the process of separating and identifying compounds from a mixture of substances based on chemical properties. This technique uses a mobile phase, liquid or gas, where the sample in question is mixed and is passed through a column filled with a solid stationary phase. This method identifies the substances by measuring how long it takes for the sample to move through the stationary phase. Therefore, gas chromatography will detect volatile substances while liquid chromatography will identify soluble molecules. In other words, gas chromatography will separate hydrophobic substances like lipids and fatty acids, while liquid chromatography will separate the rest of the soluble cell components such as proteins or other metabolites (Alseekh et al., 2021). Afterwards, mass spectrometry can be used together with liquid chromatography to determine proteins in the samples. These molecules are ionized, and their mass is calculated according to mass-to-charge ratios ( $m/z$ ). The mass spectrometer has three main components: an ion source, an analyser, and a

detector. The ionizer in this case have been ESI, which create charged droplets that increase gaseous ion production, which then are analysed (Al-Amrani et al., 2021).

Inside the mass spectrometer, the TMT tags leads to a summation of each peptide signal from all labelled and combined samples in the generated mass spectrometry spectrum. Following peptide fragmentation and ionisation, sample-specific reporter ions of different  $m/z$  values are generated from the different tags owing to the different combinations of heavy carbon and nitrogen isotopes present in the reporter ions. This enables the differentiation and relative quantification of peptides (Zecha et al., 2019).

The applications of proteomics in cancer are wide. It has been used for anticancer drug development, personalisation of cancer therapies, cancer diagnosis and biomarker discovery, and study tumour progression in the early and late stages of the disease (Al-Amrani et al., 2021).

The differences between these tumours have also been studied with metabolomics. Metabolomics is a tool that uses analytical chemistry approaches to broadly identify and quantify the metabolites present in tissues or biofluids. It allows the analysis of metabolites involved in multiple metabolic pathways including glucose and glutamine metabolism, nucleotide biosynthesis, mitochondrial function and lipid metabolism (Manchester and Anand, 2017). This technique consists of the metabolite extraction from the tissue or the cell culture to proceed with metabolite separation. This separation is carried out by liquid or gas chromatography. Afterwards, molecules are ionised to gas-phase ions so that they can be manipulated by external electric and magnetic fields. At that point, the metabolites will be separated based on their  $m/z$  in the mass spectrometer. The ions are then measured, and the  $m/z$  ratios detected together with their relative abundance is measured and a mass spectrum is built. This is the  $m/z$  ratios plotted against their intensities and the height of the peaks reflects the relative abundance of the samples (Alseekh et al., 2021). There are several tools online for metabolomics analysis, one of the most used is Metaboanalyst which was first introduced in 2009 and performs differential and pathway analysis (Xia et al., 2009, Pang et al., 2021). Metabolomics applications in cancer include the identification of biomarkers by differential analysis,

biomarkers for imaging analysis such as PET imaging, and for the study of drug therapy or the study of cancer drug resistance (Beger, 2013).

Finally, PET imaging has also been used to characterise and classified these tumours. This is a technique that measures physiological function by looking at several nutrients' uptake. The technique is based on the detection of radioactivity emitted from a radioactive tracer after this tracer is intravenously injected into the patients. The tracer is usually labelled with oxygen-15, fluorine-18, carbon-11, or nitrogen-13. As for cancer research, accumulation of the radiolabelled tracer allows measurement of the uptake of this tracer (Berger, 2003).

The most commonly used PET tracer is fluorine-18 ( $^{18}\text{F}$ ) covalently bound to the glucose analogue 2- $^{18}\text{F}$ fluoro-2-deoxyD-glucose ( $^{18}\text{F}$ FDG). Its signal indicates the increase of glucose uptake.  $^{18}\text{F}$ FDG is trapped within cells. This analogue is internalized in the cell by glucose transporters and phosphorylated by a HK but does not undergo further metabolism in glycolysis (Telo et al., 2020).  $^{18}\text{F}$ FDG PET CT is commonly used in routine as it plays an important role in tumour staging, with lung cancer (Fischer and Mortensen, 2006), although not all types of lung cancer tumours are  $^{18}\text{F}$ FDG avid. For instance, adenocarcinomas are generally less  $^{18}\text{F}$ FDG-avid than SCC, while mucinous lesions and lepidic predominant adenocarcinomas show low FDG uptake (Telo et al., 2020). This is part of the reason why other PET tracers have been or are being developed to fully understand, classify and detect lung cancer tumours. One of these tracers is  $^{11}\text{C}$ acetate.  $^{11}\text{C}$ acetate has been used to assess tricarboxylic acid (TCA) cycle activity, and to assess de novo fatty acid synthesis in tumours, depending on the acquisition time (Lewis et al., 2014).  $^{11}\text{C}$ acetate has shown promise as a tracer for slow-growing tumours that do not show high levels of  $^{18}\text{F}$ FDG uptake (Ilaria Grassi, 2012).

## 5.2 Hypothesis and Aims

As explained in chapter 1, the pure effect of lung cancer cell-of-origin in metabolic phenotype and heterogeneity has never been studied. Our hypothesis claims that lung tumours arising from different lung cancer initiating cells may result in lesions with a distinctive metabolic signature. Therefore, the aim is to

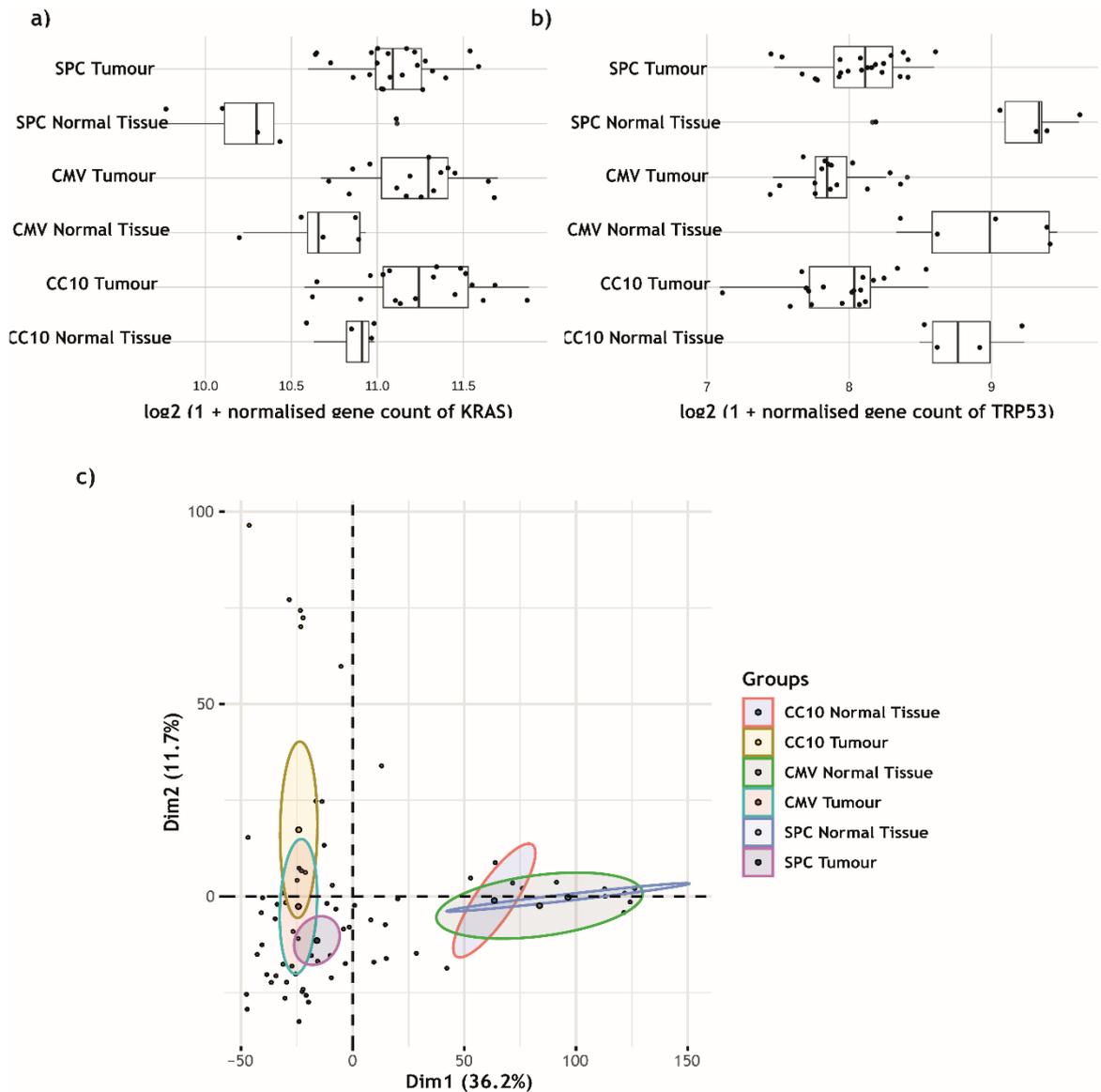
induce the mouse models developed in chapter 4, dissect and analyse tumours using transcriptomics, proteomics and metabolomics to determine the effect of the cell-of-origin on the lung tumour phenotype.

## 5.3 Results

### 5.3.1 Transcriptomic results tumour vs normal tissue

MRI screening was performed in CMV, CC10 and SPC mice and the tumour diameter was calculated. Lesions from 3 to 6 mm in diameter were dissected and cut in half. The first half was used for DNA, RNA and protein extraction and the second half was used for metabolites extraction. Lung tissue of healthy lung regions where there were no cancer lesions visible, were also dissected and used for DNA, RNA, protein and metabolite extraction. RNA samples were used for transcriptomic analysis, protein isolated was used for proteomic analysis, and extracted metabolites were used for untargeted and targeted metabolomic analysis.

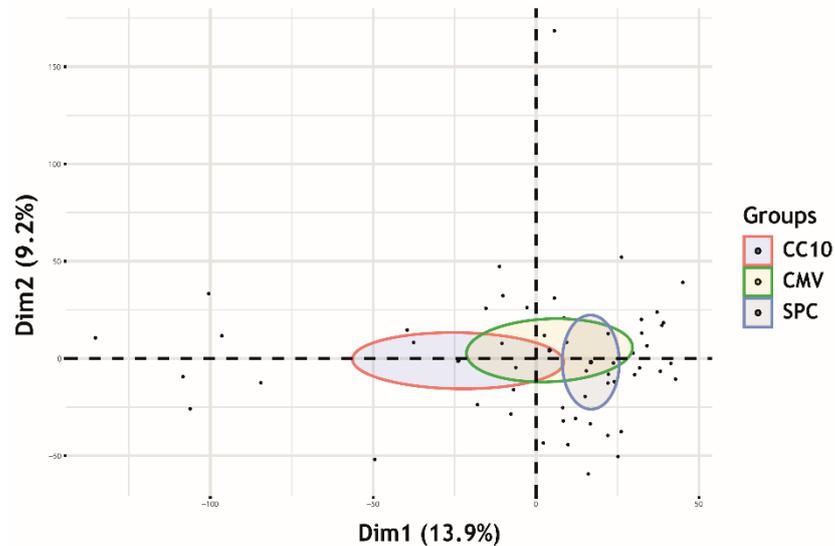
Firstly, an initial transcriptomic comparison between SPC, CC10 and CMV tumours against healthy lung tissue samples, was performed to confirm the tumorigenic process. As seen in figure 5.1c, tumour and normal lung tissue separated using PCA, showing transcriptional differences between the tumour and the healthy lung tissue and corroborating the tumorigenic process. Secondly, *p53* and *KRas* log<sub>2</sub> normalised gene counts (Figure 5.1a and Figure 5.1b) showed the efficiency of the intranasal induction as *KRas* expression is overexpressed in tumour samples compared to normal tissue samples. Also, *P53* expression is downregulated in tumour tissue samples compared to normal tissue samples.



**Figure 5.1. KRas and P53 expression in SPC and CC10 tumours and PCA showing the separation between healthy lung tissue and tumour samples.** KRAS expression (a) and P53 expression (b) bar plots indicate the mean of each group and the  $\pm$ SD. Each dot represents the data of one sample. (c) PCA was performed and plotted with factoextra R package after differential gene expression. The sample groups were set as the variable of interest and a confidence threshold was applied to each group. The elliptical around each group representing the mean and 95% confidence interval for each sample group. The mean for each group lies in the centre of each elliptical, with the outline the 95% confidence interval. Each data point represents a single tumour / normal tissue sample. SPC normal tissue n=10, CC10 normal tissue n=7, CMV normal tissue n=8, SPC n=22, CC10 n=18, CMV n=17.

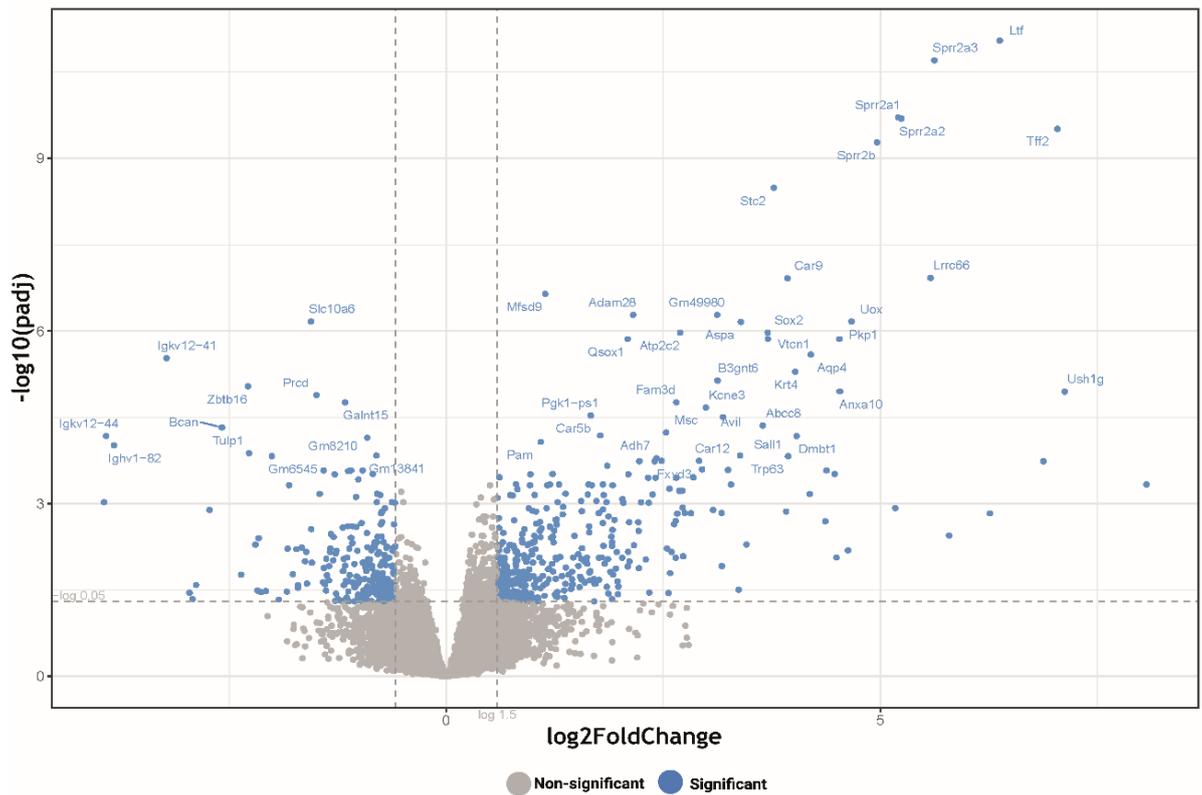
## 5.3.2 Differences between CC10 vs SPC tumours

### 5.3.2.1 Transcriptomic analysis CC10 vs SPC



**Figure 5.2.** CC10 and SPC groups are separated in an unsupervised PCA analysis. PCA was performed and plotted with factoextra R package after differential gene expression. The sample groups were set as the variable of interest and a confidence threshold was applied to each group. The elliptical around each group representing the mean and 95% confidence interval for each sample group. The mean for each group lies in the centre of each elliptical, with the outline the 95% confidence interval. Each data point represents a single tumour. SPC n=22, CC10 n=18, CMV n=17.

PCA plot in figure 5.2c showed a slight separation between SPC and CC10 tumour samples, indicating certain differences between SPC and CC10 tumours, and both groups overlap with the CMV samples. Given this information, a volcano plot was built to compare individual genes between CC10 and SPC tumours (Figure 5.3). Genes whose expression was increased in CC10 are located to the left of zero in the x-axis while genes whose expression was increased in SPC are illustrated to the right of the x-axis. These genes belong to several pathways related to tumour progression.



**Figure 5.3. Volcano plot showing significant differential genes between CC10 and SPC tumours.** Volcano plot showing significantly altered genes in CC10 tumours compared to SPC tumours. Significantly altered genes with a fold change greater than 1.5 and  $\text{padj} < 0.05$ , are overexpressed in SPC tumours. Significantly altered genes with  $\text{padj} < 0.05$  and a fold change smaller than -1.5 represent the genes overexpressed in CC10 tumours. SPC  $n=22$ , CC10  $n=18$ , CMV  $n=17$ .

A summary of highlighted proteins is shown in table 5.1.

**Table 5.1. A summary of significant genes overexpressed in SPC or CC10 tumours and the pathways they are part of.**

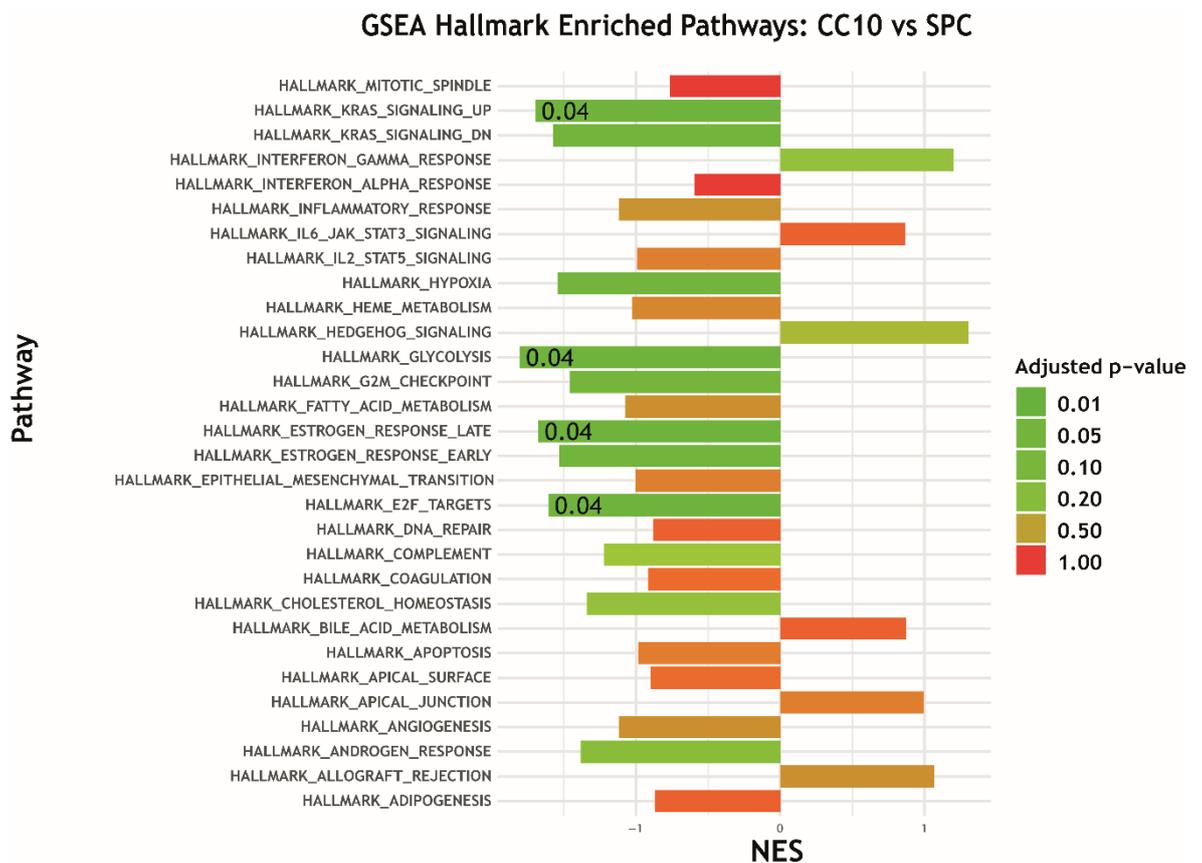
Ad5-mSPC-Cre		Ad5-CC10-Cre	
Gene	Pathway	Gene	Pathway
<i>Ltf</i>	Immune system	<i>Galnt15</i>	Protein glycosylation
<i>Vtn1</i>	Immune system	<i>Zbtb16</i>	Cel cycle
<i>Gzmd</i>	Immune system	<i>Slc10a6</i>	Lipid transport
<i>Fam3d</i>	Immune system	<i>Gm13841</i>	Ribosomal protein
<i>Lrrc66</i>	Cytoskeleton	<i>Gm8210</i>	Ribosomal protein

<i>Sprr2</i> family	Cytoskeleton	<i>EGFR</i> or <i>IGFR</i>	Growth factor receptors
---------------------	--------------	----------------------------	-------------------------

Given the variety of gene types seen above, we considered to run a GSEA. As detailed in section 2.1.16, this analysis has been run using defined gene sets from different gene databases: (a) MSigDB, or (b) KEGG.

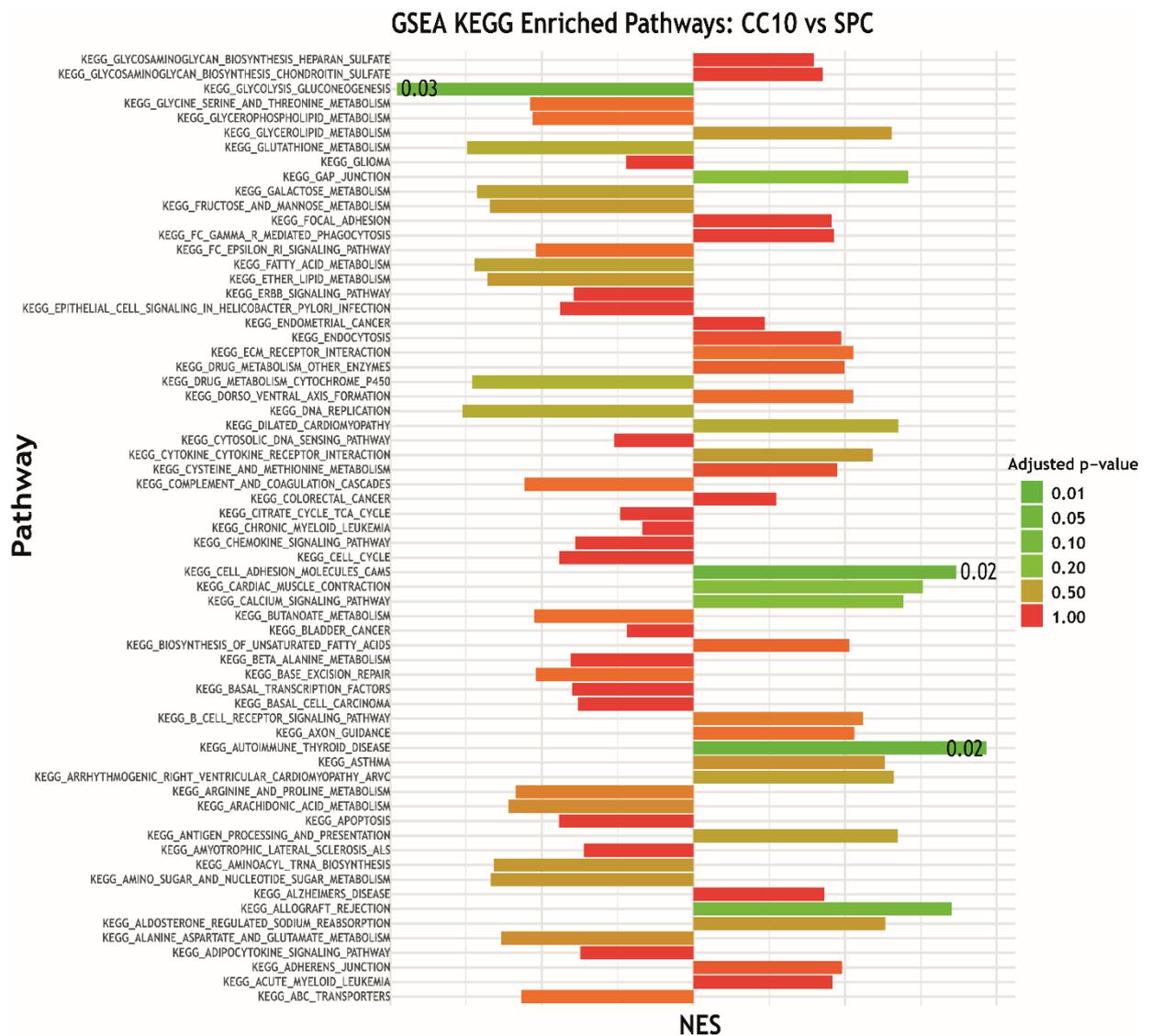
Figure 5.4 shows the results of GSEA with database (a) and figure 5.5 shows the results obtained after performing GSEA with database (b). Both databases showed that glycolysis pathway is significantly upregulated in CC10 samples compared to SPC lesions, suggesting, a difference in glycolytic metabolism in SPC and CC10 tumours.

At the same time, database (a) showed statistically significant pathways like EF2 target genes, KRAS signalling pathway and late response to estrogen pathway seem to be upregulated in CC10 tumours. There are no significant pathways significant upregulated in SPC when compared to CC10 samples in database (a).



**Figure 5.4. Hallmark GSEA transcriptomic analysis highlighting the impact of the cell-of-origin on glycolysis and proliferation-related pathways.** Enrichment plots showing gene set enrichment SPC and CC10 tumours for Hallmark database. Normalised enrichment scores (NES) and FDR-q values displayed on the graphs. RNA sequencing was done by Beatson Molecular Technologies. Pathway analysis and GSEA conducted by Robin Shaw and Adam Peters. SPC n=22, CC10 n=18, CMV n=17.

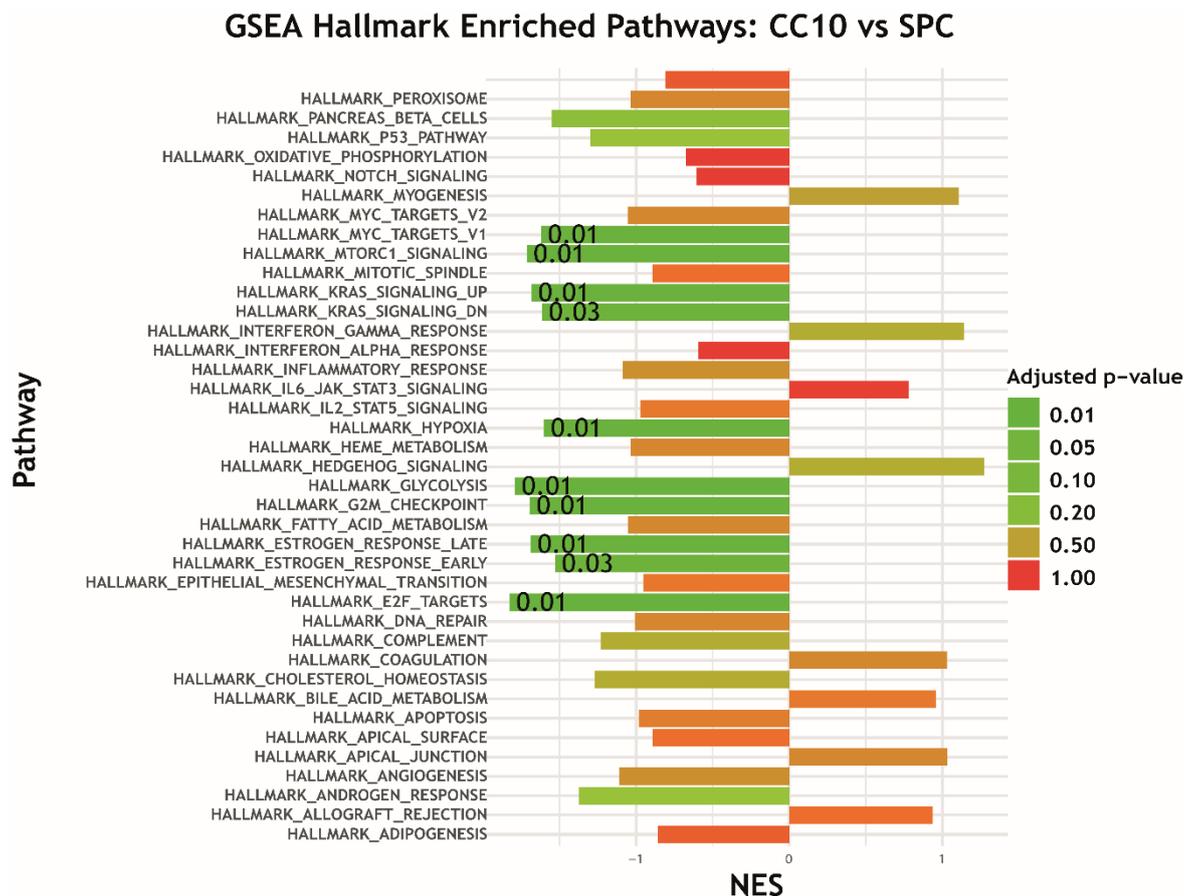
In addition, database (b) showed an upregulation of cell adhesion molecules and genes related to the autoimmune thyroid disease in SPC tumours when compared to CC10 tumours. Glycolysis was the only upregulated pathway in CC10 tumours.



**Figure 5.5. KEGG GSEA transcriptomic analysis highlighting the impact of the cell-of-origin on glycolysis and immune pathways.** Enrichment plots showing gene set enrichment SPC and CC10 tumours for KEGG database. Normalised enrichment scores (NES) and FDR-q values displayed on the graphs. RNA sequencing was done by Beatson Molecular Technologies. Pathway analysis and GSEA conducted by Robin Shaw and Adam Peters. SPC n=22, CC10 n=18, CMV n=17.

However, as seen in figure 5.2, there were several outliers in the PCA plot. We initially thought that these might indicate sampling or processing errors, thus it was decided to separate the effect/bias from the main comparisons. Therefore, these samples were excluded, and this analysis was rerun.

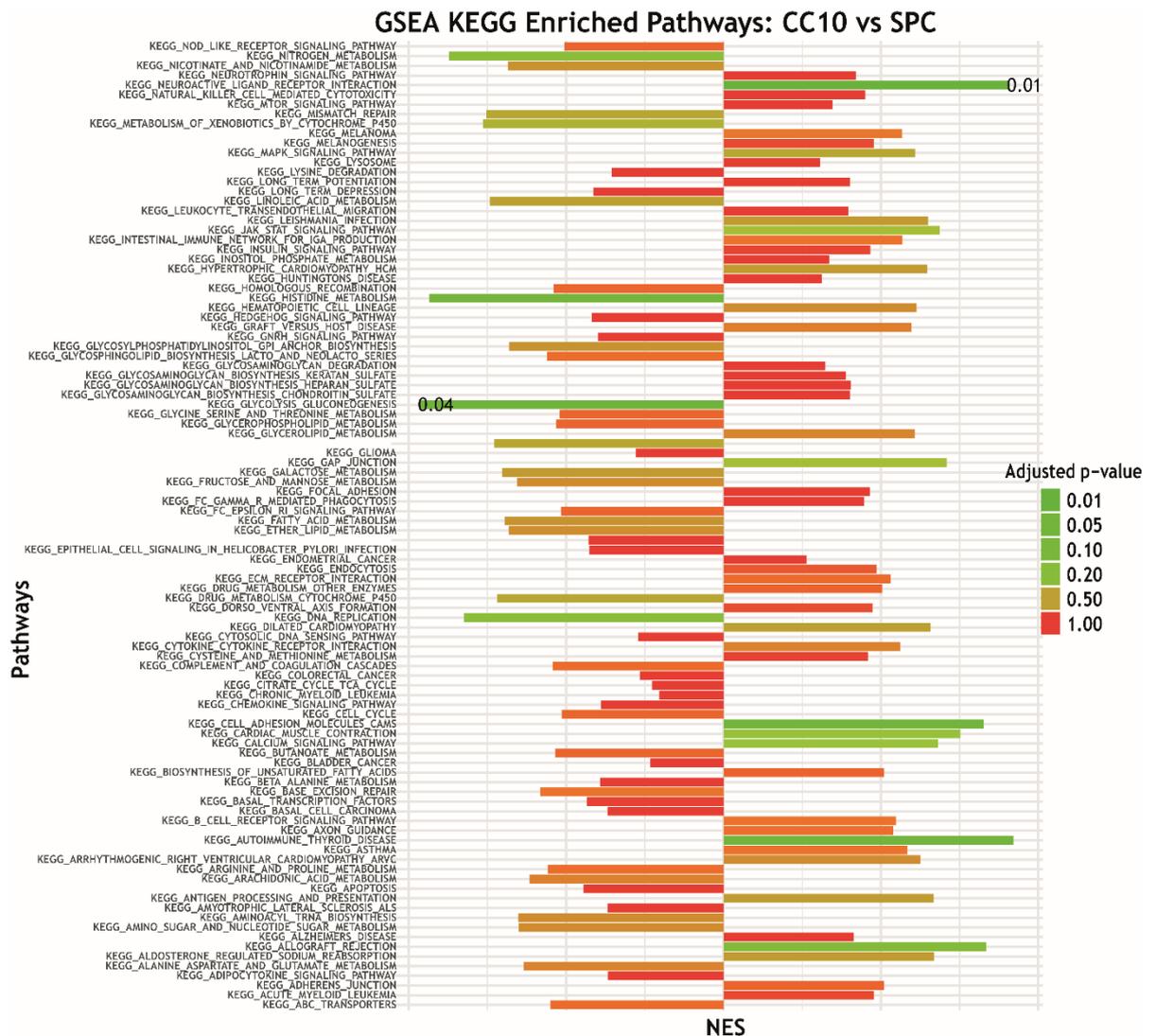
Once these outliers were removed there were several pathways that become statistically significant. The GSEA analysis was run using the same principles and the same gene databases and it showed similar results. Figure 5.6 shows the results of GSEA with database (a) and figure 5.7 shows the results obtained after performing GSEA with database (b). They both showed again a significant upregulation of the glycolysis gene set in CC10 lesions when compared against the SPC tumours, supporting the data above which suggests a difference in glycolytic metabolism in SPC and CC10 tumours.



**Figure 5.6. Hallmark GSEA transcriptomic analysis highlighting the impact of the cell-of-origin on glycolysis and proliferation-related pathways.** Enrichment plots showing gene set enrichment SPC and CC10 tumours for Hallmark database. Normalised enrichment scores (NES) and FDR-q values displayed on the graphs. RNA sequencing was done by Beatson Molecular

Technologies. Pathway analysis and GSEA conducted by Robin Shaw and Adam Peters. SPC n=22, CC10 n=11, CMV n=16.

Database (a) showed that EF2 target genes pathway, MTORC1 pathway, KRAS signalling, and MYC target genes also seem to be significantly upregulated in CC10 tumours when compared against SPC tumours. Moreover, the hypoxia hallmark, the G2M checkpoint and the estrogen related response are also upregulated in CC10 tumours. There are no significant upregulated pathways in SPC tumours.

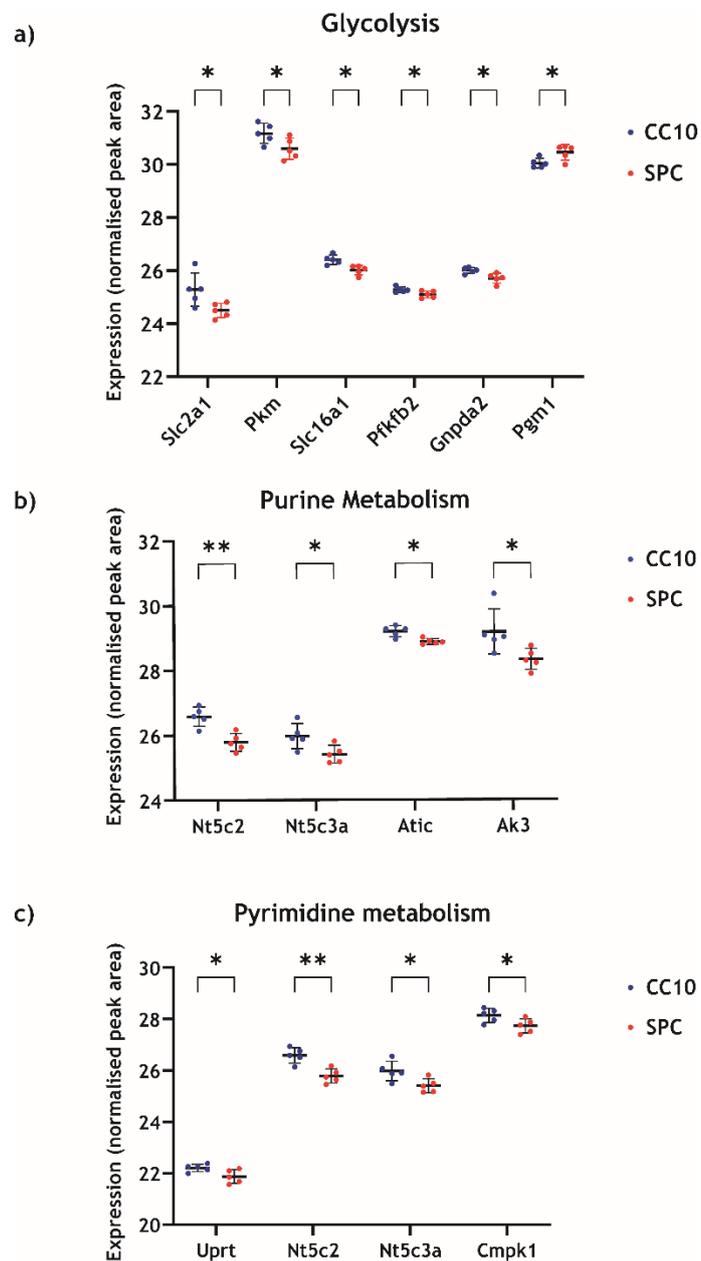


**Figure 5.7. KEGG GSEA transcriptomic analysis highlighting the impact of the cell-of-origin on glycolysis and neuroactive interactions.** Enrichment plots showing gene set enrichment SPC and CC10 tumours for KEGG database. Normalised enrichment scores (NES) and FDR-q values displayed on the graphs. RNA sequencing was done by Beatson Molecular Technologies. Pathway analysis and GSEA conducted by Robin Shaw and Adam Peters. SPC n=22, CC10 n=11, CMV n=16.

Database (b) showed that SPC tumours have a significant upregulation of the neuroactive ligand receptor interaction gene set. Similarly, this analysis did not show any other upregulated pathways in SPC tumours. There are several immune related pathways that were slightly upregulated in SPC tumours although this difference was not significant. For instance, the autoimmune thyroid disease or the allograft rejection pathway.

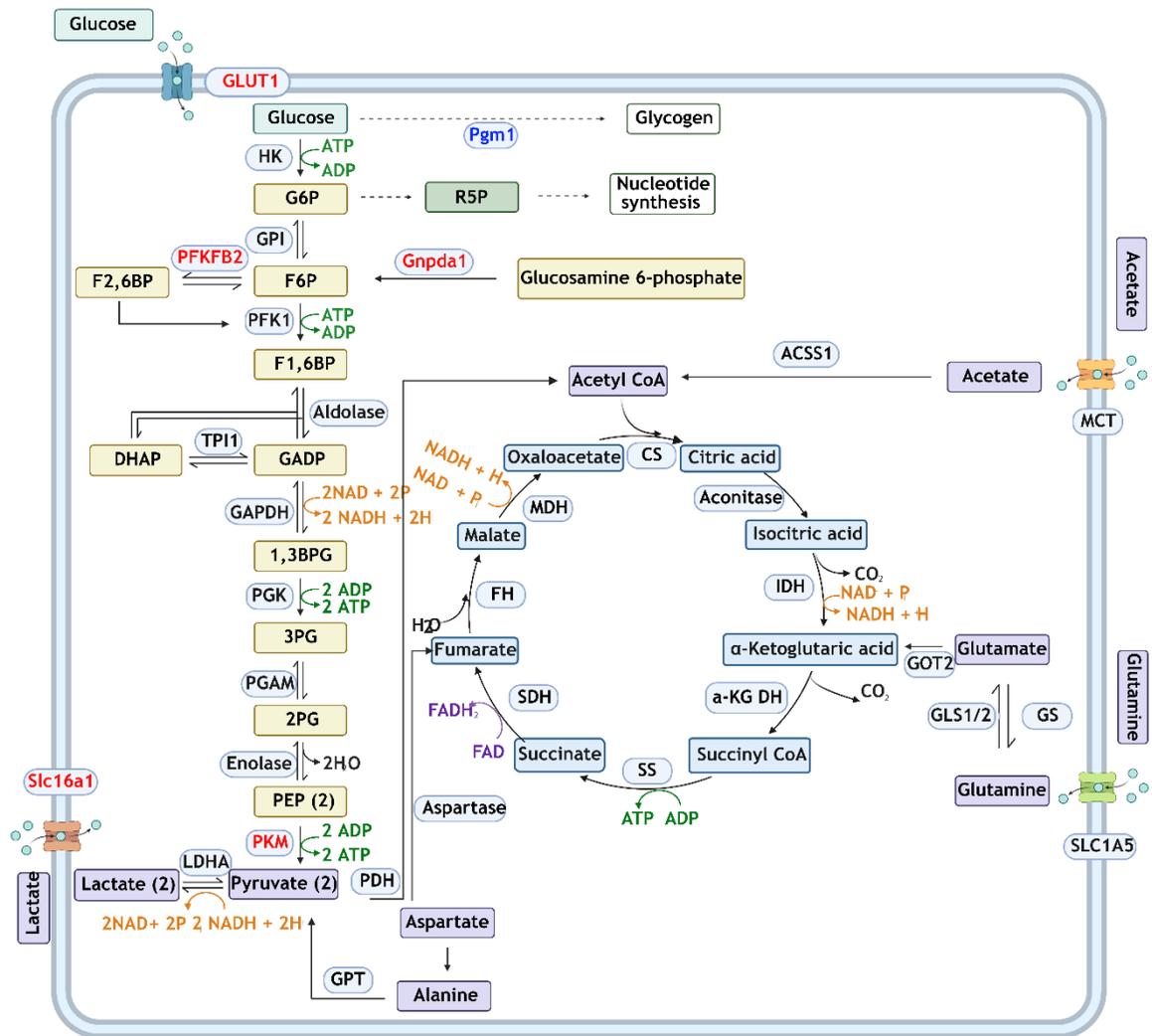
### 5.3.2.2 Proteomic analysis CC10 vs SPC

As explained above, half of the collected tumours were used for RNA and protein isolation. Randomised selection of five SPC and five CC10 samples were sent to proteomic analysis.



**Figure 5.8. Proteomic data indicates an increase of glycolytic and nucleotide synthesis related genes in CC10 tumours.** (a) Represents glycolysis proteins and its expression in SPC and CC10 tumours. (b) This graph represents the identified proteins upregulated in CC10 tumours. (c) Represents pyrimidine synthesis proteins and its expression in SPC and CC10 tumours. Dot plots where each plot represents a tumour. Bars represent the mean and the  $\pm$ SD. Statistical analysis using multiple unpaired t-test, with a 5% FDR in GraphPad Prism was used to determine a statistically significant difference across the two cohorts for each of the graphs presented. SPC n=5, CC10 n=5 and CMV n=5.

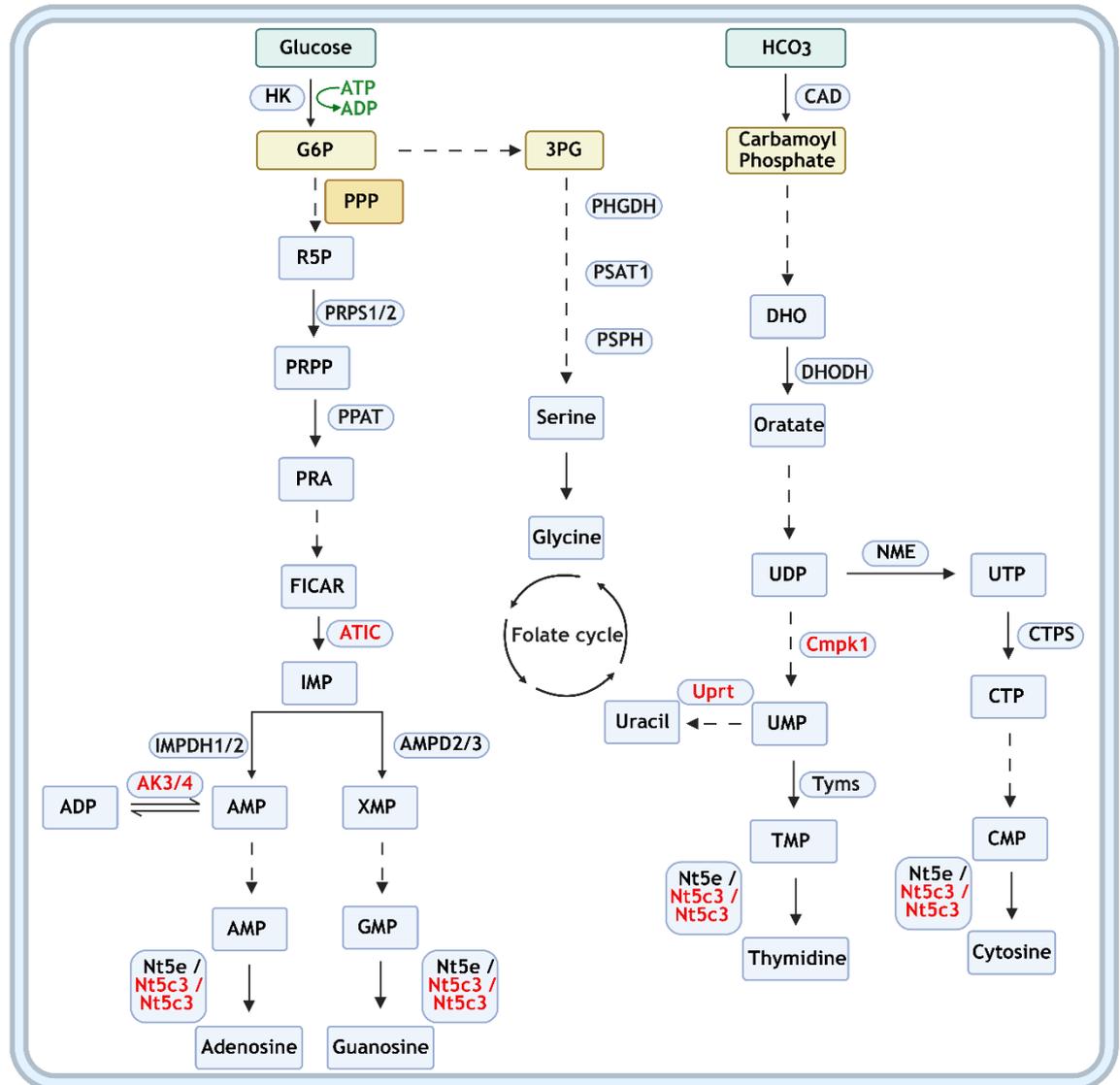
Preliminary proteomic results for some of the genes found overexpressed in the transcriptomic analysis showed a significant increase of several glycolytic proteins in CC10 tumours supporting the results seen in the GSE transcriptomic analysis, including *Slc2a1*, which encodes for GLUT1 transporter, which uptakes glucose into the cell (Ancey et al., 2018); *Pkm* which converts phosphoenolpyruvate into pyruvate in the last step of the glycolysis; *Slc16a1*, which catalyses the transport of pyruvate and lactate across membranes (Li et al., 2015); *Pfkfb2* which controls the synthesis and degradation of fructose-2-6-biphosphate (F2,6BP) from or to fructose-6-phosphate. F2,6BP is a regulator of *PFK1*, which catalyses is the major rate-limiting step in glycolysis and the first committed step. PFK1 activity is increased in the presence of F2,6BP (Lee et al., 2020a); *Gnpda1*, which catalysed the conversion of glucosamine 6-phosphate into fructose 6-phosphate. Finally, *Pgm1*, which is significantly upregulated in SPC tumours and catalyses the interconversion between glucose 1-phosphate (G1P) and G6P. G1P is a precursor of glycogen synthesis and a product of glycogen degradation (Semenza, 2013) (Figure 5.8a and 5.9).



**Figure 5.9. Schematic representation of the glycolysis and the proteins found upregulated in the analysis.** In red, genes upregulated in CC10 samples and in blue, genes upregulated in SPC samples. 1,3BPG 1,3-bisphosphoglycerate, F1,6-BP fructose 1,6- bisphosphate, F6P fructose 6-phosphate, DHAP dihydroxy acetone phosphaye, GAPD glyceraldehyde-3-phosphate, GAPDH glyceraldehyde-3-phosphate dehydrogenase, TPI triosephosphate isomerase GLS glutaminase, GOT2 glutamate oxaloacetate transaminase, GS glutamine synthetase, G6P glucose-6-phosphate, GPI phosphoglucose isomerase, GPT glutamine pyruvate transaminase, GS glutamine synthetase, HK hexokinase, LDHA lactate dehydrogenase A, MCT monocarboxylate transporter, PDH pyruvate dehydrogenase, PEP phosphoenolpyruvate, PFK phosphofructokinase, PG phosphoglycerate, PGK phosphoglycerate kinase, PGAM phosphoglucomutase, PK pyruvate kinase, SLC solute carrier family, TCA tricarboxylic acid, PFKFB2 phosphofructo kinase/fructose 2,6 biphosphatase, R5P ribose 5 phosphate, CS citrate synthase, IDH isocitrate dehydrogenase,  $\alpha$ -KG DH alpha-ketoglutarate dehydrogenase, SS succinyl-CoA synthetase, SDH succinic dehydrogenase, FH fumarate dehydrogenase, MDH malate dehydrogenase . Created with BioRender.com

Moreover, several proteins from the purine and pyrimidine biosynthesis were overexpressed in CC10 samples. *Nt5c2* and *Nt5c3a* participate in the synthesis of nucleotides; *Atic* which synthetise FAICAR and converts it to inosinic acid (IMP) a

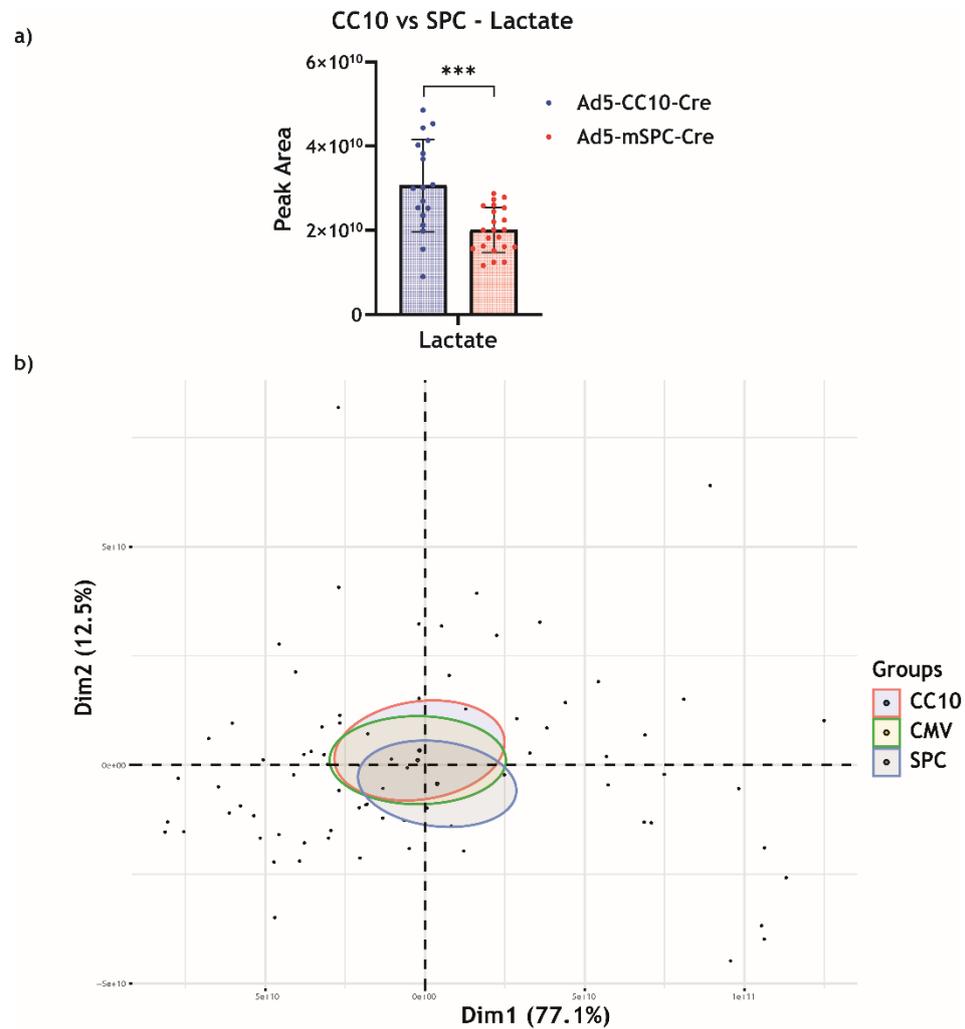
precursor of purines; *Ak3* which catalyses the interconversion of adenosine monophosphate (AMP) and adenosine diphosphate (ADP); *Uprt* which converts uracil monophosphate (UMP) to Uracil and *Cmpk1*, which catalyses the conversion of UDP and UMP (Lane and Fan, 2015) (Figure 5.8b and c and 5.10).



**Figure 5.10. Schematic representation of the nucleotide biosynthesis and the genes found upregulated in this analysis.** In red genes upregulated in CC10 samples. GMP guanosine monophosphate, glucose-6-phosphate, HK hexokinase, PPP pentose phosphate pathway, R5P ribose 5-phosphate, PRPS1/2 phosphoribosyl pyrophosphate synthetase 1, PRPP phosphoribosyl pyrophosphate, PPAT phosphoribosyl pyrophosphate amidotransferase, IMP inosine monophosphate, IMPDH inosine-5'-monophosphate dehydrogenase, AMP adenosine monophosphate, ADP adenosine biphosphate, AK3/4 Adenylate Kinase, Nt5e 5'-Nucleotidase Ecto, XMP xanthosine monophosphate, PHGDH phosphoglycerate dehydrogenase, PRA 5-phosphoribosylamine, PRPS pyrophosphate synthetase, PSPH phosphate ester hydrolysis, GSH glutathione. Created by BioRender.com.

### 5.3.2.3 Metabolomic analysis CC10 vs SPC

Based on transcriptomics data, a targeted metabolomic analysis was run to investigate the upregulation of glycolytic genes in CC10 tumours.



**Figure 5.11. CC10 and SPC samples are do not differ a lot from a metabolic point of view. (a)** Column bar representing the lactate levels in SPC and CC10 tumours. Unpaired t-test was run on GraphPad showing significant differences in lactate between SPC and CC10 tumours. Each dot represents the lactate levels from a specific tumour. Horizontal line represents the mean and the errors bars represent  $\pm$ SD (b) PCA with metabolomic data as input was performed and plotted with factoextra R package after differential gene expression. The sample groups were set as the variable of interest and a confidence threshold was applied to each group. The elliptical around each group representing the mean and 95% confidence interval for each sample group. The mean for each group lies in the centre of each elliptical, with the outline the 95% confidence interval. Each data point represents a single tumour. SPC n=22, CC10 n=18, CMV n=17. \*\*\* p<0.001

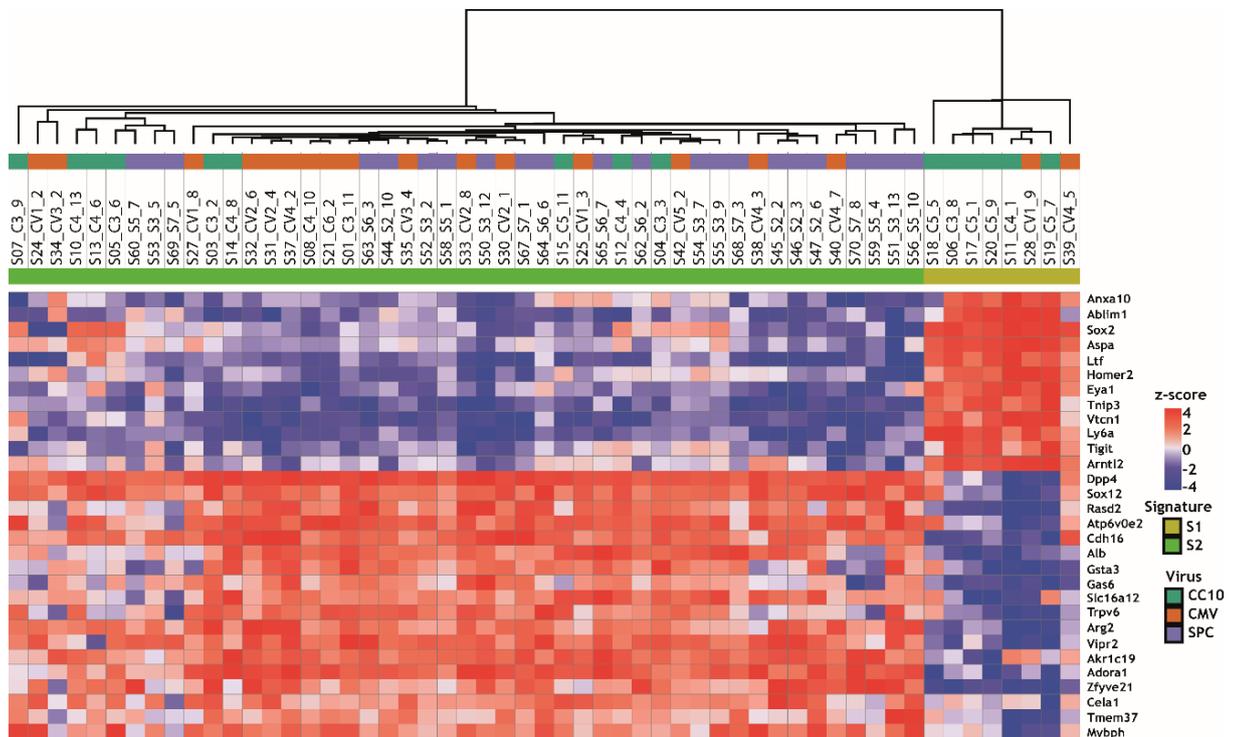
Only lactate had enough statistical power to be considered different between CC10 and SPC samples (Figure 5.11a). This metabolite is significantly higher in

CC10 tumours than in SPC tumours. PCA analysis was run to demonstrate the low separation among these three groups, suggesting few differences among them (Figure 5.11b).

### 5.3.3 Description of S1, S2 and ACE subgroups

Given the statistical changes observed after removing the outliers samples, we decided to explore this further. We came across with a paper from Melissa Junttila which identifies several transcriptional phenotypes within human and mouse lung adenocarcinomas (Daemen et al., 2021). As described in this chapter introduction, S1 and S2 subtypes were identified in KP mouse models.

Given this new information, we plotted these SPC, CC10 and CMV samples in a heatmap with S1 and S2 gene signatures. As seen in figure 5.12, there were 7 samples which fitted into the S1 tumour subgroup, 6 CC10 and 1 CMV samples. These samples turned out to be the same outliers that we excluded in the previous analysis.

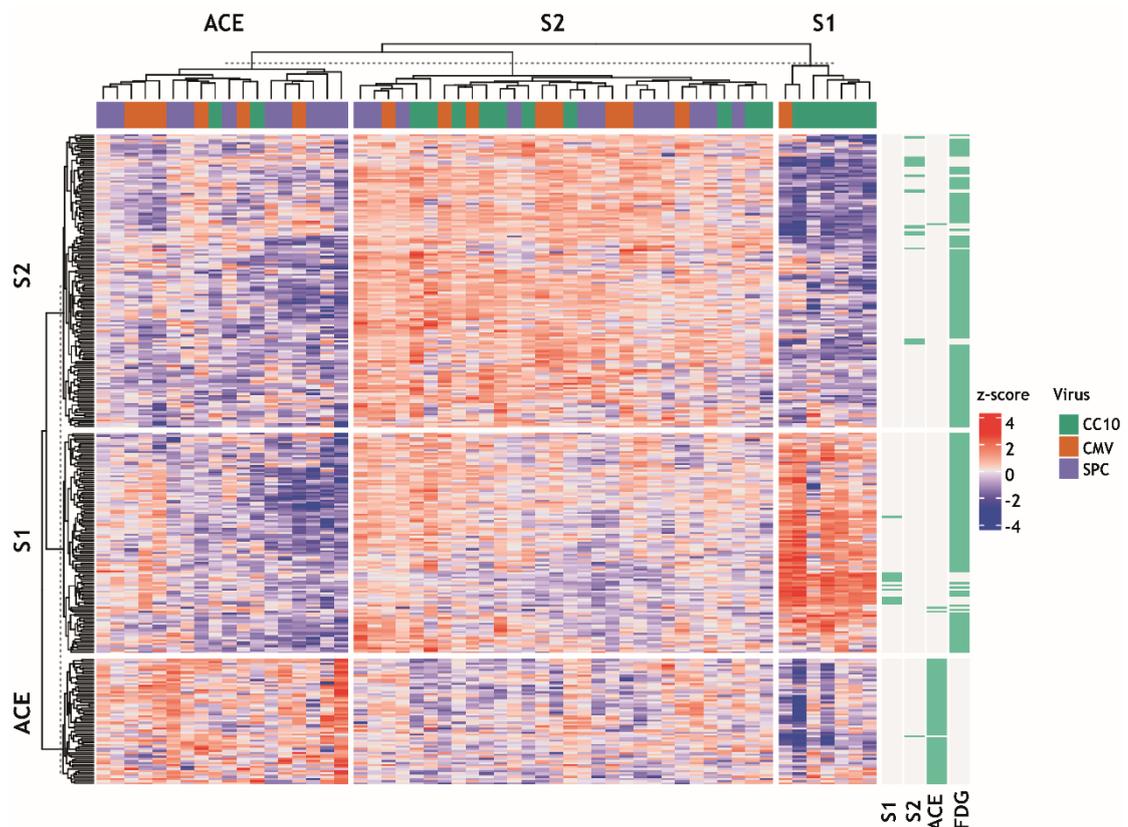


**Figure 5.12.** Gene expression patterns of S1 and S2 gene signatures, revealed at least two transcriptomic subtypes. Heatmap shows expression pattern of the S1 and S2 gene signatures in SPC, CC10 and CMV samples. Heatmap was generated using z-scores of gene expression values

acquired from RNA sequencing of 3-6mm diameter tumours. Heatmap was generated in Morpheus. SPC n=22, CC10 n=18, CMV n=17.

At this point, it was known that the tumours we collected belong to two different transcriptomic subtypes. At the same time, our research group has been working to elucidate the metabolic heterogeneity found in murine lung cancer. To this aim, KP mice induced with the CMV virus have been imaged using two PET imaging tracers, [ $^{11}\text{C}$ ]acetate and [ $^{18}\text{F}$ ]FDG. It has been observed that KP murine tumours can be classified into [ $^{11}\text{C}$ ]acetate avid tumours and [ $^{18}\text{F}$ ]FDG avid tumours depending on their uptake preferences for one tracer or the other (unpublished data). Based on these studies, [ $^{11}\text{C}$ ]acetate and [ $^{18}\text{F}$ ]FDG gene signatures have been developed to determine the uptake preferences of dissected tumours (Supplementary information).

After these findings, we included in the analysis the [ $^{11}\text{C}$ ]acetate and [ $^{18}\text{F}$ ]FDG gene signatures.



**Figure 5.13. Gene expression patterns of S1, S2, FDG and Ace gene signatures, revealed ACE as novel transcriptomic subtype.** Heatmap shows expression pattern of the S1, S2 ACE and FDG gene signatures in SPC, CC10 and CMV samples. Heatmap was generated using z-scores of gene

expression values acquired from RNA sequencing of 3-6mm diameter tumours. Heatmap was generated in R. SPC n=22, CC10 n=18, CMV n=17. Tumour subtypes were annotated by clustering the normalized gene counts of signature genes (S1, S2, ACE, FDG) into three clusters (ACE, S1, S2) using k-means. SPC n=22, CC10 n=18, CMV n=17.

Figure 5.13 shows a heatmap with three different clusters. These clusters were calculated by simple K-means on row-scaled normalised gene expression of the genes on S1, S2, [<sup>11</sup>C]acetate and [<sup>18</sup>F]FDG gene signatures. Clusters S1 and S2 are represented by the expression of S1 and S2 signature genes defined in Melissa Junttila's paper and by [<sup>18</sup>F]FDG tumour gene signature, suggesting that these two subtypes are [<sup>18</sup>F]FDG avid. S1 subgroup consists mainly of CC10 tumours. S2 subgroup is a mix of CC10, CMV and SPC samples. Finally, we introduced a third subtype of lung adenocarcinoma: the Acetate (ACE) group. This third subtype is characterised mainly for SPC samples, and they are characterised by the high expression of [<sup>11</sup>C]acetate gene signature, suggesting that they are [<sup>11</sup>C]acetate avid.

In total, 46% of SPC tumours were classified as ACE tumours and 56% of them as S2 tumours. Moreover, 89% of CC10 tumours were defined as either S1 or S2 tumours, and only an 11% was classified as ACE tumours. The enrichment of SPC tumours in the ACE group and the enrichment of CC10 tumours in S1 group is statistically significant, suggesting an association between two of these transcriptomic subtypes and the cell-of-origin of the tumour (Figure 5.14).

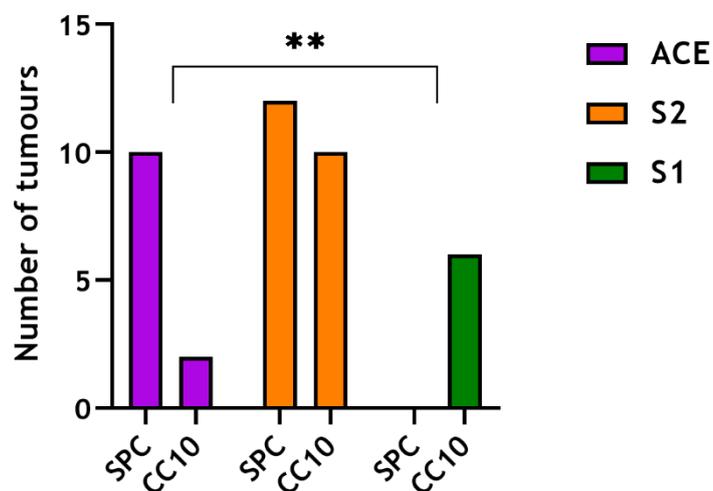


Figure 5.14. SPC tumours are significantly enriched in ACE samples while CC10 tumours are enriched in S1 group. Fisher exact test was used to analyse this data. This test was run with a

95% confidence interval and running a two-sided t-test comparison. SPC n=22, CC10 n=18, CMV n=17. \*\* p<0.001.

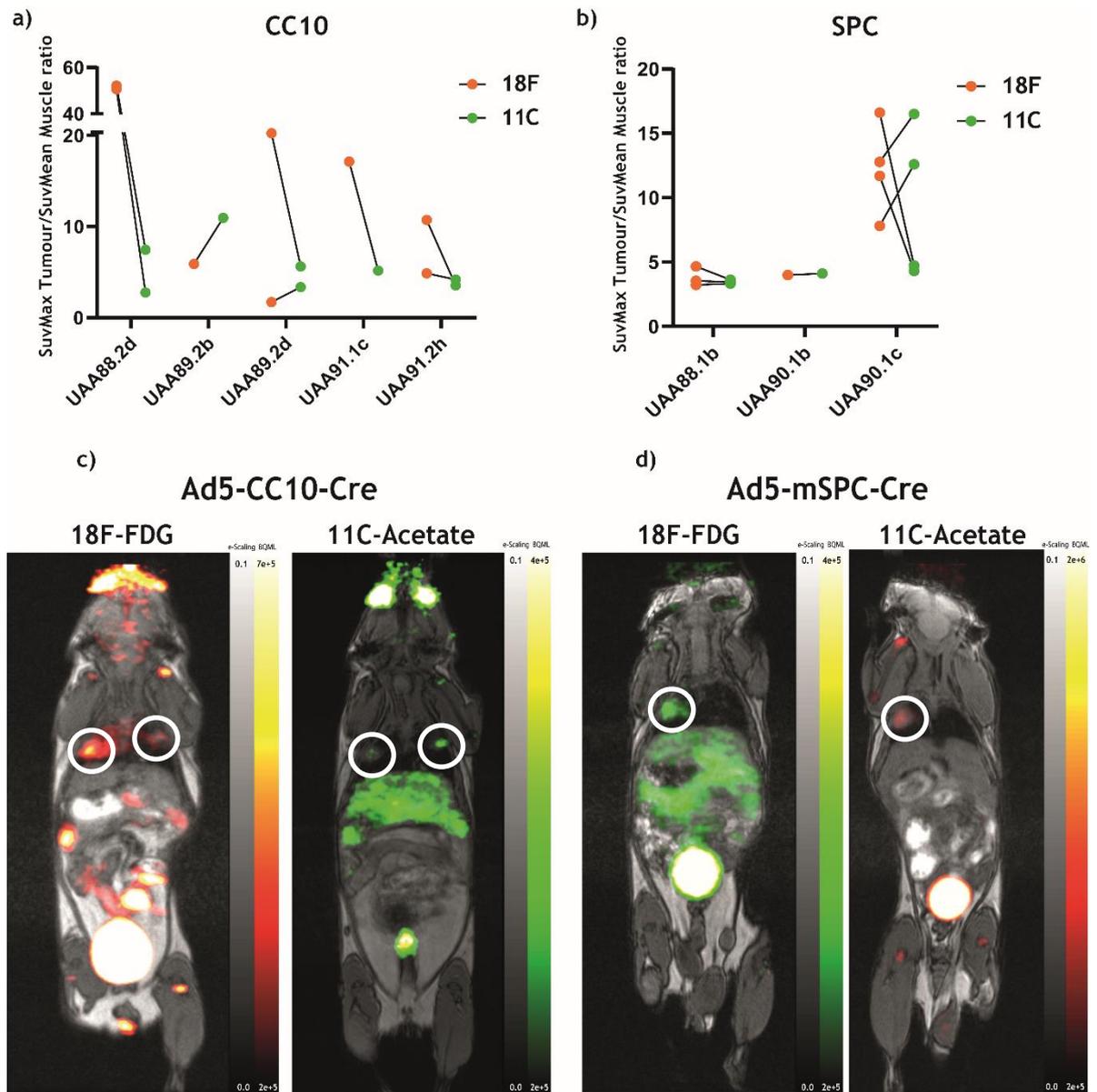
To further validate this analysis, tumours from KP mice induced with Ad5-mSPC-Cre and Ad5-CC10-Cre vectors were imaged using PET imaging and [<sup>18</sup>F]FDG and [<sup>11</sup>C]acetate tracers.

### 5.3.4 PET imaging

S1 and ACE tumours from KP mouse model has been defined due to the clusters made by simple K-means on row-scaled normalised gene expression of the genes on S1, S2, [<sup>18</sup>F]FDG and [<sup>11</sup>C]acetate gene sets and KP murine tumours classification have been refined from two to three groups.

This classification shows SPC tumours fit into two subgroups: 50% of tumours belong to ACE groups and the other half belongs to S2 groups (figure 5.14). At the same time, CC10 tumours fit mainly into two groups: over 75% of samples fit in S2 and S1 groups, while only less than 25% fit into the ACE group (figure 5.14). S1 and S2 tumour subgroups are enriched in genes from [<sup>18</sup>F]FDG gene signature and ACE tumours are enriched in genes from [<sup>11</sup>C]acetate gene set. This means that S1 and S2 tumours are [<sup>18</sup>F]FDG avid while ACE tumours are [<sup>11</sup>C]acetate avid.

Interestingly, when CC10 and SPC tumours were PET imaged with the same two tracers, the same proportions explained above were observed, corroborating this tumour classification. As seen in figure 5.15a, 8 CC10 tumours were imaged. From these tumours, six of them were showed a higher uptake of [<sup>18</sup>F]FDG than [<sup>11</sup>C]acetate compared to only two CC10 tumours which preferred [<sup>11</sup>C]acetate uptake to [<sup>18</sup>F]FDG. On the other hand, from 8 SPC imaged tumours, 4 of them had a higher uptake of [<sup>18</sup>F]FDG while the other 4 of them preferred the uptake of [<sup>11</sup>C]acetate (Figure 5.15b). These follows the same trends described in the previous section and it confirms the glycolytic phenotype of CC10 tumours identified with the omics techniques.



**Figure 5.15. SPC and Cc10 tumours differ in [ $^{18}\text{F}$ ]FDG and [ $^{11}\text{C}$ ]acetate uptake.** Mouse in pairs were scan at the same time. The first day 200-250 MBq of [ $^{11}\text{C}$ ]acetate was injecting. Then between 10-15 MBq of [ $^{18}\text{F}$ ]FDG (a) PET imaging data was analysed and the proportion of tumours in either [ $^{18}\text{F}$ ]FDG-avid or [ $^{11}\text{C}$ ]acetate-avid subgroups was the same shown in figure 5.15. (a) and (b) Paired data graph shows [ $^{18}\text{F}$ ]FDG and [ $^{11}\text{C}$ ]acetate measurement of every tumour. (c) and (d) represent the combination of T1 MRI and the PETs scan. Representative of CC10 and SPC induced mice are shown. White circle indicates the presence of a tumour.

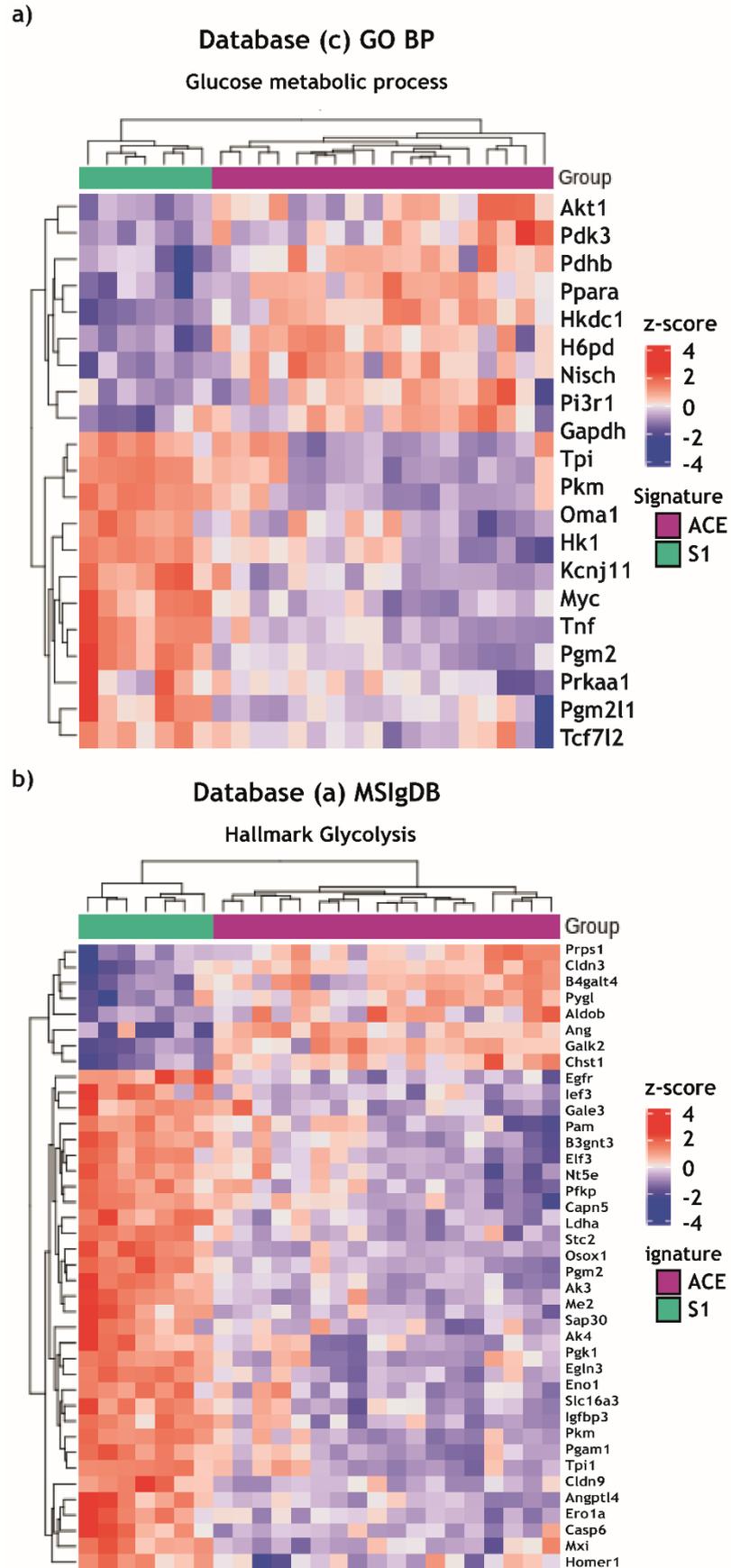
### 5.3.5 Differences between S1 and ACE tumours

#### 5.3.5.1 Transcriptomic characterisation of S1, S2 and ACE subgroups

Given the significant enrichment of CC10 and SPC tumours in S1 and ACE transcriptomic groups, respectively, pathway analysis was run to investigate the potential differences between S1 and ACE tumours. In order to characterise

these subgroups, multiple genes databases were used: (a) MSigDB, (b) KEGG and (c) GO BP. GSE analysis have been performed using these databases and the gene sets they contained. The statistically significant genes among those gene sets have been plotted as heatmaps, which allow us a deeper understanding of each individual gene expression. The genes include in every gene set varies depending on each database criteria.

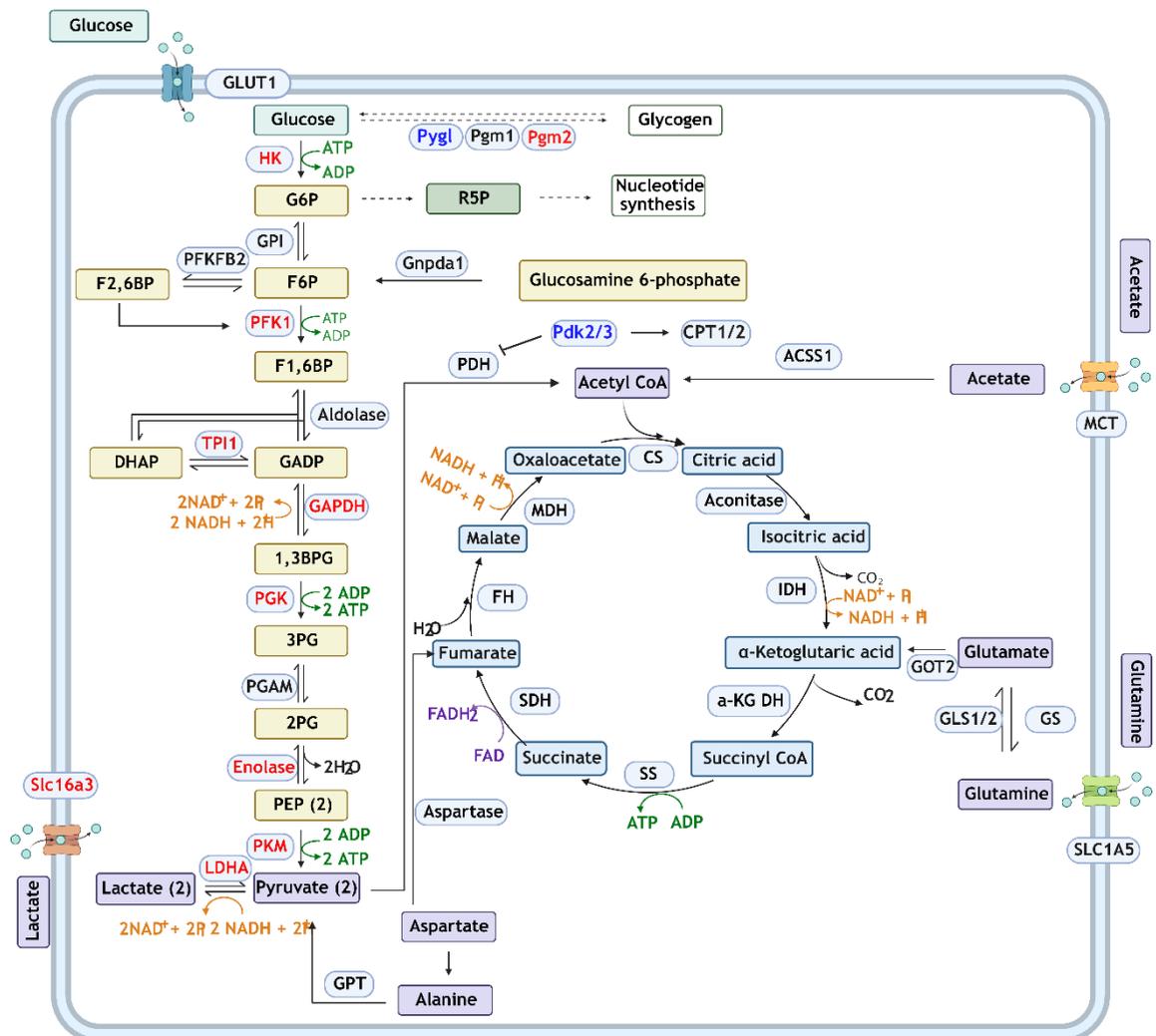
GSE analysis run with database (a) and (c) showed the most interesting results. Firstly, they both showed a significant enrichment of the glycolysis gene set in S1 subgroup, suggesting a higher glycolytic activity in S1 tumours when compared to ACE tumours. Heatmaps were built with the expression of each one of the statistically significant genes between S1 and ACE comparison of these gene sets.



**Figure 5.16.** Gene expression patterns of S1, and ACE samples of significant genes in glycolysis gene sets. Heatmap shows expression pattern of significant genes of glycolytic genes in glycolysis gene sets in S1 and ACE tumours. Heatmap was generated using z-scores of gene

expression values acquired from RNA sequencing of 3-6mm diameter tumours. Heatmap was generated in R. Samples clustering was performed using k-means. S1 n=7, ACE n=18.

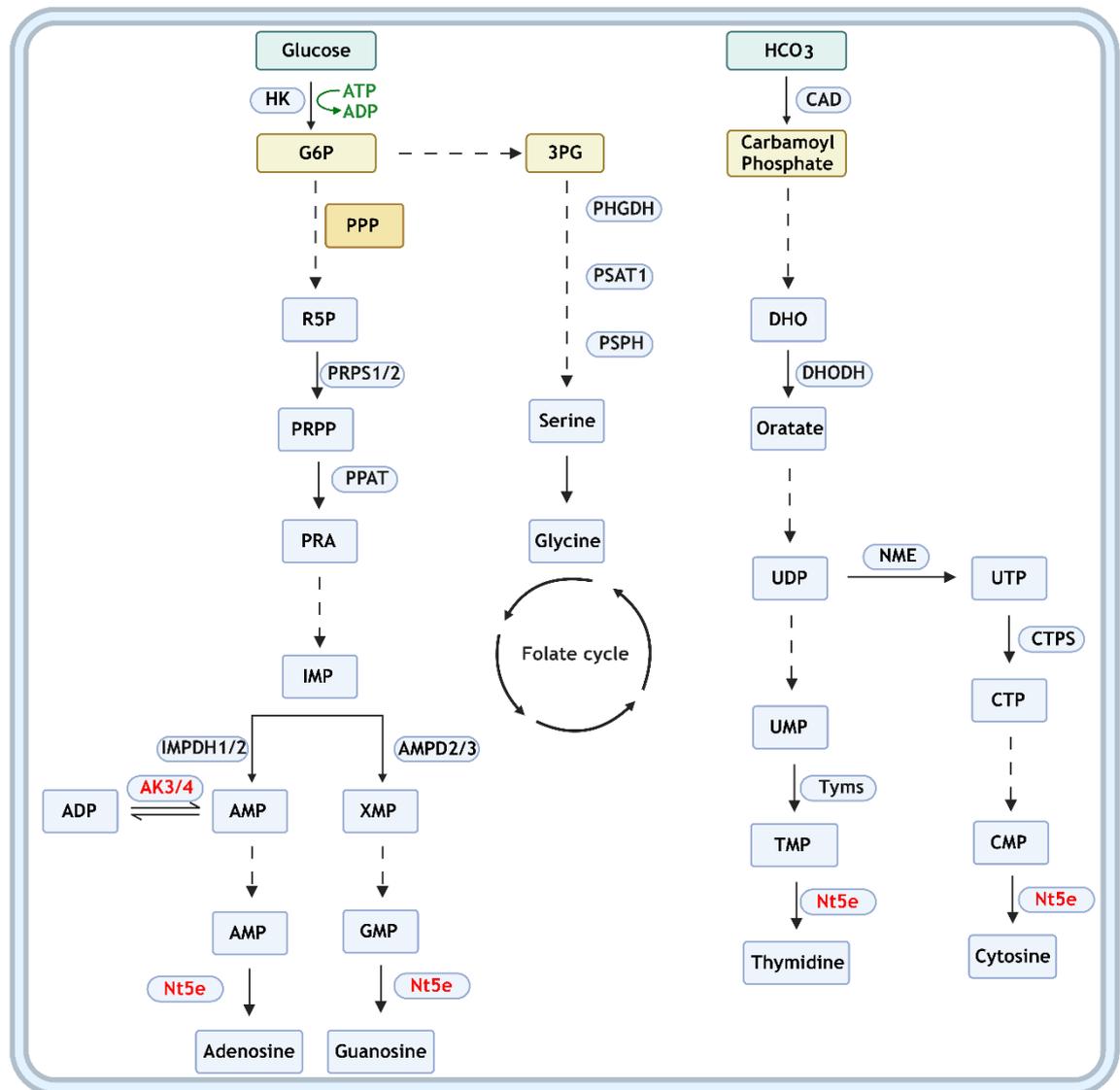
These plots showed that in database (c) glycolytic genes increased expression is predominant in S1 tumours (Figure 5.16a), including *Gapdh*, which converts glyceraldehyde-3 phosphate to glycerate-1,3 biphosphate; *Tpi1*, which incorporates glycercerone-phosphate into glycolysis; *Pkm*, which converts phosphoenolpyruvate to pyruvate and *Hk1*, which phosphorylates glucose and retains it within the cytoplasm (Li et al., 2015) (Figure 5.17). There's also an increased expression of *Myc*, *Prkaa1*(catalytic subunit of AMPK) *Pgm2l1*, which allows the conversion of glucose-1, 6 biphosphate into glucose-1 phosphate for it to be incorporated into glycolysis (Semenza, 2013).



**Figure 5.17. Glycolysis map shows the genes upregulated in both S1 and ACE tumours. Genes upregulated in S1 tumours are marked in red. Genes upregulated in ACE tumours are represented in blue. 1,3BPG 1,3-bisphosphoglycerate, F1,6-BP fructose 1,6- biphosphate, F6P fructose 6-**

phosphate, DHAP dihydroxy acetone phosphay, GAPD glyceraldehyde-3-phosphate, GAPDH glyceraldehyde-3-phosphate dehydrogenase, TPI triosephosphate isomerase GLS glutaminase, GOT2 glutamate oxaloacetate transaminase, GS glutamine synthetase, G6P glucose-6-phosphate, GPI phosphoglucose isomerase, GPT glutamine pyruvate transaminase, GS glutamine synthetase, HK hexokinase, LDHA lactate dehydrogenase A, MCT monocarboxylate transporter, PDH pyruvate dehydrogenase, PEP phosphoenolpyruvate, PFK phosphofructokinase, PG phosphoglycerate, PGK phosphoglycerate kinase, PGAM phosphoglucomutase, PK pyruvate kinase, SLC solute carrier family, TCA tricarboxylic acid, PFKFB2 phosphofructo kinase/fructose 2,6 biphosphatase, R5P ribose 5 phosphate, CS citrate synthase, IDH isocitrate dehydrogenase,  $\alpha$ -KG DH alpha-ketoglutarate dehydrogenase, SS succinyl-CoA synthetase, SDH succinic dehydrogenase, FH fumarate dehydrogenase, MDH malate dehydrogenase, PDK pyruvate dehydrogenase kinase, CPT carnitine palmitoyltransferase. Created with BioRender.com

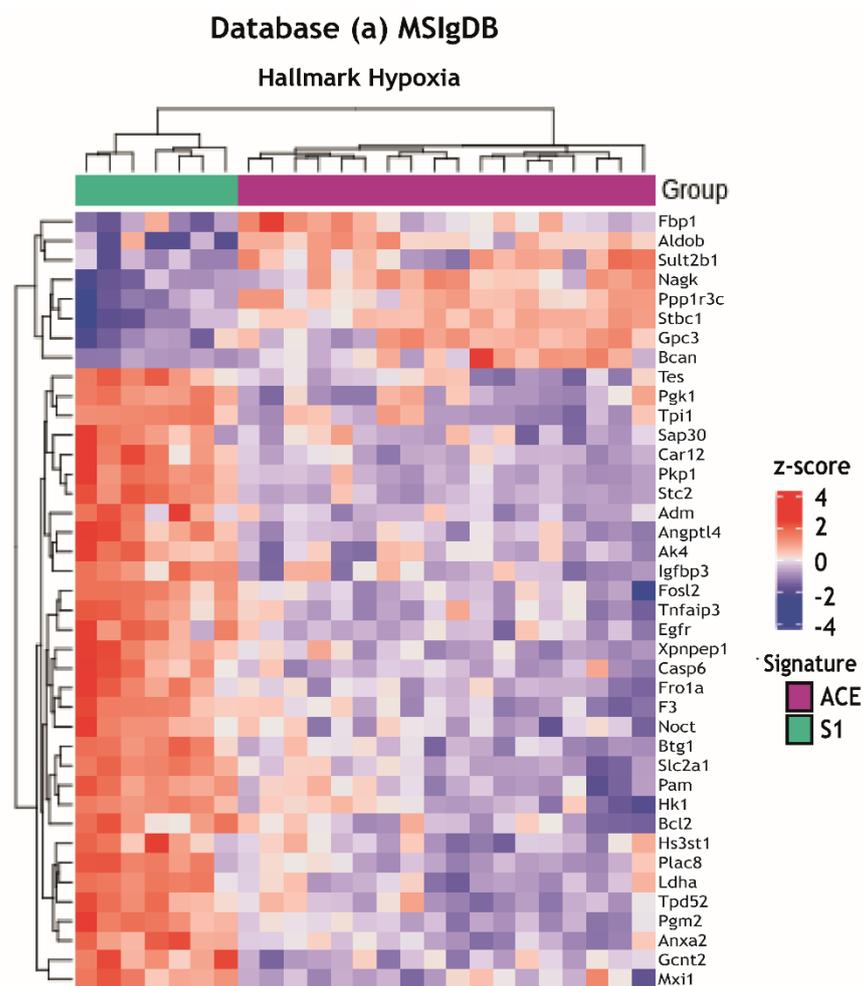
Additionally, several genes from the pyrimidine / purine biosynthesis are upregulated, such as *Nt5e*, which participates in the conversion of monophosphate ribose into guanosine, xanthosine, inosine and adenosine; *AK3* and *AK4*, which switch between AMP and ADP (Lane and Fan, 2015) (Figure 5.18).



**Figure 5.18. Nucleotide synthesis map with the upregulated genes in S1 tumours.** In red, upregulated genes in S1 samples. GMP guanosine monophosphate, glucose-6-phosphate, HK hexokinase, PPP pentose phosphate pathway, R5P ribose 5-phosphate, PRPS1/2 phosphoribosyl pyrophosphate synthetase 1, PRPP phosphoribosyl pyrophosphate, PPAT phosphoribosyl pyrophosphate amidotransferase, IMP inosine monophosphate, IMPDH inosine-5'-monophosphate dehydrogenase, AMP adenosine monophosphate, ADP adenosine diphosphate, AK3/4 Adenylate Kinase, Nt5e 5'-Nucleotidase Ecto, XMP xanthosine monophosphate, PHGDH phosphoglycerate dehydrogenase, PRA 5-phosphoribosylamine, PRPS pyrophosphate synthetase, PSPH phosphate ester hydrolysis, GSH glutathione. Created by BioRender.com.

ACE tumours showed an increase expression of *Pdk2* and *Pdk3*, which inhibits *Pdh* and block the conversion of pyruvate to Acetyl-CoA and promotes the expression of *Cpt1/2* which increases lipid metabolism (Bonnefont et al., 2004, Woolbright and Harris, 2021), *Akt1* and *Pi3kr1* genes, which promote not only glycolysis but lipid metabolism and cell growth (Miao et al., 2022) (Figure 5.16) .

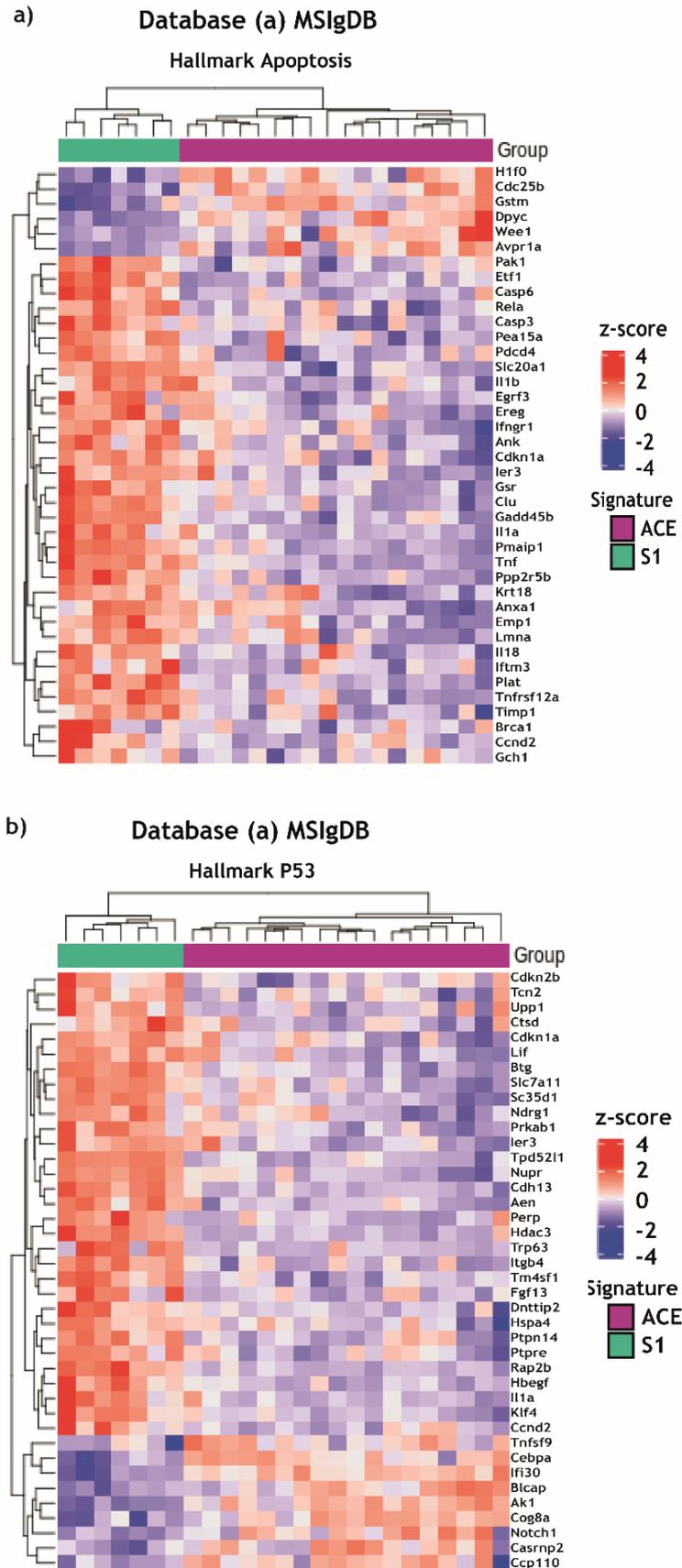
Database (a) showed again an increased expression of glycolytic genes in S1 tumours when compared to ACE tumours (Figure 5.16b), including the expression of *Pfk1*, which converts fructose-6 phosphate into fructose-1,6 biphosphate; *Pgm2*, which converts glucose-1 phosphate to glucose-6 phosphate; *Ldha*, which converts pyruvate into lactate; *ME2*, which converts pyruvate into malate; *Pgk1*, which converts glycerate-1,3 biphosphate into glycerate-3 phosphate; Enolase, which converts glycerate-2P into phosphoenolpyruvate; *Slc16a3*, lactate transporter (MCT4) *Pkm*, *Pgam1*, which converts glycerate-3 phosphate into glycerate-3 phosphate; and *Tpi1*, whereas ACE tumours showed increased expression of *Pygl* which promotes glycogen degradation (Figure 5.17) (Li et al., 2015).



**Figure 5.19. Gene expression patterns of S1, and ACE samples of significant genes in hypoxia gene set.** Heatmap shows expression pattern of significant genes of hypoxic genes in hypoxia gene sets in S1 and ACE tumours. Heatmap was generated using z-scores of gene expression values acquired from RNA sequencing of 3-6mm diameter tumours. Heatmap was generated in R. Samples clustering was performed using k-means. S1 n=7, ACE n=18.

Moreover, database (a) showed an enrichment of the hypoxia in S1 tumours, when compared against ACE group, suggesting that hypoxic pathways are activated in S1 tumours. Heatmap in figure 5.19 showed the increased expression of hypoxic genes in S1 subtype samples. For instance, glycolytic tumours are again upregulated in S1 tumours like *Pgk1*, *Tpi1*, *Slc2a1*, which encodes for glucose transporter GLUT1; *Hk1*, *Ldha* and *Pgm2*, (Li et al., 2015) together with other glycolysis promoting genes like *Maxi1*, which promoted *Myc* signalling, AMPK or EGFR which promotes proliferation in epithelial cells, as discussed above (Figure 5.20). This supports the idea that S1 tumours uptake glucose from the extracellular space and consume it through glycolysis. On the other hand, ACE tumours showed an upregulated expression of few glycolytic genes like *Fbp1*, which converts fructose-1,6 biphosphate to fructose-6 phosphate and reverses glycolysis; and aldolase, which converts to fructose-1,6 biphosphate to glyceraldehyd-3 phosphate (Li et al., 2015); together with glycogen regulator proteins, such as *Ppp1r3c* which promotes glycogen biosynthesis, suggesting again, the lower ACE tumours dependency in glycolysis and the use of intracellular storage of glucose as a source of energy (Figure 5.20) (Shen et al., 2010).





**Figure 5.21. Gene expression patterns of S1, and ACE samples of significant genes in apoptosis (a) and P53 (b) gene set.** Heatmap shows expression pattern of significant genes of apoptotic genes and genes related to P53 pathway in gene sets in S1 and ACE tumours. Heatmap was generated using z-scores of gene expression values acquired from RNA sequencing of 3-6mm

diameter tumours. Heatmap was generated in R. Samples clustering was performed using k-means. S1 n=7, ACE n=18.

In addition, database (a) showed an enrichment of apoptosis signalling pathways in S1 tumours when compared to ACE tumours. There are several genes related to apoptosis, cell cycle checkpoints and growth suppression in S1 tumours. Several examples are listed in table 5.2.

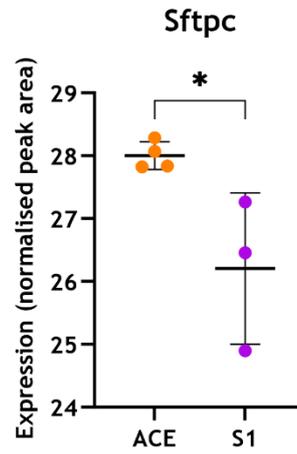
**Table 5.2. Summary of some of the statistically significant apoptotic genes upregulated in S1 tumours.**

Gene	Role
Atf1	mRNA surveillance pathway
Caspases 3 and 6	Apoptosis
Pdcd4	Growth suppression
Cdkn1a	Encodes for P21 / growth suppression
Pmaip1	Encodes for Noxa / apoptosis
Lmna	Apoptosis / loss of membrane integration
BRCA	Cell cycle arrest and DNA repair

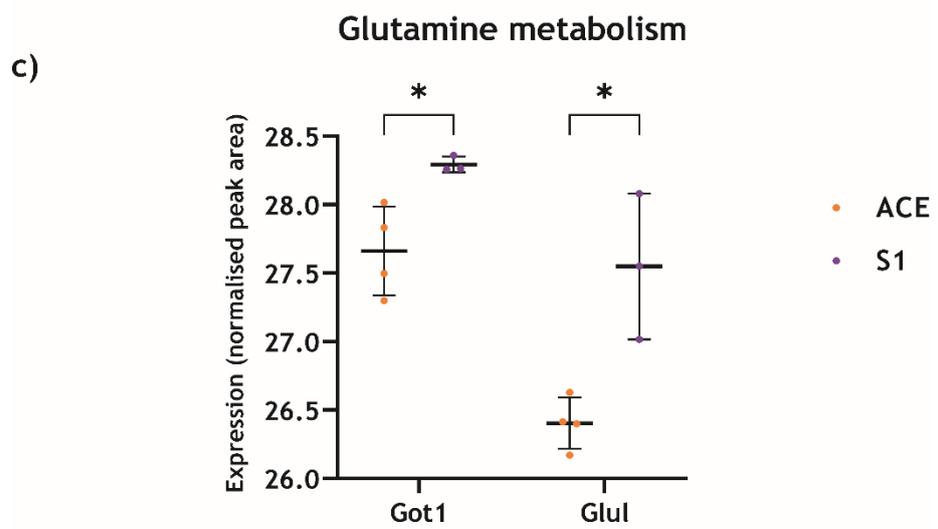
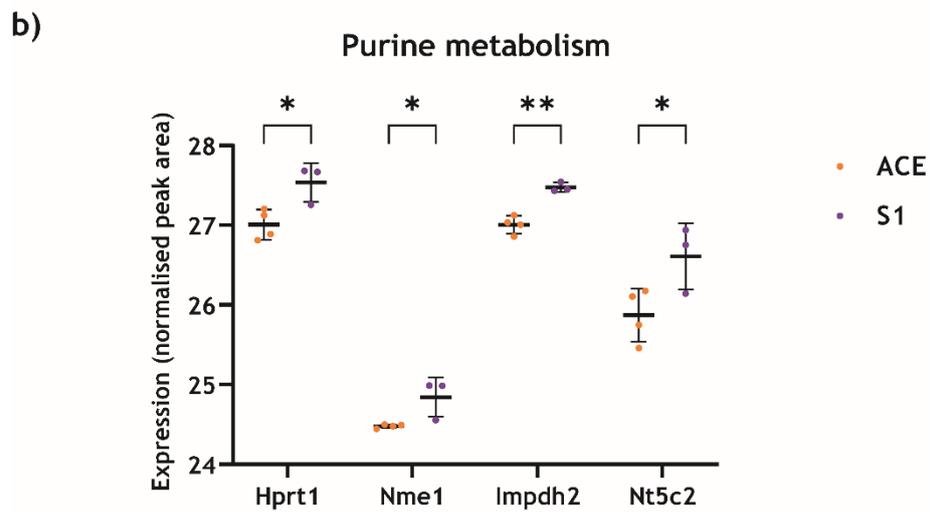
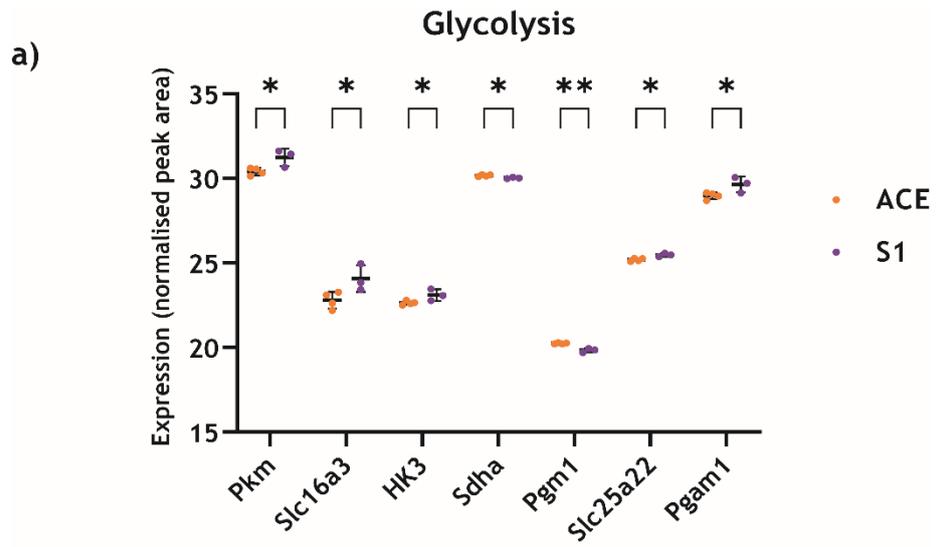
Finally, database (a) showed an increase expression of P53 pathway genes in S1 tumours, especially in the control of the cell cycle and autophagy. P53 pathway is downregulated in both S1 and ACE tumours due to P53 deletion but some of its targets may upregulate in the presence of certain stimuli, including *Cdkn2b*, *Ctsd*, *Btg1*, *Perp* and *Tp63* which regulate apoptosis and cell cycle inhibition.

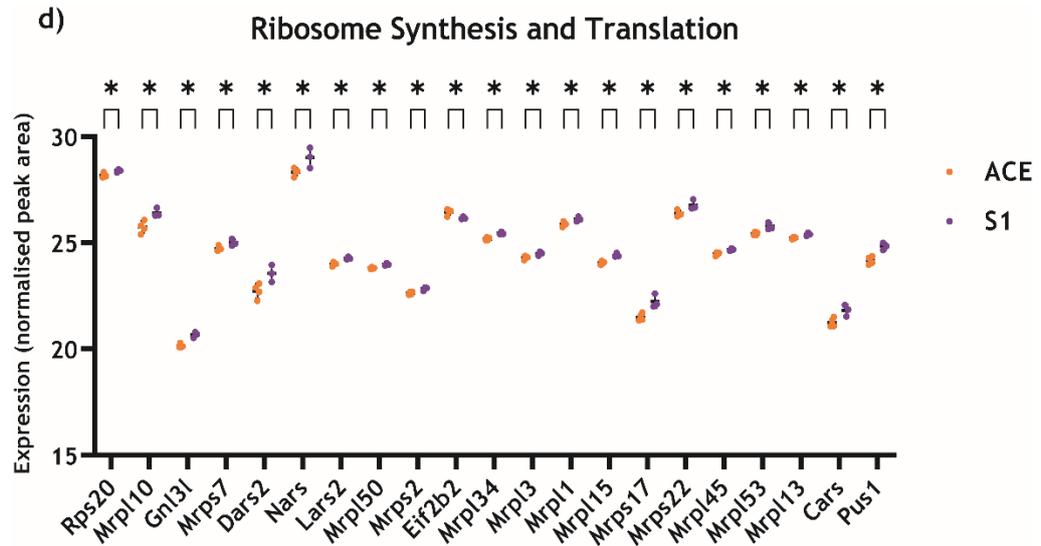
#### 5.3.5.2 Proteomic analysis of S1 and ACE samples

Four ACE samples and three S1 samples were randomly selected and sent to proteomic analysis. Preliminary analysis has focused on some of the the same genes that showed a significant difference in the transcriptomic analysis. These results revealed that *Sftpc* or SPC is upregulated in ACE samples.



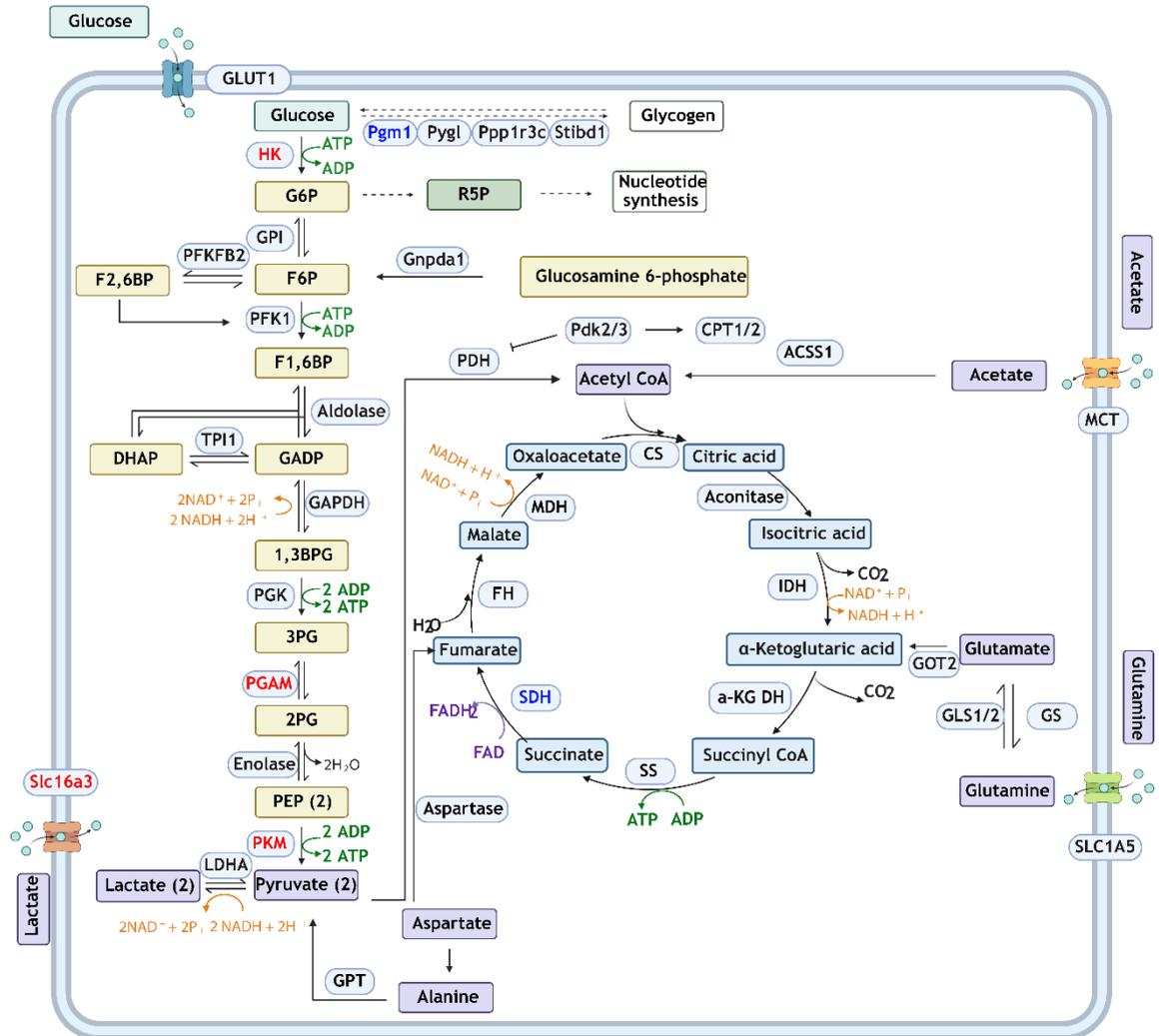
**Figure 5.22.** Sftpc expression is significantly higher in ACE than in S1 tumours. Each point represents the Sftpc value in every tumour. Unpaired two-tailed t-test performed in GraphPad Prism was used to determine a statistically significant difference between the two groups. S1 n=3, ACE n=4. \*p<0.05.





**Figure 5. 23. Proteomic data suggest a higher expression of glycolytic (a), nucleotide (b) and arginine (c) biosynthesis, one carbon metabolism (c) and translation (d) in S1 tumours.** (a) Represents glycolysis proteins and its expression in S1 and ACE tumours. (b) This graph represents the identified proteins in purine synthesis upregulated in S1 tumours. (c) Represents arginine biosynthesis and one carbon metabolism proteins and its expression in S1 and ACE tumours. (d) representation of upregulated expression of translational genes in S1 samples. Dot plots where each plot represents a tumour Bars represent the mean and the  $\pm$ SD. Statistical analysis using multiple unpaired t-test, with a 5% FDR in GraphPad Prism was used to determine a statistically significant difference across the two cohorts for each of the graphs presented. S1 n=3, ACE n=4. \*p<0.05 and \*\*p<0.01.

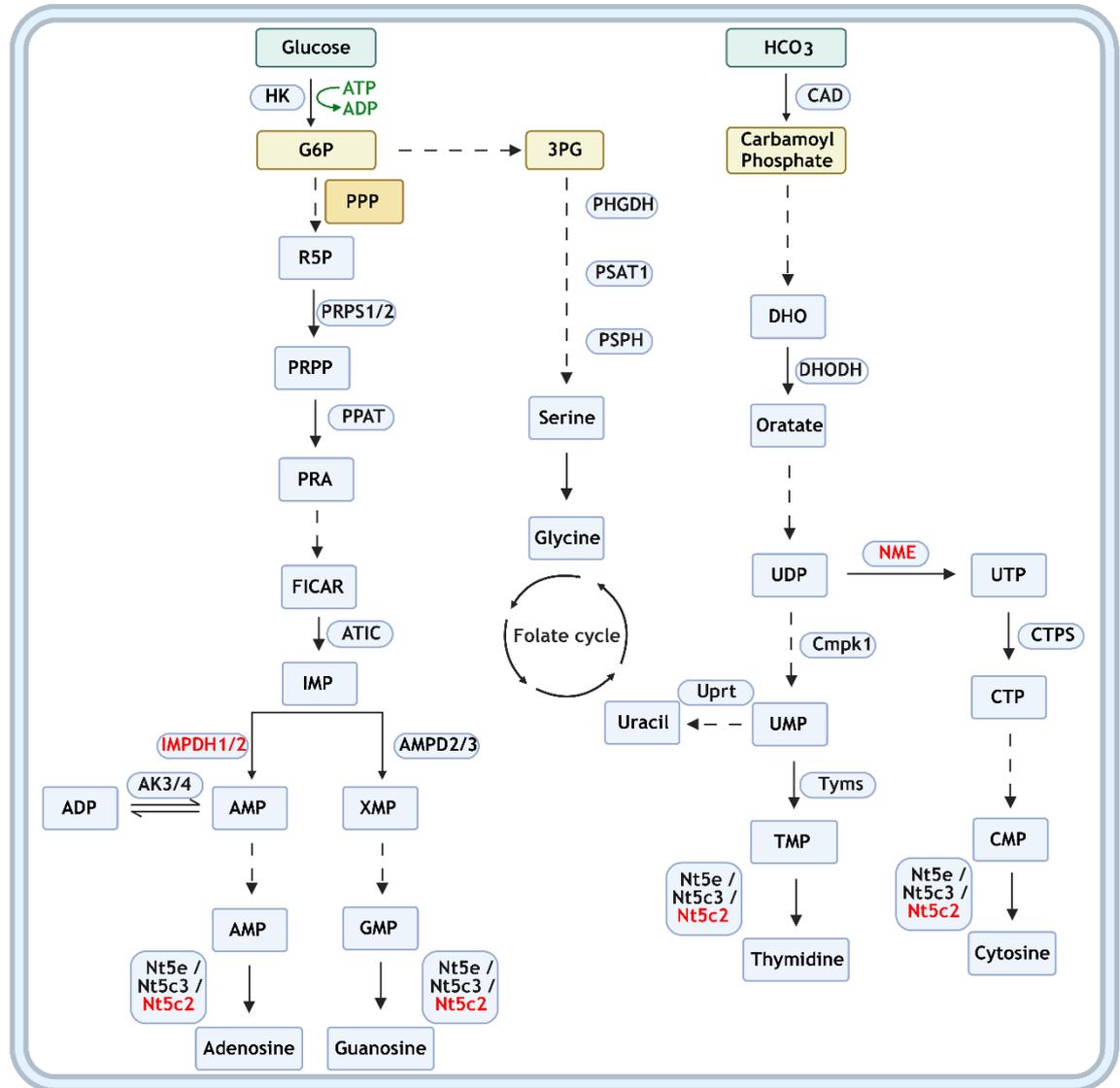
Moreover, several glycolytic enzymes were found upregulated in S1 tumours (Figure 5.23a), including PKM, which converts phosphoenolpyruvate into pyruvate; *Slc16a3*, which transports lactate out of the cell; hexokinase 3 which phosphorylates glucose into glucose 6-phosphate and retains it inside the cell; *Slc16a22* responsible for the transport of glutamate inside the mitochondria; and *Pgam1* a glycolytic enzyme which converts glycerate 3-phosphate into glycerate 2-phosphate. This data supports the idea that S1 tumours are more glycolytic than ACE tumours. On the other hand, *Sdha* and *Pgm1* were upregulated in ACE tumours. *Sdha* converts succinate into fumarate in the TCA cycle, and *Pgm1* which converts glucose 1-phosphate into glucose 6-phosphate and participates in the synthesis and degradation of glycogen (Figure 5.24) (Li et al., 2015, Shen et al., 2010).



**Figure 5.24. Schematic representation of glycolysis and TCA pathway.** In red, genes upregulated in S1 samples and in blue, genes upregulated in ACE samples. 1,3BPG 1,3-bisphosphoglycerate, F1,6-BP fructose 1,6- bisphosphate, F6P fructose 6-phosphate, DHAP dihydroxy acetone phosphaye, GAPD glyceraldehyde-3-phosphate, GAPDH glyceraldehyde-3-phosphate dehydrogenase, TPI triosephosphate isomerase GLS glutaminase, GOT2 glutamate oxaloacetate transaminase, GS glutamine synthetase, G6P glucose-6-phosphate, GPI phosphoglucose isomerase, GPT glutamine pyruvate transaminase, GS glutamine synthetase, HK hexokinase, LDHA lactate dehydrogenase A, MCT monocarboxylate transporter, PDH pyruvate dehydrogenase, PEP phosphoenolpyruvate, PFK phosphofructokinase, PG phosphoglycerate, PGK phosphoglycerate kinase, PGAM phosphoglucomutase, PK pyruvate kinase, SLC solute carrier family, TCA tricarboxylic acid, PFKFB2 phosphofructo kinase/fructose 2,6 biphosphatase, R5P ribose 5 phosphate, CS citrate synthase, IDH isocitrate dehydrogenase, α-KG DH alpha-ketoglutarate dehydrogenase, SS succinyl-CoA synthetase, SDH succinic dehydrogenase, FH fumarate dehydrogenase, MDH malate dehydrogenase, PDK pyruvate dehydrogenase kinase, CPT carnitine palmitoyltransferase. Created with BioRender.com

Moreover, supporting transcriptomic results, several nucleotide synthesis genes were upregulated in this analysis, including *Hprt1* that converts purines

monophosphate into nucleosides; *Nme1* which converts nucleosides triphosphates to nucleosides biphosphates; *Impdh2* which converts IMP to AMP, a precursor of adenosine, and finally *Nt5c2* which synthesizes nucleotides (Figure 5.25) (Lane and Fan, 2015).



**Figure 5.25. Schematic representation of nucleotides synthesis pathway.** In red upregulated proteins found in S1 tumours. GMP guanosine monophosphate, glucose-6-phosphate, HK hexokinase, PPP pentose phosphate pathway, R5P ribose 5-phosphate, PRPS1/2 phosphoribosyl pyrophosphate synthetase 1, PRPP phosphoribosyl pyrophosphate, PPAT phosphoribosyl pyrophosphate amidotransferase, IMP inosine monophosphate, IMPDH inosine-5'-monophosphate dehydrogenase, AMP adenosine monophosphate, ADP adenosine biphosphate, AK3/4 Adenylate Kinase, Nt5e 5'-Nucleotidase Ecto, XMP xanthosine monophosphate, PHGDH phosphoglycerate dehydrogenase, PRA 5-phosphoribosylamine, PRPS pyrophosphate synthetase, PSPH phosphate ester hydrolysis, GSH glutathione. Created by BioRender.com.

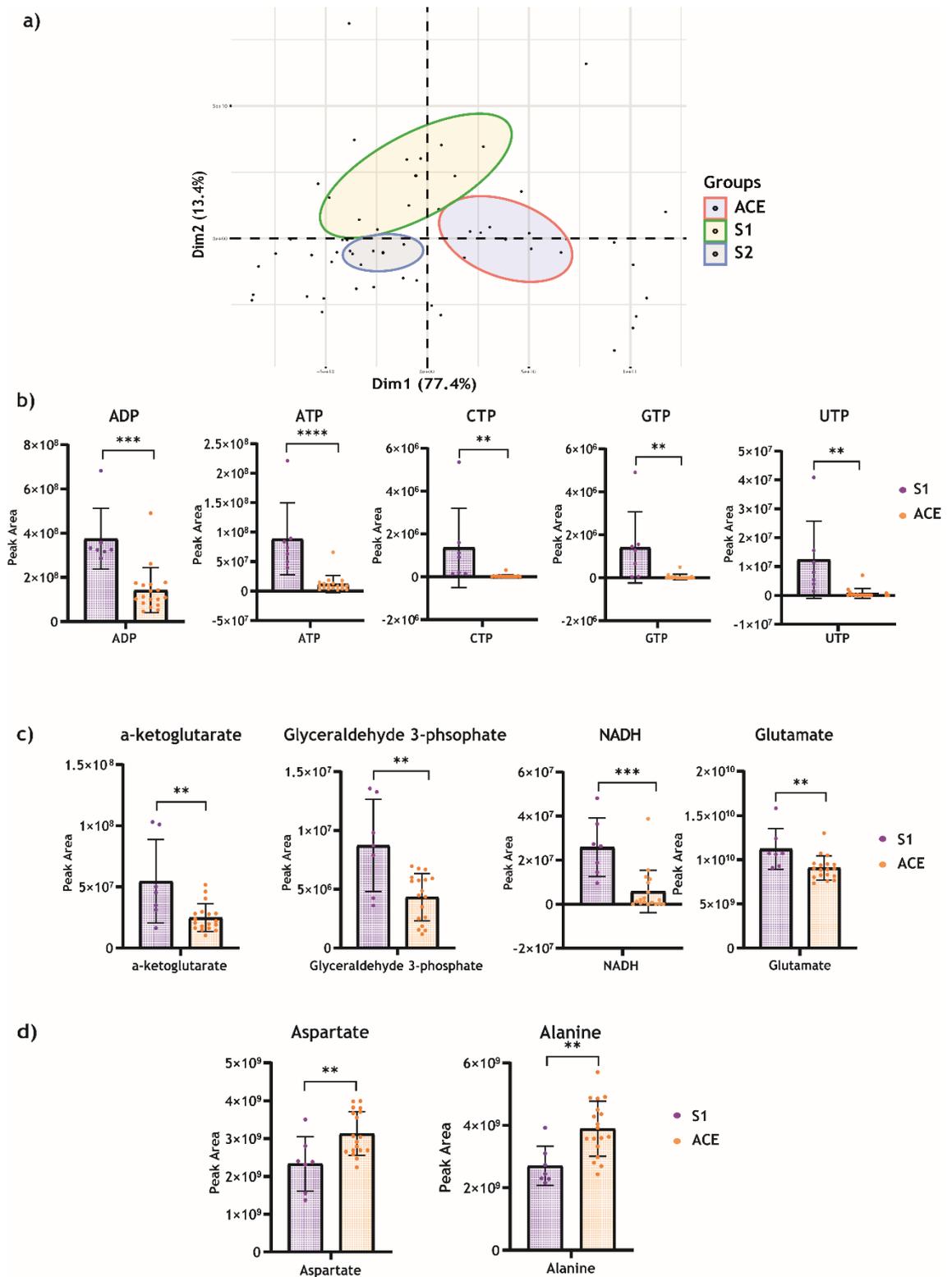
Additionally, *Got1* and *Glul* related to glutamine metabolism were increased in S1 tumours that will increase the intracellular pull of glutamine and glutamate. This can be used as an alternative source of energy to fill the TCA cycle (Yoo et al., 2020). *Phgdh* is part of the glycine, serine and threonine biosynthesis pathway (Figure 5.24) (Zhao et al., 2021a).

Finally, S1 tumours have an increase expression for translational proteins. *Mrpl*, *Mrps* and *Rps20* genes are related to ribosome synthesis and integrity, while *Gnl3l* and *Eif2b2* are part of the initiating gene machinery of gene translation. *Nars*, *Dars2*, *Lars2*, *Cars* and *Pus1* are aminoacyl-tRNA synthase (Figure 5.23) (Dever et al., 2018).

### 5.3.5.3 Metabolic characterisation of S1 and ACE

To further confirm transcriptomic data, a targeted metabolomic analysis was run in these groups. Unpaired parametric multiple t tests, which performs several t-test at once, were run to understand the statistical differences of these metabolites between S1 and ACE samples with a false discovery rate of a 5%. Targeted data was analysed in Metaboanalyst and a PCA was performed (Figure 5.26a). In this analysis ACE, S2 and S1 subgroup separate well from one another, indicating some differences in the metabolites level of these subgroups.

S1 vs ACE there are a series of metabolites which are overexpressed in S1 tumours. For instance, ATP, ADP, NADH, UTP,  $\alpha$ -ketoglutarate, glyceraldehyde 3-phosphate, GTP, CTP and glutamate. On the other hand, alanine and aspartate are upregulated in ACE tumours (Figure 5.26b, c and d).

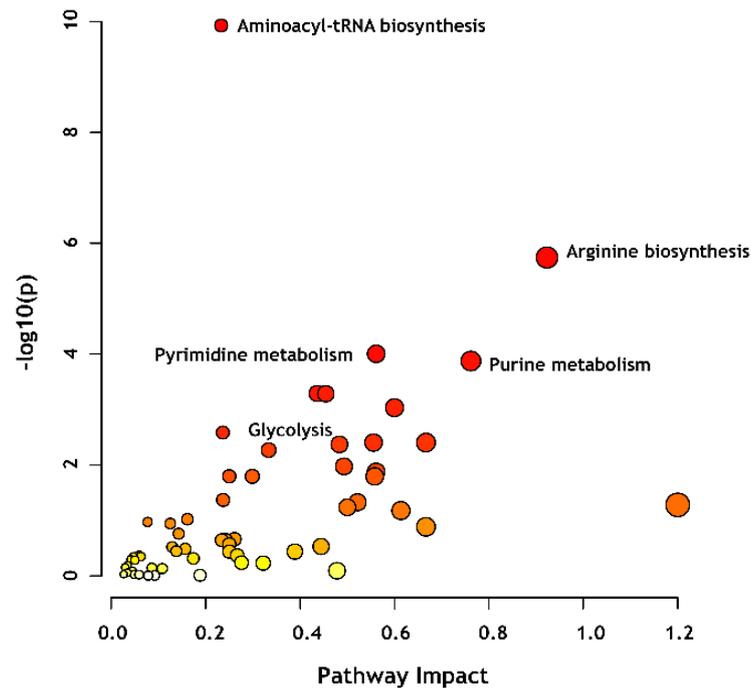


**Figure 5.26. S1 tumours are metabolic different from ACE tumours.** (a) PCA was performed and plotted with factextra R package after differential metabolite expression. The sample groups were set as the variable of interest and a confidence threshold was applied to each group. The elliptical around each group representing the mean and 95% confidence interval for each sample group. The mean for each group lies in the centre of each elliptical, with the outline the 95% confidence interval. Each data point represents a single tumour. (b) Bar plots representing several metabolites related to nucleotide synthesis in S1 tumours. (c) Bar plots with

glycolytic metabolites upregulated in S1 tumours. (d) and (e) Bar plots representing several intermediate metabolites of the urea cycle, TCA among others. Dot plots where each plot represents a tumour Bars represent the mean and the  $\pm$ SD. Statistical analysis using multiple unpaired t-test, with a 5% FDR in GraphPad Prism was used to determine a statistically significant difference across the two cohorts for each of the graphs presented. S1 n=7, ACE n=18. \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.0001.

#### 5.3.5.4 Joint pathway analysis

Following the results obtained in section 5.3.7, we ran a joint pathway analysis with transcriptomics and metabolomic data. Interestingly, pyrimidine and purine metabolism were upregulated in S1 tumours when compared against ACE tumours, together with aminoacyl-tRNA biosynthesis, glycolysis and arginine biosynthesis. Moreover, D-glutamine and D-glutamate metabolism, glycolysis, glycine, serine and threonine metabolism, TCA cycle are upregulated in S1 tumours.



**Figure 5.27. Joint pathway analysis of transcriptomics and metabolomics revealed differences in the glycolytic metabolism and nucleotide synthesis.** Metabolic pathway analysis was run with Metaboanalyst 5.0 with Joint Pathway analysis feature. Hypergeometric test was chosen as enrichment analysis, the topology measure was Degree-Centrality measures the number of links that connect to a node. For integration methods, tight integration by combining queries was chosen in which genes and metabolites are pooled into a single query and used to perform enrichment analysis within their "pooled universe".

## 5.4 Discussion

### 5.4.1 Transcriptomics results tumour vs normal tissue

CMV virus targets any cell type in the lung (Schlabach et al., 2010) thus, it was used as positive control for infection. Therefore, CMV-induced KP mice are expected to have both SPC and CC10 tumours or even tumours from other potential lung cells-of-origin. Consequently, this CMV was used ensure that CC10 and SPC virus vectors are behaving as expecting.

In the PCA plot, CC10 and SPC tumours lie separately as they hit different cells-of-origin and therefore, it is expected for them to have transcriptional differences. However, as CMV tumours are comprised of both CC10 and SPC tumours, it was expected to see CMV samples are overlapping both CC10 and SPC samples, as seen in figure 5.1. This corroborates the expected behaviour of these adenovirus vectors.

This PCA plot also shows the normal tissue samples for KP mice induced with CC10, SPC and CMV. As these samples are all normal tissue sections, these were expected to be more overlapping than the tumour samples, as was seen. Finally, the separation between tumour and normal tissue samples confirmed the expected transcriptomic changes during the tumorigenesis.

*KRas* and *P53* expression was analysed and plotted to check the efficiency of the intranasal induction. As explained in section 4.1, once the virus has targeted the either CC10 or SPC cells, *Cre* recombinase is expressed. This CRE recombinase cut the *loxP* sequences and introduces the *KRas*<sup>G12D</sup> mutation and deletes *P53*. These effects are represented in figure 5.2a and 5.2b and they confirmed the genetic alterations driven by CRE recombinase.

### 5.4.2 Differences between CC10 and SPC tumours

#### 5.4.2.1 Transcriptomic analysis CC10 vs SPC

At this point, the transcriptomic characterisation was performed comparing CC10 versus SPC tumours. At first instance, it highlighted several immune-related genes and immune pathways seem to be enriched in SPC tumours, which

may indicate a highly active immune system in the tumour. It is unclear though if this immune activity is pro-tumorigenic or anti-proliferative. *Ltf* (Zhao et al., 2021b) and *Vtcn1* (Altan et al., 2018) have been described to help tumours elude the immune system, while *Famd3* (He et al., 2018) and *Gzmd* (Cao et al., 2018) are cytotoxic. Moreover, autoimmune thyroid gene set include genes pro and anti-tumorigenic (Walker and Sansom, 2011, Hazini et al., 2021). Moreover, SPC tumours had upregulated cell adhesion molecules that have been associated with the immune response in the skin epithelium, aggregating more significance to the immune response differences described above (Tamoutounour et al., 2019). This gene set contains again genes related to the immune and inflammatory response including cell migration and transendothelial cell migration (Woolbright and Harris, 2021).

Overall, these would indicate an active immune response in SPC tumours and a higher capacity for cell migration. More experiments are needed to understand if this cell migration happens from the blood vessels to the tumour, indicating an accumulation of immune cells in the tumour or if the tumour cells use this mechanism to migrate and invade outside the tumour, which will also correlate with broken elastin fibres in figure 4.9d. This high immune system activity may explain the latency found in SPC tumours, as tumour may need to overcome the immune checkpoints before adapting itself and evolve evasive strategies and start proliferating.

Transcriptomics is a starting point and a useful tool to generate hypothesis. In this case, it clearly indicates that there is an involvement of the immune system in tumour growth and microenvironment between SPC and CC10 tumours, but more experiments are needed to fully understand the immune system role. It is also useful to generate gene signatures to identify tumours arising from CC10 or SPC cells.

After removing the outliers, neuroactive ligand-receptor interaction gene set becomes significantly enriched in SPC genes. The neuroactive ligand-receptor interaction signalling pathway is a collection of receptors and ligands on the plasma membrane that are associated with intracellular and extracellular signalling pathways. This gene set includes genes related to metabolite channels like glutamine, glycine, alanine, aspartate among different hormones and

peptides which are involved in tumour progression. It has been demonstrated that this pathway is associated with progression of bladder, prostate cancer and renal cell carcinoma, although its mechanisms are not very well understood (Zhaohui He, 2018).

Among the up regulated genes in CC10 it can be found *Galnt15*, a glycosyltransferase, which are usually dysregulated in cancer. They can constitutively activate several growth factor receptors such as *Egfr* or *Igfr* by blocking their internalisation, which activate cell proliferation and glucose consumption via glycolysis (Rajoria et al., 2023). However, the rest of the highlighted genes in CC10 tumour did not show an obvious relation among them. To introduce more sense in this data, a GSEA was run.

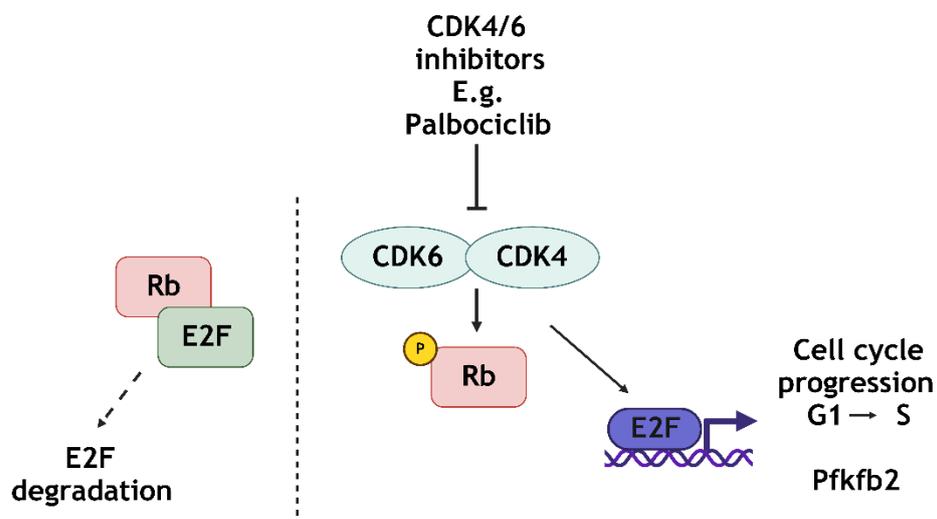
First of all, differences in glycolysis between SPC and CC10 tumours were corroborated as both databases indicated an upregulation of this pathway in CC10 tumours. This is supported by the statistically significant increase of *Igfr* expression in CC10 tumours as it drives metabolic changes, including glucose metabolism (Rajoria et al., 2023). *Igfr* expression in CC10 positive or SPC positive normal cells has not been investigated. This is also supported by the upregulation of *Myc* and *EF2* targets in CC10. It has been shown that they both upregulate glycolytic genes. *Myc* overexpression in CC10 tumours may explain the upregulation of genes related to translation and protein biosynthesis genes, as *Myc* promotes proliferation and gene expression (Chen et al., 2018). Glycolysis has also been linked with *KRas* constitutive activation (Kerr et al., 2016) as it activates mTORC1 pathway which, increases cells glycolytic activity. *KRas* signalling is upregulated in both SPC and CC10 tumours due to G12D mutation introduced. However, CC10 tumours have a significant increased upregulation compared to SPC tumours which could be explained by a distinctive activation of *KRas* in the cell-of-origin. Finally, the upregulation of hypoxic targets promotes glycolytic changes in CC10 tumours. (Lee et al., 2020b). Hypoxia in CC10 tumours or hypoxic regions in these tumours can be confirmed with IHC staining for hypoxia markers like HIF-1 $\alpha$ .

Estrogen signalling pathway, is upregulated in CC10 tumours when compared against SPC tumours. Estrogen receptor can activate *Cyclin D* and induce cell proliferation (Xu et al., 2017). In addition, CC10 tumours seems to have an

upregulated expression of G2M checkpoint. This may seem contradictory as G2M checkpoint gene set prevents cells from entering in the mitosis and maintains genomic stability (Schönthal, 2004) while all pathways described above not only induce glycolysis but also cell proliferation.

Overall, CC10 tumours seems to be more glycolytic and more proliferative than SPC tumours, given the proliferating pathways significantly upregulated in CC10 tumours. Although IHC staining for Ki67, mitotic markers or BrdU are needed to confirm this proliferative profile of CC10 tumours. This proliferative machinery may explain why CC10 tumour have a shorter latency than SPC tumours. Moreover, being CC10 tumours might be more proliferative, it is likely that they grow up to be more hypoxic tumours leading to *HIF-1a* activation.

These differences may reveal certain druggable targets. For example, cyclin-dependent kinases 4 and 6 (CDK4/6) play a crucial role in cell proliferation. They phosphorylate members of the RB family, inactivating these tumour suppressor proteins. This phosphorylation leads to the release of E2F transcription factor which is normally inactivated by retinoblastoma (Wagner and Gil, 2020). E2F targets expression is significantly upregulated in CC10 tumours suggesting an activation of CDK4/6. There have been several CDK4/6 inhibitors, including Palbociclib (David W. Fry, 2004), which has been extensively used in breast cancer. This data indicates that this drug might also be effective in CC10 tumours.



**Figure 5.28. Mechanism of action of E2F transcription factor in cell proliferation.** In G1 phase CDK4/6 activate E2F by phosphorylating retinoblastoma. This allows E2F to move to the nucleus and progress from G1 to S phase. These activation can be inhibited by CDK4/6. Rb retinoblastoma, CDK cyclin dependent kinase. Created with BioRender.com.

#### 5.4.2.2 Proteomic analysis CC10 vs SPC

Preliminary proteomic results show a significant increase of several glycolytic genes in CC10 tumours. This finding supports the transcriptomic results, which show a significant increase of glycolysis data sets in CC10 tumours.

An interesting result is the fact that CC10 tumours have an upregulation of *Slc2a1* which encodes for GLUT1 glucose transporter. On the other hand, SPC tumours have a significant increase in the expression of *Pgm1* related to glycogen metabolism, which may indicate a different source of glucose for each one of these tumours.

The expression of *Slc16a1*, which may indicate the entrance of pyruvate into the mitochondria to continue the TCA cycle or the transport of lactate to the extracellular space. Moreover, *Pfkfb2* is increasing PFK1 activity. Finally, *Gnpda1* expression in CC10 tumours proves that there might be other sources of glycolytic intermediates rather than glucose.

Overall, this may indicate a different utilisation of glucose or glucose origin. CC10 tumours seem to be taking up glucose from the extracellular space and, based on the upregulation of glycolytic genes in both transcriptomics and proteomic analysis, it is used for glycolytic purposes, whereas SPC tumours either use glycogen as a source of glucose or store the glucose they uptake as glycogen.

Moreover, several genes from the purine and pyrimidine biosynthesis are overexpressed in CC10 samples, suggesting that glucose can be derived to the PPP and be used for nucleotide synthesis which will support proliferation. This supports the results seen above, where it showed that CC10 tumours have an increased expression of *KRas*, estrogen or E2F or *Myc* target gene sets, which are all related to cell proliferation. Cell growth is dependent on de novo nucleotide synthesis to complete the DNA synthesis.

### 5.4.2.3 Metabolomic analysis CC10 vs SPC

In this analysis, lactate is significantly increased in in CC10 tumours, which again supports the upregulation of glycolytic gene sets in CC10 tumours and the upregulation of hypoxia and MYC targets, as pyruvate conversion to lactate is promoted by these pathways (Dong et al., 2020). Hypoxic cells express MCT4 lactate transport which takes out lactate out of the cell when it accumulates (Semenza, 2013), suggesting an increased lactate production and the transport of it outside the cell.

No more glycolytic metabolites have been detected with enough statistical power between CC10 and SPC. This can have different explanations. The most likely explanation is the fact that glycolysis chemical reactions occur quickly enough to accumulate the final product of the pathway. Another possible explanation, as explained above, glycolytic intermediates can be derived to serine synthesis to reduce oxidative stress in hypoxic conditions (Semenza, 2013), which would explain the lack of differences in glycolytic metabolites. However, most of them are not statistical significance. For instance, serine is not statistically significant ( $p$ -value  $> 0.05$ ) in CC10 when compared to SPC tumours, thus it is not likely. Secondly, proteomics showed a possible diversion of glucose to the nucleotide synthesis. Again, these metabolites are not statistically significant, but their protein expression was upregulated in CC10 tumours. (Figure 5.8).

### 5.4.3 Description of S1, S2 and ACE subgroups

A few outliers samples were first identified among the tumour samples. After some investigation, there were no experimental or technical explanation that would explain why these samples were not fitting with the rest. They belong to different mice (both males and females), they came from both CMV and CC10 virus vectors, mice were similar age at the time of the induction, which was performed on different days as well as the tumour dissection and collection. Therefore, a human error during the sample collection or during the induction process or a mouse/sex/age effect was discarded. When comparing our data to Juntilla's paper (Daemen et al., 2021), we identified these samples as a different transcriptional subtype. In order to have a deeper understanding of the

cell-of-origin effect of these transcriptional subtypes, a heatmap and clusters were created by simple K-means on row-scaled normalised gene expression of the genes on those gene sets. The results suggested that there are three clusters: S1 and S2, previously described in Juntilla's paper, and an ACE subgroup which has been identified for the first time.

The tumour cell-of-origin defines at least two of these subgroups and they present imaging differences, which may indicate metabolic differences: S1 subgroup is dominated by CC10 tumours, and it presents with high expression of [<sup>18</sup>F]FDG signature genes, suggesting S1 tumours might be [<sup>18</sup>F]FDG avid. On the other hand, ACE subgroups is dominated by SPC tumours. Moreover, this subgroup is enriched in [<sup>11</sup>C]acetate signature genes, which indicates the preference of these tumours for acetate compared to glucose. The enrichment of SPC tumours in ACE transcriptomic subgroup and the CC10 sample enrichment in S1 subgroup is statistically significant, indicating an association between the cell-of-origin and these two transcriptional subtypes. Therefore, the cell-of-origin variable allows us to refine the transcriptional classification of NSCLC tumours by adding a third group in this classification and remove heterogeneity.

S2 subgroups is not dominated by a particular tumour type, as it is a combination of CMV, SPC and CC10 tumours. There are several explanations why not all SPC samples lie in ACE group or all CC10 tumours are classified as S1 samples. A likely explanation might be the location of the tumour. It has been described that CC10 expressing cells are distributed along the murine airways and that there are several club cells subtypes (Barkauskas et al., 2017). SPC expressing cells are mainly located in the alveolar space, although BASCs are present in the BADJ (Kim et al., 2005). These multiple location and cell subtypes might influence the transcriptional subgroup they belong to.

Most importantly, it is important to consider that Ad5-mSPC-Cre expresses CRE recombinase in AT2 cells and BASCs while Ad5-CC10-Cre targets club cells and BASCs. Thus, there is a slightly overlap with these two vectors as BASCs express both SPC and CC10 (Lee et al., 2014). It might be the case where S2 subgroup consists of tumours arising from BASCs induced from both Ad5-mSPC-Cre and Ad5-CC10-Cre vectors. If this is true, the presence of CC10 and SPC tumours in

the S2 subgroup is due to an artefact of the model and the way these viruses work. Thus, it was decided to only further characterise S1 and ACE groups.

Moreover, Ad5-CMV-Cre vector is able to induce Cre recombinase in every cell type in the lung (Schlabach et al., 2010). It might be the case that CMV-S1 tumours are arising from CC10 club cells targeted by Ad5-CMV-Cre, CMV-ACE tumours are arising from AT2 cells infected with Ad5-CMV-Cre and CMV-S2 tumours arise from BASCs cells targeted from Ad5-CMV-Cre.

#### 5.4.4 PET imaging

S1 and S2 tumours have a higher glucose uptake than ACE tumours. At the same time, ACE tumours expressed genes from [<sup>11</sup>C]acetate gene set, indicating a preference for acetate uptake over glucose. [<sup>18</sup>F]FDG is a tracer of the Warburg effect (S1 tumours) (Telo et al., 2020). [<sup>11</sup>C]acetate gene signature is expressed in tumours which preferred acetate uptake over glucose (ACE subgroup), [<sup>11</sup>C]acetate being a tracer for both oxidative metabolism and fatty acid synthesis (Lewis et al., 2014).

PET imaging data, on 8 different CC10 tumours further validates this transcriptomic classification, and it showed that from eight imaged tumours, six of them (75%) have a higher uptake of [<sup>18</sup>F]FDG. This supports data that shows that CC10 tumours are taking up glucose. This may indicate a higher glycolytic activity although it is not measured with [<sup>18</sup>F]FDG, as it does not further metabolise. Only 11% of the CC10 tumours fell into the ACE subgroup in the transcriptomic data, and only two out of 8 imaged CC10 tumours seemed to prefer acetate uptake, corroborating CC10 tumour classification.

Transcriptomic data showed that 46% of SPC tumours are classified in the ACE subgroup while 54% of them fell into the S2 subgroup. PET data validates this classification as from eight tumours, four of them seemed to have preference for acetate uptake while the other four showed a higher uptake of glucose, corroborating SPC tumour classification. It is also known that acetate is a precursor of acetyl-CoA, and this metabolite is not only part of the TCA cycle but also a precursor of lipid biosynthesis (Sitaula and Burris, 2016).

PET imaged tumours were dissected and sent for IHC staining. Staining for GLUT1, HK, FASN or other genes related to fatty acid metabolism will help us to fully understand the metabolic profiles of these tumours.

### 5.4.5 Differences between S1 and ACE tumours

#### 5.4.5.1 Transcriptomic characterisation of S1 and ACE tumours

The GSE analysis between S1 and ACE tumour subgroups showed the most interesting results.

Firstly, there's a significant enrichment of glycolysis pathway in S1 tumours in two of the chosen databases. This correlates with the K-mean clusters seen in figure 5.13, as they showed an increased uptake of [<sup>18</sup>F]FDG (equivalent to glucose) in S1 tumours, and the PET imaging [<sup>18</sup>F]FDG uptake seen in CC10 tumours. Among these two databases there are several glycolysis genes significantly upregulated in S1 tumours. At the same time, there's several genes from the second half of the glycolysis and the lactate metabolism upregulated again in S1 tumour. Finally, genes that promote glycolytic metabolic reprogramming to facilitate cell proliferation and survival are overexpressed in S1 tumours such as *Myc*, *Homer*, which promotes PI3K pathway, resulting in glucose uptake genes transcription; or *AMPK* which promotes anaerobic metabolism (Li et al., 2015).

Additionally, genes related to the purine and pyrimidine biosynthesis have been identified among the S1 upregulated genes in these two gene sets, suggesting a possible deviation of glucose towards the synthesis of nucleotides to guarantee proliferation resources.

On the other hand, some of the genes on this gene set were significantly upregulated in ACE tumours. Few of them are part of the glycolysis pathway, which may suggest a minor utilisation of glucose for glycolysis. However, ACE tumours have a higher expression of *Pdk2* and 3 in ACE tumours. These last two enzymes inhibit PDH, blocking the entrance of pyruvate in the TCA cycle as it cannot be converted into Acetyl-CoA (Woolbright and Harris, 2021). Moreover, *Pdk2* and 3 increases the activity of *CPT1/2*, which drives fatty acid  $\beta$ -oxidation (Bonfont et al., 2004), which may suggest an alternative way to produce

energy. Interestingly, *Pygl* gene, which participates in glycogen degradation (Chandel, 2021) is overexpressed in ACE tumours, which may indicate that the glucose source in these tumours is not extracellular but intracellular, although glycogen tracer experiments are needed to prove this.

In summary, these transcriptomic results suggest the uptake of glucose from the extracellular environment, the utilisation of this glucose in glycolysis in S1 tumours, and the conversion of it to lactate which is subsequently transported outside the cell. Whereas ACE tumours may use glycolysis as a source of energy but given the lower glycolytic enzymes expression, these tumours might be less dependent on it. Given the expression of Pdk2 and 3, it is reasonable to think that ACE tumours also use  $\beta$ -oxidation as a source of energy. Moreover, glucose source for ACE tumours might differ from S1 tumours as the expression of *Pygl* gene indicates glycogen mobilisation. This correlates with the CC10 versus SPC analysis, which show that CC10 tumours were more glycolytic, as S1 subgroups is dominated by CC10 tumours.

Hypoxia targets are enriched in S1 tumours. As hypoxia promoted anaerobic metabolism, glycolytic genes are among those targets upregulated in. Moreover, there are apoptotic genes and survival genes overexpressed in S1 tumours. This may seem contradictory, but apoptosis is a cell strategy to deal with acute hypoxia (Greijer and van der Wall, 2004). This indicates a possible balance between cells which undergo apoptosis and cells which proliferate in S1 tumours.

ACE tumours also display an upregulation of few of these hypoxia targets, including genes from glycogen metabolism which were significantly upregulated in ACE tumours.

Overall, this demonstrates that ACE tumours follow the same pattern describe above. Few glycolytic genes are overexpressed but the expression of *Ppp1r3c* and *Stbd1* indicates the utilisation of glycogen as a source of glucose (Zois et al., 2014), instead of uptake it from the circulation. S1 tumours are hypoxic, or they have hypoxic pathways activated. This data also shows that S1 tumours uptake glucose from the extracellular environment given the expression of *Slc2a1* and they use it for anaerobic metabolism, as it has been shown with PET imaging data, where CC10 tumours are more [ $^{18}\text{F}$ ]FDG avid. Again, this corroborates the

data suggesting that CC10 tumours are more glycolytic than SPC tumours as S1 subtype is dominated by CC10 samples. Moreover, S1 tumours have an upregulated expression of genes related with both survival and apoptosis, both pathways related to the hypoxic response, suggesting a balance between cell death and proliferation within the tumour. This correlates with the finding of proliferative (Myc, Kras E2F) and non-proliferative (G2-M checkpoint) pathways upregulated in CC10 tumours. On first instance, this seems contradictory to the growth dynamic analysis performed in chapter four. However, those measurements were made on a tumour volume basis, which is not sensitive enough to estimate the cell death rate. Nevertheless, cell proliferation must be superior to cell death to allow tumour growth.

To further confirm this, apoptosis and P53 gene sets were enriched in S1 tumours. P53 has been deleted in this mouse model and is downregulated in both tumour types, but its signalling pathway is enriched in S1 tumours. In these pathways tumour, apoptotic and autophagy genes are overexpressed in S1 tumours. As said above, this can be explained by the fact the S1 tumours can have a higher apoptotic rate than ACE tumours, explained by the upregulation of apoptotic genes.

The upregulation of these pathway supports the idea expressed above that as CC10 tumours may evade immune system way quicker than SPC tumours, they seem to be more proliferative and glycolytic and producing more lactate, they have a lower latency, and they might be able to expand quicker than SPC lesions. This may induce hypoxic regions within the tumour, explaining perhaps the upregulation of hypoxic targets in these tumours. The expression of apoptotic genes and *p53* target gene sets in S1 tumours include autophagy apoptotic genes. These are two cell strategies to deal with moderate and acute hypoxia. Therefore, these differences should be visible in an H&E staining where necrotic and hypoxic areas, expressing apoptotic and cell suppression genes should be differentiated from oxygen rich areas where high proliferating pathways such as MYC or PI3K can be upregulated. This would explain the contradiction of having apoptotic and cell suppression genes within the same tumour type as proliferation pathways.

Nevertheless, it should be considered that RNA sequencing results is a very useful technique to generate hypothesis, but these findings should be confirm later. Western blot or real-time PCR should be used to clarify this aspect. For instance, it would be useful to understand if there is a specific cell expression within the tumour microenvironment, quantify the apoptosis in S1 tumour compared to ACE samples by check the activation state of the upregulated caspases or quantify the mRNA degradation.

#### **5.4.5.2 Proteomic analysis of S1 and ACE tumours**

Firstly, ACE tumours have a significant increased expression of *Sftpc* gene. This gene is also known as SPC, and it is specifically expressed in AT2 and BASCs. This data reinforces the idea that SPC tumours are the main tumour type in the ACE group.

Preliminary proteomic data showed again an increased expression of *Pgm1* in ACE tumours. This supports the data showed in the CC10 vs SPC proteomic analysis where SPC tumours also present an increased expression of *Pgm1* and it also provides support for the transcriptomic data shown above where ACE tumours overexpressed genes related to glycogen metabolism.

Moreover, S1 tumours use of glucose for glycolysis is supported on the expression of several glycolytic proteins. This may indicate a conversion of glucose into lactate and a later transport of it outside the cell in S1 tumours. This is supported by the expression of hypoxia targets in the transcriptomic data, as it has been shown above, where the entrance of pyruvate into the TCA cycle is inhibited and its conversion to lactate is promoted under low oxygen conditions. Additionally, hexokinase 3 supports the idea that S1 tumours preferentially uptake glucose from the extracellular space.

The expression of genes related to the nucleotide synthesis in S1 tumours reveal a diversion of the glucose towards the PPP and the synthesis of nucleotides to support cell proliferation. S1 subgroup is dominated by CC10 tumour, and it has been argued before that their expression of MYC, KRAS and E2F targets gene sets may indicate a higher proliferative rate in these samples. This proliferation needs to be supported by the nucleotide synthesis.

Additionally, *Got1* and *Glul* expression on S1 tumour suggest an alternative source of energy to fill the TCA cycle, as discussed before by the upregulated expression of glutamate transporter in the mitochondria (Nagamani and Erez, 2016).

Finally, S1 tumours have an increase expression for translational proteins. This may highlight that S1 tumours are transcriptionally more active than ACE tumours. It has been proved shown that tumours arise from genetic alterations that may lead to dysregulated transcriptional programs. These dysregulated programs can cause cancer cells to become highly dependent on certain regulators of gene expression and will lead to a higher transcriptional activity on these tumour cells (Bradner et al., 2017).

#### 5.4.5.3 Metabolomic characterisation of S1 and ACE tumours

Supporting GSEA analysis that identified upregulated glycolytic genes in S1 and CC10 tumours, glyceraldehyde 3-phosphate, a glycolysis intermediate and NADH, a glycolysis product have a higher abundance in S1 tumours when compared to ACE samples. As suggested above, glucose might be converted into lactate, although this metabolite is not significantly upregulated in S1 tumours. This could be explained by the increase expression of MCT lactate transport (*Slc16a3*) seen in the proteomic analysis, which transport lactate outside the cell and avoids its accumulation.

More interestingly, several products from purine and pyrimidine metabolites have been found in S1 tumours. This is interesting as it may indicate that not all glucose carbons are used to obtain energy, some are diverted into the pentose phosphate pathway for nucleotide biosynthesis. This is supported by the overexpression of PPP, purine and pyrimidine synthesis genes in S1 tumours. The accumulation of glyceraldehyd 3-phosphate also supports this idea as it can be converted into ribose 5-phosphate by TKT enzyme which support the nucleotide synthesis (Ge et al., 2020). Additionally, the [<sup>18</sup>F]FDG uptake seen in PET imaging for these tumours and the positive p value in the pyrimidine and purine metabolism and glycolysis in the joint analysis pathway also support this theory.

Moreover, S1 tumours show an increased expression of glutamate, which may indicate a glutamate contribution to the TCA cycle. This is supported by the

upregulation of *Got1* and *Glul* enzymes seen in the proteomic results, the increased of *Slc16a22* glutamate transport into the mitochondria observed in RNAseq analysis, and the accumulation of  $\alpha$ -ketoglutarate in S1 cells. Moreover, joint pathway analysis revealed differences in the glutamine/glutamate metabolism, supporting this theory. Although more experiments are needed to confirm this TCA cycle support by multiple intermediates. This suggest that S1 tumours use glucose to sustain nucleotide biosynthesis and they use secondary TCA precursors to sustain energy production.

On the other hand, aspartate and alanine are significantly upregulated in ACE samples. These metabolites are able to supply the TCA cycle by being converted to fumarate, oxalacetate or pyruvate. Moreover, from PET data it is know that these tumours uptake acetate (Alkan and Bogner-Strauss, 2019, Adeva-Andany et al., 2014). This acetate can be converted into palmitate in the cytoplasm, but it can also be converted into Acetyl-CoA in the mitochondria through the action of ACSS1 and tis can be incorporated into the TCA cycle. Acetate tracing might be necessary to clarify this, even though ACE tumours may be able to rely on other nutrients apart from glucose to feed into the TCA, acetate may be used for fatty acid production or lipid metabolism, as indicates the cholesterol/fatty acid related genes expressed in ACE tumours, although the utilisation of lipid metabolism must be confirmed with lipidomics. At the same time, data seems to indicate that ACE tumours my rely on the TCA cycle sustained by amino acids to obtain energy and they rely less on glycolysis, although the expression of upregulated glycolytic genes in this samples and the lack of expression of LDHA may indicate certain glucose contribution to the TCA.

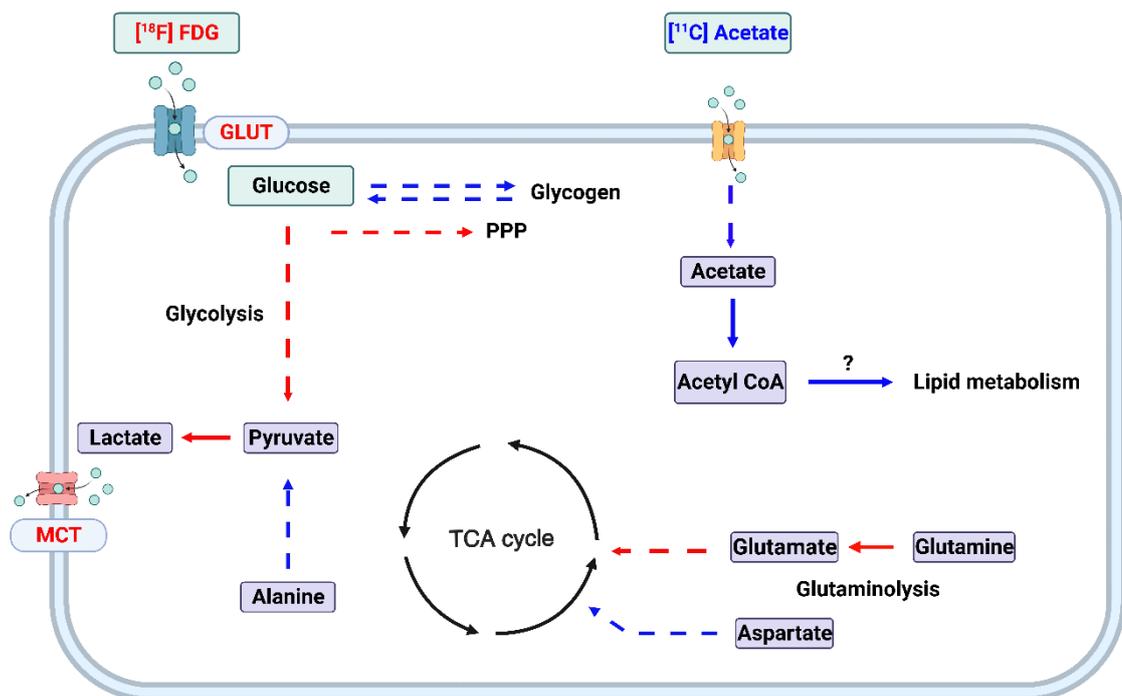
## 5.5 Conclusion

In conclusion, transcriptomic and PET data showed that GEMMs KP adenocarcinomas can be classified within three transcriptomic subgroups (S1, S2 and ACE). S1 and ACE subgroup are significantly enriched of tumours arising from different cells-of-origin, CC10 positive and SPC positive cells, respectively. This tumour classification may determine different metabolic phenotypes.

Firstly, a differential nutrient uptake, S1 tumours being more FDG avid and ACE tumour more acetate avid. Moreover, transcriptomic and metabolomic data

suggest some differences between these subgroups that need to be confirmed. Data indicates that S1 tumours may be more glycolytic, proliferative, hypoxic and apoptotic than ACE tumours. Additionally, data indicates that S1 tumours uptake glucose from the extracellular environment, and they use it for glycolysis and the synthesis and transport of lactate and nucleotide biosynthesis through the pentose phosphate pathway, while they use glutamine to sustain TCA cycle.

ACE tumours, which are dominated by SPC tumours seems to be less dependent on glycolysis, although data indicates they may use of inner glucose storage to sustain a lower glycolytic activity. The expression of *Pdk2/3* may indicate the deviation of acetate into lipid metabolism. Although quantification of fatty acid such as palmitate is needed to confirm this. The accumulation of alanine and aspartate may indicate an alternative source of carbon for the TCA cycle.



**Figure 5.29. Schematic representation of the overall differences between S1 and ACE tumours.** In red pathways upregulated in S1 tumours. In blue pathways upregulated in ACE tumours. PPP pentose phosphate pathway, MCT monocarboxylate transporter. Created with BioRender.com.

## Chapter 6 Discussion

The combined analysis of transcriptomic, proteomic, and metabolomic data has revealed an insight into the potential differences between S1 and ACE tumours, which are predominantly enriched with CC10 and SPC tumours, respectively. The initial findings suggest that S1 tumours may exhibit characteristics such as increased glycolysis, heightened proliferation, hypoxia, and enhanced apoptotic activity compared to ACE tumours. However, to gain a more comprehensive understanding of the true differences between these two tumour types, further experimental investigations are needed. These additional experiments will be crucial to confirm the precise underlying mechanisms that differentiate S1 and ACE tumours.

### 6.1 Effect of the cell-of-origin in metabolism of the resultant tumour

Throughout this project it has been discussed differences in expression for CC10 and SPC tumours induced in KP mice. Among these differences it has been observed a significant enrichment in proliferative pathways such as *KRas*, *MYC* or oestrogen related genes and an increased in glycolysis for these tumours. On the other hand, SPC tumours present a higher immune activity and further experiments are needed to confirm an enhanced lipid metabolism. These results are also supported by S1 and ACE tumours classification, which are significantly enriched in CC10 and SPC tumours, respectively.

However, more experiments are needed to understand the why *KRas* and *P53* mutations induce a different set of transcriptomic changes when they are mutated in different cell types. The mechanism behind this is not known and has not been explored in this project. However, there are several explanations for this. CC10 positive cells (club cells and BASCs) and SPC positive cells (AT2 cells and BASCs) role in the healthy lung epithelium is different and the cells they differentiate into vary (Chen et al, 2014). Therefore, pathways such as *KRas*, *P53* or *MYC* or metabolic processes like glycolysis might be intrinsically more active in CC10 positive than in SPC positive cells and these features might be inherited by tumour cells. Considering this hypothesis, differences in the transcriptome,

the metabolome or in the epigenetics of these normal cells-of-origin regarding these pathways should be analysed.

Non-transformed cells-of-origin have not been studied in this project. However, it has been proved that each one of them can be isolated by sorting and seeded in 3D cultures. The intrinsic characteristics of these cells and their effect in the transcriptomic and metabolomic features observed in the resultant tumours could be studied using these 3D models.

## **6.2 Limitations of the study**

This study provided valuable insights into murine lung cancer metabolic heterogeneity. However, this research presents a few limitations. First of all, Ad5-mSPC-Cre and Ad5-CC10-Cre vectors have an overlap, as they both target BASCs. Therefore, the pure effect of AT2, club and BASC cells as tumour-initiating cells has not been studied. This overlap might be represented by the S2 transcriptomic subgroup, which may group SPC and CC10 tumours arising from BASCs in mice induced with each virus. However, considering that SPC and CC10 markers may not be conserved in grown tumours and CC10 and SPC tumour signatures have not been developed yet, it can be challenging to interpret the specific cell-of-origin in vivo from late-stage tumours. As AT2, club and BASC cells can be individually isolated, lineage-restricted 3D cultures should be developed to properly assess the differences between these three cell types as cells-of-origin of lung cancer.

Secondly, the virus specificity has not been tested in detail. Data shown in chapter three shows a high virus specificity for each cell type, although this specificity does not reach 100%. The possibility that these viruses are targeting other tumour-initiating cells has not been explored and it might be the case that the tumours arising from SPC and CC10 mice might not be exclusively arising from AT2, club or BASC cells.

Thirdly, S1, S2 and ACE tumour classification has been calculated using K-means on row-scaled normalised gene expression. K-means analysis assumes that the data can be classified into a number of groups previously defined, in this case, three groups. It can be argued that this is a biased analysis and therefore, the

resulting groups lack of biological value. However, during this research project it has been argued that the ACE subgroup is biologically meaningful as it is distinctively enriched in genes from [<sup>11</sup>C]acetate gene signature while S1 and S2, previously defined in the literature, are enriched in genes from [<sup>18</sup>F]FDG gene set. This proves that ACE tumours have imaging differences which may indicate differences in their tumour metabolism.

Finally, it is important to consider that most tumour metabolic characterisation data have been obtained using multiple omics techniques, which constitute a very useful way to generate hypothesis. In this case, omic data together with imaging data, have suggested differences in glycolysis and differences and fatty acid metabolism are suspected. However, metabolic changes need to be confirmed by glucose and acetate tracing experiments so glucose and acetate utilisation can be accurately estimate. Therefore, no strong conclusion can be made with the current available data.

This tracing could be performed in vivo, but due to the current overlap between Ad5-CC10-Cre and Ad5-mSPC-Cre vectors, 3D culture in vitro tracing is being considered. Nonetheless, in vitro tracing also faces some challenges. The 3D culture developed is complex with a lower and an upper chamber, two different mediums in place, Matrigel and support cells and organoids in the same upper chamber. Therefore, 2D tracing protocols would need to be adapted. Moreover, 3D in vitro metabolic results and findings are highly dependent on the chosen medium for that cell culture. This may prevent in vitro result to replicate in vivo findings.

## **6.3 Future work**

### **6.3.1 Lung Cancer Organoids**

To further validate the lineage-restricted culture protocol, another checkpoint is necessary as it is essential to verify the efficiency of cell activation to ensure that all cells have been successfully targeted by the vector. To address this, we can capitalize on the MTMGKP mouse model, where activated cells express EGFP (green fluorescence), while non-activated ones express tdTomato (red fluorescence). Flow cytometry will be used to detect and quantify green and red

fluorescent at different time points and Incucyte S3 (Sartorius) live-cell analysis instrument can be used to detect and quantify these fluorescence signals, providing insights into the effectiveness of cell transformation. Furthermore, while we have successfully seeded AT2 cells and seen some progress, our ultimate aim is to generate organoids from club, basal, and BASCs cells too in order to as to create lineage-restricted organoids.

The next step would be to collaborate with the metabolomic department at the Beatson Institute to develop a metabolomic analysis protocol for these LTOs. However, this protocol development comes with several challenges. Firstly, the limited cell number in 3D cultures is a challenge for metabolomic techniques, as they typically require a minimum number of cells for accurate analysis. Unlike 2D cell lines that continue to grow as long as they are split and have access to nutrients, 3D cultures are believed to have significantly lower cell numbers (Temple et al., 2022). Secondly, for a precise metabolic analysis of the target cells, it becomes crucial to isolate the epithelial cells from the surrounding endothelial support cells in the organoids. Thirdly, the presence of Matrigel in the culture can complicate the metabolomic analysis. Matrigel may interfere with chromatography and isotope detection or be identified as background noise. Lastly, 3D cultures present difficulties in data normalization. In 2D in vitro metabolomic analysis, data is typically normalized by the number of cells. However, as explained in section 3.1, cell counting may not always be possible for LTO cultures. However, DNA quantification has raised as an effective normalisation technique for 3D culture. Once accomplished, this analysis will provide valuable insights into the metabolic alterations and characteristics of the tumours derived from these different cell types, and it we will be able to expand this knowledge to transcriptomic or proteomic analysis which will help us to better understand the tumour biology.

### **6.3.2 Metabolic, Transcriptomic and Proteomic characterisation**

The transcriptomic data has revealed a potentially higher level of immunological activity within the SPC tumour environment when compared to CC10 tumours. However, the nature of the immune system's involvement and its role in tumour progression require further investigation and clarification. Towards this aim, we are seeking for collaboration with the BAIR department and Leo Carlin's research

group, known for their expertise in lung cancer and the immune system's role in cancer. We aim to gain deeper insights into the immunological aspects specific to SPC tumours, ultimately enhancing our understanding of their behaviour and response to the immune system. Furthermore, there is a hypothesis that CC10 tumours may exhibit higher proliferation rates than SPC tumours, as suggested by the upregulation of proliferative pathways such as KRAS, MYC, and E2F targets. To validate this hypothesis, we plan to perform staining experiments using 5'-bromo-2'-deoxyuridine (BrdU) (Gratzner, 1982) and Ki-67 markers (Miller et al., 2018) in both CC10 and SPC tumours. These experiments will provide empirical evidence to confirm whether CC10 tumours indeed display increased cellular proliferation, shedding light on the distinctive characteristics of these tumour types.

Sftpc (SPC) and Scgb1a1 (CC10) staining techniques will also be used to investigate whether these markers remain preserved within the tumour cells. Our aim is to classify these tumours into distinct categories based on their cell-of-origin, which includes SPC positive lesions (originating from AT2 cells), CC10 positive tumours (arising from club cells), and SPC and CC10 positive tumours (originating from BASCs). By doing so, we hope to shed light not only on the cell-of-origin for CMV tumours but also for S2 tumours, providing valuable insights into how the cell-of-origin, particularly BASCs, may define the transcriptomic subtype of these tumours. Furthermore, lungs from individual PET imaged mice were dissected and fixed in embedded paraffin blocks. Our aim is to stain them with FASN, SOX2 and Cadherin-16 to further classify them into ACE, S1 and S2 respectively.

S1 tumours are shown to be more glycolytic than ACE tumours and given that this transcriptomic subtype is dominated by CC10 samples (characterised by KRAS, MYC and E2F target expression) they are likely to be more proliferative than ACE lesions. However, they also have an enriched expression of apoptosis and P53 targets and hypoxia related genes. Our hypothesis suggests that as S1 progress quicker than ACE lesions consequently they may develop some hypoxic areas where apoptosis might be induced by hypoxic target genes while more irrigated areas are more proliferative. There are several probes and assays to detect hypoxic areas in fixed tissues and several staining to detect apoptosis

markers or P53 targets such as caspase 3 or p21 respectively. The aim is to identify these areas and study their colocalization with apoptotic areas to establish an association between hypoxia and cell death in S1 tumours.

The proteomic data in this study has undergone a preliminary analysis, but there is potential for further in-depth exploration. Conducting pathway analysis and integrating it with metabolomic data could offer substantial support for the findings presented in this thesis. For instance, it may corroborate the presence of immunogenic genes in SPC tumours and a deep understanding on apoptotic pathways in S1 tumours. Additionally, this approach may yield insights into acetate utilization by ACE tumours and its potential role in lipid metabolism. By delving deeper into these data sets and their interactions, we can enhance our understanding of the mechanisms underlying these tumour types, contributing valuable information to the overall thesis.

To provide additional validation for the observed alterations in glycolytic and nucleotide synthesis pathways within S1 tumours, we plan to employ *in vivo* glucose tracing techniques. This approach will help to understand the utilization of glucose by S1 tumours. Simultaneously, we will employ lipidomic methods and *in vivo* acetate tracing to investigate and quantify acetate utilization within ACE tumours. By utilizing these complementary techniques, we aim to gain a comprehensive understanding of how S1 and ACE tumours metabolize glucose and acetate, respectively, and further solidifying our findings in these specific tumour types.

Finally, it would be important to corroborate the existence of these three transcriptomic subtypes (S1, S2 and ACE) in human tumours with G12D mutation in KRAS and deletion in P53 and their imaging differences using multiple transcriptomic databases available online. At the same time, we could use the data obtained in this project to obtain gene signature which allows us to identify the cell-of-origin of a developed KP tumour. IHC or transcriptomic data in human tumour samples will then be used to identify the tumour initiating cell.

## Supplementary information

Genes included in [<sup>11</sup>C]acetate gene signature:

Notum, Cst8, Klf15, Bex2, Bex1, Cacna1s, Pcdh20, Tnfaip8l3, Nkd1, Rarb, Slc22a18, S100g, Npw, Elovl1, Edar, Ascl2, Gjb6, Nol3, Tnni3k, Panx2, Sftpc, Lrrc14b, Nphs2, Tdrd5, Wnt3a, Ang4, Ang5, Ang6, Tmod1, Nrxa1, Sec14l4, Mesp1, Rtkn2, Trim67, Fbxo2, Dcl1, Hnmt, Kcnj15, Pmp22, Blvrb, M1ap, Lcn2, Lrrc3b, Rilp, Sybu, Vnn3, Mesp2, Prmt8, Capn11, Chchd10, Ramp1, Ifit3b, Sema4f, Sftpa1, Rbp7, Dcxr, Epdr1, Tspyl4, Ang, Ang2, Pcdhga3, Azin2, Ptn, Cd207, Khk, P2rx5, Cd300lg and Slc30a2.

Genes included in [<sup>18</sup>F]FDG gene signature:

Ggt6, Gm5414, Krt6b, Krt6a, Aldh1a3, Vil1, Avil, Calml4, Rgs7, Gkn2, Pthlh, Scara3, Unc13c, Fhit, Rhobtb3, Lman1l, Cplx3, Mst1r, Vstm5, Fsd1, Scn8a, Prrt1, Clstn3, Rps6ka6, Spp2, Sox6, Col6a6, Has2, Sgk2, Apoa5, Cfi, Atp8a2, Cyp11a1, Il22ra1, Amot, Adm2, Fhl2, Entpd8, Iqcf5, Nup62cl, Absg5, Cdhr2, Fga, Slc6a19, Lrrc31, Clca1, Mmp9, Tmprss11e, Pla2g12b, Soat2, Arhgap40, Slc17a4, Dscaml1, Bcas1, Hnf4a, B4galnt3, Apoc2, Gm44805, Cyp24a1, Cbln3, Trim71, Drd3, Mkrn2os, Ephx2, Nr1h3, Gluc, Gm20547, Cfb, Sdsl, Syt14, Apoh, Sox21, Prprt, Paqr5, Tpbp, Kcnh7, Vsig1, Elovl6, Prtg, Apoa1, Fstl4, Nlrc4, Eps8l3, Amer2, Acsbg1, Gstm5, Igsf11, Capn8, Pkp1, Gm15262, Hpx, Agt, Myo1a, Ankrd34c, Fapb2, Nkain1, Snap91, Degs2, Pklr, Abcg8, Sgms2, Gcnt3, Papss2, Trim10, Ctse, Pdzk1, Tmed6, Slc7a9, Fam186b, Otc, Inpp5j, F2, Prom2, Foxa3, Ankrd1, Baiap2l2, Pdyn, Galnt16, Rnd183, Ogdhl, Plekhg3, Gpr142, Sdr9c7, Dnajc22, Cubn, Csf3, Ddx25, Mtmr11, Evolv2, Fsol1, Serpina10, Crev3l3, Brbd17, Ano7, Cdh17, Dcd, Fxyd3, Cntn2, Zpbp2, Snai3, Pscn, Vwce, Gcnt1, Fgfr4, Add2, Ngf, Slc25a48, Trhde, Tns4, Zan, Hecw1, Prom1, Tm4sf20, Unc5cl, Pcsk9, Snap25, Misp, Syt16, Adgrg7, Snph, Krt20, Pipox, Cacna1h, Srrm4, Acta1, Mtmr7, Itga2, Adcy1, Poln, Kcnip2, Fa2h, Rbp2, Ryr3, Ak4, Cxxc4, Srd5a1, Styk1, Phyhipl, Onecut2, Cdhr5, Habp2, Stxbp5l, Card14, Smpd3, Arc, Agr2, Can, Nr1i2, Lypd8, Pde6c, Tspo2, Cyb5r2, Fermt1, Cdh6, Cldn23, Hnf1a, Stk32a, Anxa10, Tmc3, Frem2, Dsc3, Bche, Fer1l6, Sptssb, Tff1, Nfe2l3, Ambp, Alpl2, Alpi, Akp3, Apob, Nlrp10, Tmc5, Lrrn1, Tph1, Kpna7, Pyy, Kif12, Ttr, Pclo, hsd17b2, Prrx2, Lypd6,

Ank1, Mmp13, Prss33, Rundc3b, Myrf, Arl14, Inhba, Atg9b, Lrrc9, Bco1, Anks4b, Cntnap2, Gpa33, Igfals, Cldn6, Tenm4, Rab3b, Tm4sf4, Igsf23 and Bfsp1.

## References

2013. Suspected cancer: recognition and referral. National Institute for Health and Care Excellence.
- ADEVA-ANDANY, M., LOPEZ-OJEN, M., FUNCASTA-CALDERON, R., AMENEIROS-RODRIGUEZ, E., DONAPETRY-GARCIA, C., VILA-ALTESOR, M. & RODRIGUEZ-SEIJAS, J. 2014. Comprehensive review on lactate metabolism in human health. *Mitochondrion*, 17, 76-100.
- AHMAD, A. 2016. *Lung Cancer and Personalized Medicine Current Knowledge and Therapies*
- AKBAY, E. A., KOYAMA, S., CARRETERO, J., ALTABEF, A., TCHAICHA, J. H., CHRISTENSEN, C. L., MIKSE, O. R., CHERNIACK, A. D., BEAUCHAMP, E. M., PUGH, T. J., WILKERSON, M. D., FECCI, P. E., BUTANEY, M., REIBEL, J. B., SOUCHERAY, M., COHOON, T. J., JANNE, P. A., MEYERSON, M., HAYES, D. N., SHAPIRO, G. I., SHIMAMURA, T., SHOLL, L. M., RODIG, S. J., FREEMAN, G. J., HAMMERMAN, P. S., DRANOFF, G. & WONG, K. K. 2013. Activation of the PD-1 pathway contributes to immune escape in EGFR-driven lung tumors. *Cancer Discov*, 3, 1355-63.
- AL-AMRANI, S., AL-JABRI, Z., AL-ZAABI, A., ALSHEKAILI, J. & AL-KHABORI, M. 2021. Proteomics: Concepts and applications in human medicine. *World J Biol Chem*, 12, 57-69.
- ALKAN, H. F. & BOGNER-STRAUSS, J. G. 2019. Maintaining cytosolic aspartate levels is a major function of the TCA cycle in proliferating cells. *Mol Cell Oncol*, 6, e1536843.
- ALSEEKH, S., AHARONI, A., BROTMAN, Y., CONTREPOIS, K., D'AURIA, J., EWALD, J., J, C. E., FRASER, P. D., GIAVALISCO, P., HALL, R. D., HEINEMANN, M., LINK, H., LUO, J., NEUMANN, S., NIELSEN, J., PEREZ DE SOUZA, L., SAITO, K., SAUER, U., SCHROEDER, F. C., SCHUSTER, S., SIUZDAK, G., SKIRYCZ, A., SUMNER, L. W., SNYDER, M. P., TANG, H., TOHGE, T., WANG, Y., WEN, W., WU, S., XU, G., ZAMBONI, N. & FERNIE, A. R. 2021. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat Methods*, 18, 747-756.
- ALTAN, M., KIDWELL, K. M., PELEKANOU, V., CARVAJAL-HAUSDORF, D. E., SCHALPER, K. A., TOKI, M. I., THOMAS, D. G., SABEL, M. S., HAYES, D. F. & RIMM, D. L. 2018. Association of B7-H4, PD-L1, and tumor infiltrating lymphocytes with outcomes in breast cancer. *NPJ Breast Cancer*, 4, 40.
- ANCEY, P. B., CONTAT, C. & MEYLAN, E. 2018. Glucose transporters in cancer - from tumor cells to the tumor microenvironment. *FEBS J*, 285, 2926-2943.
- ANCRILE, B., LIM, K. H. & COUNTER, C. M. 2007. Oncogenic Ras-induced secretion of IL6 is required for tumorigenesis. *Genes Dev*, 21, 1714-9.

- ANDREW THOMPSON, J. S., KARSTEN KUHN 2003. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Analytical Chemistry*, 75.
- ARAVIND SUBRAMANIAN, P. T. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *PNAS*.
- ASLAM, B., BASIT, M., NISAR, M. A., KHURSHID, M. & RASOOL, M. H. 2017. Proteomics: Technologies and Their Applications. *J Chromatogr Sci*, 55, 182-196.
- BALMAIN, B. H. A. A. 2001. Building 'validated' mouse models of human cancer. *Current Opinion in Cell Biology*, 13.
- BARCELO, J., SAMAIN, R. & SANZ-MORENO, V. 2023. Preclinical to clinical utility of ROCK inhibitors in cancer. *Trends Cancer*, 9, 250-263.
- BARKAUSKAS, C. E., CHUNG, M. I., FIORET, B., GAO, X., KATSURA, H. & HOGAN, B. L. 2017. Lung organoids: current uses and future promise. *Development*, 144, 986-997.
- BARTA, J. A., POWELL, C. A. & WISNIVESKY, J. P. 2019. Global Epidemiology of Lung Cancer. *Ann Glob Health*, 85.
- BARTRONS, R., SIMON-MOLAS, H., RODRIGUEZ-GARCIA, A., CASTANO, E., NAVARRO-SABATE, A., MANZANO, A. & MARTINEZ-OUTSCHOORN, U. E. 2018. Fructose 2,6-Bisphosphate in Cancer Cell Metabolism. *Front Oncol*, 8, 331.
- BEGER, R. D. 2013. A review of applications of metabolomics in cancer. *Metabolites*, 3, 552-74.
- BERGER, A. 2003. Positron emission tomography. *British Medical Journal*.
- BLANPAIN, C. & FUCHS, E. 2014. Stem cell plasticity. Plasticity of epithelial stem cells in tissue regeneration. *Science*, 344, 1242281.
- BONNEFONT, J. P., DJOUADI, F., PRIP-BUUS, C., GOBIN, S., MUNNICH, A. & BASTIN, J. 2004. Carnitine palmitoyltransferases 1 and 2: biochemical, molecular and medical aspects. *Mol Aspects Med*, 25, 495-520.
- BOUMELHA, J. D. C. T., SOPHIE LAW, EMILY K. ROMERO-CLAVIJO, PABLO COELHO, MATTHEW A. NG, KEVIN MUGARZA, EDURNE MOORE, CHRISTOPHER RANA, SAREENA CASWELL, DEBORAH R. MURILLO, MIGUEL HANCOCK, DAVID C. ARGYRIS, PROKOPIOS P. BROWN, WILLIAM L. DURFEE, CAMERON LARSON, LINDSAY K. VOGEL, RACHEL I. SUÁREZ-BONNET, ALEJANDRO PRIESTNALL, SIMON L. EAST, PHILIP ROSS, SARAH J. KASSIOTIS, GEORGE MOLINA-ARCAS, MIRIAM SWANTON, CHARLES HARRIS, REUBEN DOWNWARD, JULIAN 2022. An immunogenic model of KRAS-mutant lung cancer for study of targeted therapy and immunotherapy combinations. *Cancer Research*.

- BRADNER, J. E., HNISZ, D. & YOUNG, R. A. 2017. Transcriptional Addiction in Cancer. *Cell*, 168, 629-643.
- BRAWLEY, O. W., GLYNN, T. J., KHURI, F. R., WENDER, R. C. & SEFFRIN, J. R. 2014. The first Surgeon General's report on smoking and health: the 50th anniversary. *CA Cancer J Clin*, 64, 5-8.
- BROWN, K. F., RUMGAY, H., DUNLOP, C., RYAN, M., QUARTLY, F., COX, A., DEAS, A., ELLISS-BROOKES, L., GAVIN, A., HOUNSOME, L., HUWS, D., ORMISTON-SMITH, N., SHELTON, J., WHITE, C. & PARKIN, D. M. 2018. The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. *Br J Cancer*, 118, 1130-1141.
- BUZZEO, M. P., SCOTT, E. W. & COGLE, C. R. 2007. The hunt for cancer-initiating cells: a history stemming from leukemia. *Leukemia*, 21, 1619-27.
- CANCER GENOME ATLAS RESEARCH, N. 2014. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, 511, 543-50.
- CAO, B., YANG, S., LI, W., CHEN, H., CHEN, Y., LIU, Y. & LIU, B. 2018. GMZ-1 is a podophyllotoxin derivative that suppresses growth and induces apoptosis in adriamycin-resistant K562/A02 cells through modulation of MDR1 expression. *Mol Med Rep*, 17, 474-478.
- CHANDEL, N. S. 2021. Glycolysis. *Cold Spring Harb Perspect Biol*, 13.
- CHANDEL, R. J. D. A. N. S. 2016. Fundamentals of cancer metabolism. *Science Advances*.
- CHANG, D. R., MARTINEZ ALANIS, D., MILLER, R. K., JI, H., AKIYAMA, H., MCCREA, P. D. & CHEN, J. 2013. Lung epithelial branching program antagonizes alveolar differentiation. *Proc Natl Acad Sci U S A*, 110, 18042-51.
- CHATURVEDI, G., SIMONE, P. D., AIN, R., SOARES, M. J. & WOLFE, M. W. 2009. Noggin maintains pluripotency of human embryonic stem cells grown on Matrigel. *Cell Prolif*, 42, 425-33.
- CHEN, H., LIU, H. & QING, G. 2018. Targeting oncogenic Myc as a strategy for cancer treatment. *Signal Transduct Target Ther*, 3, 5.
- CHEN, Y., TOTH, R., CHOCARRO, S., WEICHENHAN, D., HEY, J., LUTSIK, P., SAWALL, S., STATHOPOULOS, G. T., PLASS, C. & SOTILLO, R. 2022. Club cells employ regeneration mechanisms during lung tumorigenesis. *Nat Commun*, 13, 4557.
- CHEN, Z., FILLMORE, C. M., HAMMERMAN, P. S., KIM, C. F. & WONG, K. K. 2014. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat Rev Cancer*, 14, 535-46.

- CHEON, D. J. & ORSULIC, S. 2011. Mouse models of cancer. *Annu Rev Pathol*, 6, 95-119.
- CHOI, J., IICH, E. & LEE, J. H. 2016. Organogenesis of adult lung in a dish: Differentiation, disease and therapy. *Dev Biol*, 420, 278-286.
- CHUNG, W. J., DAEMEN, A., CHENG, J. H., LONG, J. E., COOPER, J. E., WANG, B. E., TRAN, C., SINGH, M., GNAD, F., MODRUSAN, Z., FOREMAN, O. & JUNTILLA, M. R. 2017. Kras mutant genetically engineered mouse models of human cancers are genomically heterogeneous. *Proc Natl Acad Sci U S A*, 114, E10947-E10955.
- CLEVERS, H. 2016. Modeling Development and Disease with Organoids. *Cell*, 165, 1586-1597.
- CONDE, E., ANGULO, B., REDONDO, P., TOLDOS, O., GARCIA-GARCIA, E., SUAREZ-GAUTHIER, A., RUBIO-VIQUEIRA, B., MARRON, C., GARCIA-LUJAN, R., SANCHEZ-CESPEDES, M., LOPEZ-ENCUENTRA, A., PAZ-ARES, L. & LOPEZ-RIOS, F. 2010. The use of P63 immunohistochemistry for the identification of squamous cell carcinoma of the lung. *PLoS One*, 5, e12209.
- COSTA, D. B., HALMOS, B., KUMAR, A., SCHUMER, S. T., HUBERMAN, M. S., BOGGON, T. J., TENEN, D. G. & KOBAYASHI, S. 2007. BIM mediates EGFR tyrosine kinase inhibitor-induced apoptosis in lung cancers with oncogenic EGFR mutations. *PLoS Med*, 4, 1669-79; discussion 1680.
- CUNNIFF, B., DRUSO, J. E. & VAN DER VELDEN, J. L. 2021. Lung organoids: advances in generation and 3D-visualization. *Histochem Cell Biol*, 155, 301-308.
- DAEMEN, A., COOPER, J. E., MYRTA, S., WONGCHENKO, M. J., LIN, E., LONG, J. E., FOREMAN, O., MODRUSAN, Z., TREMAYNE, J. R., DE LA CRUZ, C. C., MERCHANT, M., MARTIN, S. E., YAN, Y. & JUNTILLA, M. R. 2021. Transcriptional Subtypes Resolve Tumor Heterogeneity and Identify Vulnerabilities to MEK Inhibition in Lung Adenocarcinoma. *Clin Cancer Res*, 27, 1162-1173.
- DANG, L., WHITE, D. W., GROSS, S., BENNETT, B. D., BITTINGER, M. A., DRIGGERS, E. M., FANTIN, V. R., JANG, H. G., JIN, S., KEENAN, M. C., MARKS, K. M., PRINS, R. M., WARD, P. S., YEN, K. E., LIAU, L. M., RABINOWITZ, J. D., CANTLEY, L. C., THOMPSON, C. B., VANDER HEIDEN, M. G. & SU, S. M. 2009. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature*, 462, 739-44.
- DANIEL L. HAMILTON, K. A. 1984. Site-specific Recombination by the Bacteriophage P1 lox-Cre system. *Journal of Molecular Biology*, 78.
- DANOPOULOS, S., SHIOSAKI, J. & AL ALAM, D. 2019. FGF Signaling in Lung Development and Disease: Human Versus Mouse. *Front Genet*, 10, 170.
- DAVID W. FRY, P. J. H., PAUL R. KELLER, WILLIAM L. ELLIOTT, MARY ANNE MEADE, ERIN TRACHET, MUDHER ALBASSAM, XIANXIAN ZHENG, WILBUR R.

- LEOPOLD, NANCY K. PRYER, AND PETER L. TOOGOOD 2004. Specific inhibition of cyclin-dependent kinase 4/6 by PD 0332991 and associated antitumor activity in human tumor xenografts. *Molecular Cancer Therapeutics*, 3.
- DAVIDSON, M. R., GAZDAR, A. F. & CLARKE, B. E. 2013. The pivotal role of pathology in the management of lung cancer. *J Thorac Dis*, 5 Suppl 5, S463-78.
- DAVIDSON, S. M., PAPAGIANNAKOPOULOS, T., OLENCHOCK, B. A., HEYMAN, J. E., KEIBLER, M. A., LUENGO, A., BAUER, M. R., JHA, A. K., O'BRIEN, J. P., PIERCE, K. A., GUI, D. Y., SULLIVAN, L. B., WASYLENKO, T. M., SUBBARAJ, L., CHIN, C. R., STEPHANOPOULOS, G., MOTT, B. T., JACKS, T., CLISH, C. B. & VANDER HEIDEN, M. G. 2016. Environment Impacts the Metabolic Dependencies of Ras-Driven Non-Small Cell Lung Cancer. *Cell Metab*, 23, 517-28.
- DE TORRES, J. P., MARIN, J. M., CASANOVA, C., COTE, C., CARRIZO, S., CORDOBA-LANUS, E., BAZ-DAVILA, R., ZULUETA, J. J., AGUIRRE-JAIME, A., SAETTA, M., COSIO, M. G. & CELLI, B. R. 2011. Lung cancer in patients with chronic obstructive pulmonary disease-- incidence and predicting factors. *Am J Respir Crit Care Med*, 184, 913-9.
- DEVER, T. E., DINMAN, J. D. & GREEN, R. 2018. Translation Elongation and Recoding in Eukaryotes. *Cold Spring Harb Perspect Biol*, 10.
- DODD, R. D., ANO, L., BLUM, J. M., LI, Z., VAN MATER, D. & KIRSCH, D. G. 2015. Methods to generate genetically engineered mouse models of soft tissue sarcoma. *Methods Mol Biol*, 1267, 283-95.
- DONG, Y., TU, R., LIU, H. & QING, G. 2020. Regulation of cancer cell metabolism: oncogenic MYC in the driver's seat. *Signal Transduct Target Ther*, 5, 124.
- DOOLEY, A. L., WINSLOW, M. M., CHIANG, D. Y., BANERJI, S., STRANSKY, N., DAYTON, T. L., SNYDER, E. L., SENNA, S., WHITTAKER, C. A., BRONSON, R. T., CROWLEY, D., BARRETINA, J., GARRAWAY, L., MEYERSON, M. & JACKS, T. 2011. Nuclear factor I/B is an oncogene in small cell lung cancer. *Genes Dev*, 25, 1470-5.
- DOST, A. F. M., MOYE, A. L., VEDAIE, M., TRAN, L. M., FUNG, E., HEINZE, D., VILLACORTA-MARTIN, C., HUANG, J., HEKMAN, R., KWAN, J. H., BLUM, B. C., LOUIE, S. M., ROWBOTHAM, S. P., SAINZ DE AJA, J., PIPER, M. E., BHETARIYA, P. J., BRONSON, R. T., EMILI, A., MOSTOSLAVSKY, G., FISHBEIN, G. A., WALLACE, W. D., KRYSAN, K., DUBINETT, S. M., YANAGAWA, J., KOTTON, D. N. & KIM, C. F. 2020. Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells. *Cell Stem Cell*, 27, 663-678 e8.
- DRAGIC, H., CHAVEROUX, C., COSSET, E. & MANIE, S. N. 2022. Modelling cancer metabolism in vitro: current improvements and future challenges. *FEBS J*.

- DROST, J. & CLEVERS, H. 2018. Organoids in cancer research. *Nat Rev Cancer*, 18, 407-418.
- DUMA, N., SANTANA-DAVILA, R. & MOLINA, J. R. 2019. Non-Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. *Mayo Clin Proc*, 94, 1623-1640.
- DUPAGE, M., DOOLEY, A. L. & JACKS, T. 2009. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nat Protoc*, 4, 1064-72.
- ELTAYEB, K., LA MONICA, S., TISEO, M., ALFIERI, R. & FUMAROLA, C. 2022. Reprogramming of Lipid Metabolism in Lung Cancer: An Overview with Focus on EGFR-Mutated Non-Small Cell Lung Cancer. *Cells*, 11.
- FAN, H., WU, Y., YU, S., LI, X., WANG, A., WANG, S., CHEN, W. & LU, Y. 2021. Critical role of mTOR in regulating aerobic glycolysis in carcinogenesis (Review). *Int J Oncol*, 58, 9-19.
- FERONE, G., SONG, J. Y., KRIJGSMAN, O., VAN DER VLIET, J., COZIJNSEN, M., SEMENOVA, E. A., ADAMS, D. J., PEEPER, D. & BERNS, A. 2020. FGFR1 Oncogenic Activation Reveals an Alternative Cell of Origin of SCLC in Rb1/p53 Mice. *Cell Rep*, 30, 3837-3850 e3.
- FERONE, G., SONG, J. Y., SUTHERLAND, K. D., BHASKARAN, R., MONKHORST, K., LAMBOOIJ, J. P., PROOST, N., GARGIULO, G. & BERNS, A. 2016. SOX2 Is the Determining Oncogenic Switch in Promoting Lung Squamous Cell Carcinoma from Different Cells of Origin. *Cancer Cell*, 30, 519-532.
- FISCHER, B. M. & MORTENSEN, J. 2006. The future in diagnosis and staging of lung cancer: positron emission tomography. *Respiration*, 73, 267-76.
- FRANCIS, G. L. 2010. Albumin and mammalian cell culture: implications for biotechnology applications. *Cytotechnology*, 62, 1-16.
- FUHRMANN, S., ZOU, C. & LEVINE, E. M. 2014. Retinal pigment epithelium development, plasticity, and tissue homeostasis. *Exp Eye Res*, 123, 141-50.
- GAO, S. P., MARK, K. G., LESLIE, K., PAO, W., MOTOI, N., GERALD, W. L., TRAVIS, W. D., BORNMANN, W., VEACH, D., CLARKSON, B. & BROMBERG, J. F. 2007. Mutations in the EGFR kinase domain mediate STAT3 activation via IL-6 production in human lung adenocarcinomas. *J Clin Invest*, 117, 3846-56.
- GARG, A., SUI, P., VERHEYDEN, J. M., YOUNG, L. R. & SUN, X. 2019. Consider the lung as a sensory organ: A tip from pulmonary neuroendocrine cells. *Curr Top Dev Biol*, 132, 67-89.
- GAVISH, A., TYLER, M., GREENWALD, A. C., HOFFFLIN, R., SIMKIN, D., TSCHERNICHOVSKY, R., GALILI DARNELL, N., SOMECH, E., BARBOLIN, C., ANTMAN, T., KOVARSKY, D., BARRETT, T., GONZALEZ CASTRO, L. N., HALDER, D., CHANOCH-MYERS, R., LAFFY, J., MINTS, M., WIDER, A., TAL,

- R., SPITZER, A., HARA, T., RAITSES-GUREVICH, M., STOSSEL, C., GOLAN, T., TIROSH, A., SUVA, M. L., PURAM, S. V. & TIROSH, I. 2023. Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature*, 618, 598-606.
- GE, T., YANG, J., ZHOU, S., WANG, Y., LI, Y. & TONG, X. 2020. The Role of the Pentose Phosphate Pathway in Diabetes and Cancer. *Front Endocrinol (Lausanne)*, 11, 365.
- GOODWIN, J., NEUGENT, M. L., LEE, S. Y., CHOE, J. H., CHOI, H., JENKINS, D. M. R., RUTHENBORG, R. J., ROBINSON, M. W., JEONG, J. Y., WAKE, M., ABE, H., TAKEDA, N., ENDO, H., INOUE, M., XUAN, Z., YOO, H., CHEN, M., AHN, J. M., MINNA, J. D., HELKE, K. L., SINGH, P. K., SHACKELFORD, D. B. & KIM, J. W. 2017. The distinct metabolic phenotype of lung squamous cell carcinoma defines selective vulnerability to glycolytic inhibition. *Nat Commun*, 8, 15503.
- GOVINDAN, R., DING, L., GRIFFITH, M., SUBRAMANIAN, J., DEES, N. D., KANCHI, K. L., MAHER, C. A., FULTON, R., FULTON, L., WALLIS, J., CHEN, K., WALKER, J., MCDONALD, S., BOSE, R., ORNITZ, D., XIONG, D., YOU, M., DOOLING, D. J., WATSON, M., MARDIS, E. R. & WILSON, R. K. 2012. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell*, 150, 1121-34.
- GRATZNER, H. G. 1982. Monoclonal Antibody to 5-Bromo- and 5-Iododeoxyuridine: A New Reagent for Detection of DNA Replication. *Science*, 218.
- GREEN, J. G. R. A. H. 1975. Serial Cultivation of Strains of Human Epidermal Keratinocytes: the Formation of Keratinizing Colonies from Single Cells. *Cell*, 6.
- GREIJER, A. E. & VAN DER WALL, E. 2004. The role of hypoxia inducible factor 1 (HIF-1) in hypoxia induced apoptosis. *J Clin Pathol*, 57, 1009-14.
- HANAHAN, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. *Cell*, 144, 646-74.
- HARRIS, M. A., CLARK, J., IRELAND, A., LOMAX, J., ASHBURNER, M., FOULGER, R., EILBECK, K., LEWIS, S., MARSHALL, B., MUNGALL, C., RICHTER, J., RUBIN, G. M., BLAKE, J. A., BULT, C., DOLAN, M., DRABKIN, H., EPPIG, J. T., HILL, D. P., NI, L., RINGWALD, M., BALAKRISHNAN, R., CHERRY, J. M., CHRISTIE, K. R., COSTANZO, M. C., DWIGHT, S. S., ENGEL, S., FISK, D. G., HIRSCHMAN, J. E., HONG, E. L., NASH, R. S., SETHURAMAN, A., THEESFELD, C. L., BOTSTEIN, D., DOLINSKI, K., FEIERBACH, B., BERARDINI, T., MUNDODI, S., RHEE, S. Y., APWEILER, R., BARRELL, D., CAMON, E., DIMMER, E., LEE, V., CHISHOLM, R., GAUDET, P., KIBBE, W., KISHORE, R., SCHWARZ, E. M., STERNBERG, P., GWINN, M., HANNICK, L., WORTMAN, J., BERRIMAN, M., WOOD, V., DE LA CRUZ, N., TONELLATO, P., JAISWAL, P., SEIGFRIED, T., WHITE, R. & GENE ONTOLOGY, C. 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*, 32, D258-61.

- HASEGAWA, K., SATO, A., TANIMURA, K., UEMASU, K., HAMAKAWA, Y., FUSEYA, Y., SATO, S., MURO, S. & HIRAI, T. 2017. Fraction of MHCII and EpCAM expression characterizes distal lung epithelial cells for alveolar type 2 cell isolation. *Respir Res*, 18, 150.
- HAZINI, A., FISHER, K. & SEYMOUR, L. 2021. Deregulation of HLA-I in cancer and its central importance for immunotherapy. *J Immunother Cancer*, 9.
- HE, L., FU, Y., DENG, J., SHEN, Y., WANG, Y., YU, F., XIE, N., CHEN, Z., HONG, T., PENG, X., LI, Q., ZHOU, J., HAN, J., WANG, Y., XI, J. & KONG, W. 2018. Deficiency of FAM3D (Family With Sequence Similarity 3, Member D), A Novel Chemokine, Attenuates Neutrophil Recruitment and Ameliorates Abdominal Aortic Aneurysm Development. *Arterioscler Thromb Vasc Biol*, 38, 1616-1631.
- HECHT, S. S. 1999. Tobacco Smoke Carcinogens and Lung Cancer. *Journal of the National Cancer Institute*, 91.
- HERZOG, E. L., BRODY, A. R., COLBY, T. V., MASON, R. & WILLIAMS, M. C. 2008. Knowns and unknowns of the alveolus. *Proc Am Thorac Soc*, 5, 778-82.
- HOGAN, B. L., BARKAUSKAS, C. E., CHAPMAN, H. A., EPSTEIN, J. A., JAIN, R., HSIA, C. C., NIKLASON, L., CALLE, E., LE, A., RANDELL, S. H., ROCK, J., SNITOW, M., KRUMMEL, M., STRIPP, B. R., VU, T., WHITE, E. S., WHITSETT, J. A. & MORRISEY, E. E. 2014. Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. *Cell Stem Cell*, 15, 123-38.
- HONG, M., TAO, S., ZHANG, L., DIAO, L. T., HUANG, X., HUANG, S., XIE, S. J., XIAO, Z. D. & ZHANG, H. 2020. RNA sequencing: new technologies and applications in cancer research. *J Hematol Oncol*, 13, 166.
- HSU, H. S., LIU, C. C., LIN, J. H., HSU, T. W., SU, K. & HUNG, S. C. 2014. Repair of naphthalene-induced acute tracheal injury by basal cells depends on beta-catenin. *J Thorac Cardiovasc Surg*, 148, 322-32.
- HUA GU, J. D. M., PAUL C. ORBAN, HORST MOSSMANN, KLAUS RAJEWSKY 1994. Deletion of a DNA Polymerase IP Gene Segment in T Cells Using Cell Type-Specific Gene Targeting. *Science*, 265.
- HUANG, F., NI, M., CHALISHAZAR, M. D., HUFFMAN, K. E., KIM, J., CAI, L., SHI, X., CAI, F., ZACHARIAS, L. G., IRELAND, A. S., LI, K., GU, W., KAUSHIK, A. K., LIU, X., GAZDAR, A. F., OLIVER, T. G., MINNA, J. D., HU, Z. & DEBERARDINIS, R. J. 2018. Inosine Monophosphate Dehydrogenase Dependence in a Subset of Small Cell Lung Cancers. *Cell Metab*, 28, 369-382 e5.
- ILARIA GRASSI, C. N., VINCENZO ALLEGRI, JOSHUA JAMES MORIGI, GIAN CARLO MONTINI, PAOLO CASTELLUCCI, STEFANO FANTI 2012. The clinical use of PET with 11C-acetate. *The American Journal of Nuclear Medicine and Molecular Imaging*.

- J.D. WILLIAMSON, P. C. 1968. Use of a New Buffer in the Culture of Animal Cells. *Journal of General Virology* 2.
- JACKSON, E. L., OLIVE, K. P., TUVESON, D. A., BRONSON, R., CROWLEY, D., BROWN, M. & JACKS, T. 2005. The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res*, 65, 10280-8.
- JACKSON, E. L., WILLIS, N., MERCER, K., BRONSON, R. T., CROWLEY, D., MONTOYA, R., JACKS, T. & TUVESON, D. A. 2001. Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev*, 15, 3243-8.
- JASON R ROCK, M. W. O., EMMA L RAWLINS, YUN LU, CHERYL P CLARK, YAN XUE, SCOTT H RANDELL, BRIGID L M HOGAN 2009. Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proceedings of the National Academy of Sciences*, 106, 12771-5.
- JASON R. ROCK, M. W. O., EMMA L. RAWLINS, YUN LU, CHERYL P. CLARK, YAN XUE, SCOTT H. RANDELL AND BRIGID L. M. HOGAN 2009. Basal cells as stem cells of the mouse trachea and human airway epithelium. *PNAS*.
- JASON R. ROCK, M. W. O., EMMA L. RAWLINS, YUN LU, CHERYL P. CLARK, YAN XUE, SCOTT H. RANDELL AND BRIGID L. M. HOGAN 2009. Basal cells as stem cells of the mouse trachea and human airway epithelium. *PNAS*, 106.
- JI, H., RAMSEY, M. R., HAYES, D. N., FAN, C., MCNAMARA, K., KOZLOWSKI, P., TORRICE, C., WU, M. C., SHIMAMURA, T., PERERA, S. A., LIANG, M. C., CAI, D., NAUMOV, G. N., BAO, L., CONTRERAS, C. M., LI, D., CHEN, L., KRISHNAMURTHY, J., KOIVUNEN, J., CHIRIEAC, L. R., PADERA, R. F., BRONSON, R. T., LINDEMAN, N. I., CHRISTIANI, D. C., LIN, X., SHAPIRO, G. I., JANNE, P. A., JOHNSON, B. E., MEYERSON, M., KWIATKOWSKI, D. J., CASTRILLON, D. H., BARDEESY, N., SHARPLESS, N. E. & WONG, K. K. 2007. LKB1 modulates lung cancer differentiation and metastasis. *Nature*, 448, 807-10.
- JOSEF M. PENNINGER, J. I.-S., TAKEHIKO SASAKI AND ANTONIO J. OLIVEIRA-DOS-SANTOS 2001. CD45: new jobs for an old acquaintance. *Nature Immunology*, 2.
- KALOF, A. N., EVANS, M. F., DANTEY, K. & COOPER, K. 2022. Special Diagnostic Techniques in Surgical Pathology. *Gattuso's Differential Diagnosis in Surgical Pathology*.
- KANEHISA, M., SATO, Y., KAWASHIMA, M., FURUMICHI, M. & TANABE, M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*, 44, D457-62.
- KAZUHIRO ITO, P. D., MISAKO ITO, B.A., W. MARK ELLIOTT, PH.D., BORJA COSIO, M.D., GAETANO CARAMORI, P. D., ONN MIN KON, M.D., ADAM BARCZYK, M.D., SHIZU HAYASHI, P. D., IAN M. ADCOCK, PH.D., JAMES C. HOGG, M.D., & AND PETER J. BARNES, D. M., D.SC 2005. Decreased Histone Deacetylase Activity in Chronic Obstructive Pulmonary Disease. *The new england journal of medicine*, 352.

- KELSEY, H. H. N. A. K. T. 2002. The molecular epidemiology of asbestos and tobacco in lung cancer. *Oncogene*, 21, 7284 - 7288.
- KENT, L. N. & LEONE, G. 2019. The broken cycle: E2F dysfunction in cancer. *Nat Rev Cancer*, 19, 326-338.
- KERR, E. M., GAUDE, E., TURRELL, F. K., FREZZA, C. & MARTINS, C. P. 2016. Mutant Kras copy number defines metabolic reprogramming and therapeutic susceptibilities. *Nature*, 531, 110-3.
- KERR, E. M. & MARTINS, C. P. 2018. Metabolic rewiring in mutant Kras lung cancer. *FEBS J*, 285, 28-41.
- KHEDKAR, G. D., PRAKASH, B., KHEDKAR, C. D. & CHOPADE, B. A. 2016. Nucleic Acids. *Encyclopedia of Food and Health*.
- KIM, C. F. 2017. Intersections of lung progenitor cells, lung disease and lung cancer. *Eur Respir Rev*, 26.
- KIM, C. F., JACKSON, E. L., WOOLFENDEN, A. E., LAWRENCE, S., BABAR, I., VOGEL, S., CROWLEY, D., BRONSON, R. T. & JACKS, T. 2005. Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell*, 121, 823-35.
- KIM, H. J., LEE, H. N., JEONG, M. S. & JANG, S. B. 2021. Oncogenic KRAS: Signaling and Drug Resistance. *Cancers (Basel)*, 13.
- KIM, J. & DEBERARDINIS, R. J. 2019. Mechanisms and Implications of Metabolic Heterogeneity in Cancer. *Cell Metab*, 30, 434-446.
- KIM, J., HU, Z., CAI, L., LI, K., CHOI, E., FAUBERT, B., BEZWADA, D., RODRIGUEZ-CANALES, J., VILLALOBOS, P., LIN, Y. F., NI, M., HUFFMAN, K. E., GIRARD, L., BYERS, L. A., UNSAL-KACMAZ, K., PENA, C. G., HEYMACH, J. V., WAUTERS, E., VANSTEENKISTE, J., CASTRILLON, D. H., CHEN, B. P. C., WISTUBA, I., LAMBRECHTS, D., XU, J., MINNA, J. D. & DEBERARDINIS, R. J. 2017. CPS1 maintains pyrimidine pools and DNA synthesis in KRAS/LKB1-mutant lung cancer cells. *Nature*, 546, 168-172.
- KIM, M., MUN, H., SUNG, C. O., CHO, E. J., JEON, H. J., CHUN, S. M., JUNG, D. J., SHIN, T. H., JEONG, G. S., KIM, D. K., CHOI, E. K., JEONG, S. Y., TAYLOR, A. M., JAIN, S., MEYERSON, M. & JANG, S. J. 2019. Patient-derived lung cancer organoids as in vitro cancer models for therapeutic screening. *Nat Commun*, 10, 3991.
- KISHIMOTO, T. 2005. Interleukin-6: from basic science to medicine--40 years in immunology. *Annu Rev Immunol*, 23, 1-21.
- KLEBE, S., LEIGH, J., HENDERSON, D. W. & NURMINEN, M. 2019. Asbestos, Smoking and Lung Cancer: An Update. *Int J Environ Res Public Health*, 17.
- KRYUKOV, G. V. 2016. MTAP deletion confers enhanced dependency on the PRMT5 arginine methyltransferase in cancer cells. *Science*, 351.

- KUNG, H. N., MARKS, J. R. & CHI, J. T. 2011. Glutamine synthetase is a genetic determinant of cell type-specific glutamine independence in breast epithelia. *PLoS Genet*, 7, e1002229.
- KWON, M. C. & BERNS, A. 2013. Mouse models for lung cancer. *Mol Oncol*, 7, 165-77.
- KYUNG U. HONG, S. D. R., SIMON WATKINS, ELAINE FUCHS, AND BARRY R. STRIPP 2003. In vivo differentiation potential of tracheal basal cells: evidence for multipotent and unipotent subpopulations. *The American Journal of Physiology-Lung Cellular and Molecular Physiology*, 286.
- LABABEDE, O. & MEZIANE, M. A. 2018. The Eighth Edition of TNM Staging of Lung Cancer: Reference Chart and Diagrams. *Oncologist*, 23, 844-848.
- LABUSCHAGNE, C. F., ZANI, F. & VOUSDEN, K. H. 2018. Control of metabolism by p53 - Cancer and beyond. *Biochim Biophys Acta Rev Cancer*, 1870, 32-42.
- LANCASTER, M. A. & KNOBLICH, J. A. 2014. Organogenesis in a dish: modeling development and disease using organoid technologies. *Science*, 345, 1247125.
- LANE, A. N. & FAN, T. W. 2015. Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic Acids Res*, 43, 2466-85.
- LAU, S. C. M., PAN, Y., VELCHETI, V. & WONG, K. K. 2022. Squamous cell lung cancer: Current landscape and future therapeutic options. *Cancer Cell*, 40, 1279-1293.
- LEE, J. H., BHANG, D. H., BEEDE, A., HUANG, T. L., STRIPP, B. R., BLOCH, K. D., WAGERS, A. J., TSENG, Y. H., RYEOM, S. & KIM, C. F. 2014. Lung stem cell differentiation in mice directed by endothelial cells via a BMP4-NFATc1-thrombospondin-1 axis. *Cell*, 156, 440-55.
- LEE, M., HARLEY, G., KATERELOS, M., GLEICH, K., SULLIVAN, M. A., LASKOWSKI, A., COUGHLAN, M., FRASER, S. A., MOUNT, P. F. & POWER, D. A. 2020a. Mutation of regulatory phosphorylation sites in PFKFB2 worsens renal fibrosis. *Sci Rep*, 10, 14531.
- LEE, P., CHANDEL, N. S. & SIMON, M. C. 2020b. Cellular adaptation to hypoxia through hypoxia inducible factors and beyond. *Nat Rev Mol Cell Biol*, 21, 268-283.
- LEEMAN, K. T., PESSINA, P., LEE, J. H. & KIM, C. F. 2019. Mesenchymal Stem Cells Increase Alveolar Differentiation in Lung Progenitor Organoid Cultures. *Sci Rep*, 9, 6479.
- LEISA JOHNSON, K. M., DORON GREENBAUM, RODERICK T. BRONSONK, DENISE CROWLEY, DAVID A. TUVESON AND TYLER JACKS 2001. Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. *Nature*, 410.

- LERTKIATMONGKOL, P., LIAO, D., MEI, H., HU, Y. & NEWMAN, P. J. 2016. Endothelial functions of platelet/endothelial cell adhesion molecule-1 (CD31). *Curr Opin Hematol*, 23, 253-9.
- LEWIS, D. Y., BOREN, J., SHAW, G. L., BIELIK, R., RAMOS-MONTOYA, A., LARKIN, T. J., MARTINS, C. P., NEAL, D. E., SOLOVIEV, D. & BRINDLE, K. M. 2014. Late Imaging with [1-(11)C]Acetate Improves Detection of Tumor Fatty Acid Synthesis with PET. *J Nucl Med*, 55, 1144-9.
- LI, S., WANG, Y., ZHANG, Y., LU, M. M., DEMAYO, F. J., DEKKER, J. D., TUCKER, P. W. & MORRISEY, E. E. 2012. Foxp1/4 control epithelial cell fate during lung development and regeneration through regulation of anterior gradient 2. *Development*, 139, 2500-9.
- LI, X. B., GU, J. D. & ZHOU, Q. H. 2015. Review of aerobic glycolysis and its key enzymes - new targets for lung cancer therapy. *Thorac Cancer*, 6, 17-24.
- LIBERZON, A., BIRGER, C., THORVALDSDOTTIR, H., GHANDI, M., MESIROV, J. P. & TAMAYO, P. 2015. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*, 1, 417-425.
- LIBERZON, A., SUBRAMANIAN, A., PINCHBACK, R., THORVALDSDOTTIR, H., TAMAYO, P. & MESIROV, J. P. 2011. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, 27, 1739-40.
- LOMMEL, A. V. 2001. Pulmonary neuroendocrine cells (PNEC) and neuroepithelial bodies (NEB): chemoreceptors and regulators of lung development. *PAEDIATRIC RESPIRATORY REVIEWS*, 2, 171-176.
- LORI C. GOWEN, B. L. J., ANNE M. LATOUR, KATHLEEN K. SULIK, BEVERLY H. KOLLER 1996. Brea 1 deficiency results in early embryonic lethality characterized by neuroepithelial abnormalities. *Nature genetics*.
- LOUIE, S. M., MOYE, A. L., WONG, I. G., LU, E., SHEHAJ, A., GARCIA-DE-ALBA, C., ARARAT, E., RABY, B. A., LU, B., PASCHINI, M., BRONSON, R. T. & KIM, C. F. 2022. Progenitor potential of lung epithelial organoid cells in a transplantation model. *Cell Rep*, 39, 110662.
- MA, C., WEI, S. & SONG, Y. 2011. T790M and acquired resistance of EGFR TKI: a literature review of clinical reports. *J Thorac Dis*, 3, 10-8.
- MANABU SODAA, S. T., KENGO TAKEUCHIC, YOUNG LIM CHOIA, MUNEHIRO ENOMOTOA, TOSHIHIDE UENOA, & HIDENORI HARUTAA, T. H., YOSHIHIRO YAMASHITAA, YUICHI ISHIKAWAC, YUKIHIKO SUGIYAMAB, AND HIROYUKI MANOA 208. A mouse model for EML4-ALK-positive lung cancer. *PNAS*, 105.
- MANAFI-FARID, R., KARAMZADE-ZIARATI, N., VALI, R., MOTTAGHY, F. M. & BEHESHTI, M. 2021. 2-[(18)F]FDG PET/CT radiomics in lung cancer: An overview of the technical aspect and its emerging role in management of the disease. *Methods*, 188, 84-97.

- MANCHESTER, M. & ANAND, A. 2017. Metabolomics: Strategies to Define the Role of Metabolism in Virus Infection and Pathogenesis. *Adv Virus Res*, 98, 57-81.
- MANTOVANI, F., COLLAVIN, L. & DEL SAL, G. 2019. Mutant p53 as a guardian of the cancer cell. *Cell Death Differ*, 26, 199-212.
- MATTHEW G. VANDER HEIDEN, L. C. C., CRAIG B. THOMPSON 2009. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science*.
- MAYERS, J. R. 2016. Tissue of origin dictates branched-chain amino acid metabolism in mutant Kras-driven cancers. *Science*, 353.
- MCCMAHON, J. R. A. A. 1999. ‘‘Cre’’-ating mouse mutants—a meeting review on conditional mouse genetics. *Genes and Development*, 13.
- MCQUALTER, J. L., YUEN, K., WILLIAMS, B. & BERTONCELLO, I. 2010. Evidence of an epithelial stem/progenitor cell hierarchy in the adult mouse lung. *Proc Natl Acad Sci U S A*, 107, 1414-9.
- MIAO, R., FANG, X., WEI, J., WU, H., WANG, X. & TIAN, J. 2022. Akt: A Potential Drug Target for Metabolic Syndrome. *Front Physiol*, 13, 822333.
- 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, 518-540.
- MILLER, I., MIN, M., YANG, C., TIAN, C., GOOKIN, S., CARTER, D. & SPENCER, S. L. 2018. Ki67 is a Graded Rather than a Binary Marker of Proliferation versus Quiescence. *Cell Rep*, 24, 1105-1112 e5.
- MINNITI, G., GOLDSMITH, C. & BRADA, M. 2012. Radiotherapy. *Handb Clin Neurol*, 104, 215-28.
- MORIMOTO, M., NISHINAKAMURA, R., SAGA, Y. & KOPAN, R. 2012. Different assemblies of Notch receptors coordinate the distribution of the major bronchial Clara, ciliated and neuroendocrine cells. *Development*, 139, 4365-73.
- MUKHERJEE, P., ROY, S., GHOSH, D. & NANDI, S. K. 2022. Role of animal models in biomedical research: a review. *Lab Anim Res*, 38, 18.
- MUZUMDAR, M. D., TASIC, B., MIYAMICHI, K., LI, L. & LUO, L. 2007. A global double-fluorescent Cre reporter mouse. *Genesis*, 45, 593-605.
- NAGAMANI, S. C. & EREZ, A. 2016. A metabolic link between the urea cycle and cancer cell proliferation. *Mol Cell Oncol*, 3, e1127314.
- NAGY, A. 2000. Cre Recombinase: The Universal Reagent for Genome Tailoring. *Genesis*, 26.

- NAKANO, H., NAKANO, K. & COOK, D. N. 2018. Isolation and Purification of Epithelial and Endothelial Cells from Mouse Lung. *Methods Mol Biol*, 1799, 59-69.
- NICHOLSON, A. G., TSAO, M. S., BEASLEY, M. B., BORCZUK, A. C., BRAMBILLA, E., COOPER, W. A., DACIC, S., JAIN, D., KERR, K. M., LANTUEJOL, S., NOGUCHI, M., PAPOTTI, M., REKHTMAN, N., SCAGLIOTTI, G., VAN SCHIL, P., SHOLL, L., YATABE, Y., YOSHIDA, A. & TRAVIS, W. D. 2022. The 2021 WHO Classification of Lung Tumors: Impact of Advances Since 2015. *J Thorac Oncol*, 17, 362-387.
- NIKOLIC, M. Z., SUN, D. & RAWLINS, E. L. 2018. Human lung development: recent progress and new challenges. *Development*, 145.
- NOORELDEEN, R. & BACH, H. 2021. Current and Future Development in Lung Cancer Diagnosis. *Int J Mol Sci*, 22.
- NORDEN, C. 2017. Pseudostratified epithelia - cell biology, diversity and roles in organ formation at a glance. *J Cell Sci*, 130, 1859-1863.
- ONUMA, Y., HIGUCHI, K., AIKI, Y., SHU, Y., ASADA, M., ASASHIMA, M., SUZUKI, M., IMAMURA, T. & ITO, Y. 2015. A stable chimeric fibroblast growth factor (FGF) can successfully replace basic FGF in human pluripotent stem cell culture. *PLoS One*, 10, e0118931.
- OTTO WARBURG, F. W., AND ERWI'N NEGELEIN 1926. THE METABOLISM OF TUMORS IN THE BODY. . *The Journal of General Physiology*.
- PANG, Z., CHONG, J., ZHOU, G., DE LIMA MORAIS, D. A., CHANG, L., BARRETTE, M., GAUTHIER, C., JACQUES, P. E., LI, S. & XIA, J. 2021. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res*, 49, W388-W396.
- PARVIZ MINOO, G. S., HONG DRUM, & PABLO BRINGAS, A. S. K. 1999. Defects in Tracheoesophageal and Lung Morphogenesis in Nkx2.1(2/2) Mouse Embryos. *Developmental Biology*, 209.
- PAUL, M. K., BISHT, B., DARMAWAN, D. O., CHIOU, R., HA, V. L., WALLACE, W. D., CHON, A. T., HEGAB, A. E., GROGAN, T., ELASHOFF, D. A., ALVA-ORNELAS, J. A. & GOMPERS, B. N. 2014. Dynamic changes in intracellular ROS levels regulate airway basal stem cell homeostasis through Nrf2-dependent Notch signaling. *Cell Stem Cell*, 15, 199-214.
- PEROU, C. M. 2000. Molecular portraits of human breast tumours. *Nature*, 406.
- R HAKEM, J. L. D. L. P., C SIRARD, R MO, M WOO, A HAKEM, A WAKEHAM, J POTTER, A REITMAIR, F BILLIA, E FIRPO, C C HUI, J ROBERTS, J ROSSANT, T W MAK 1996. The Tumor Suppressor Gene Brca1 Is Required for Embryonic Cellular Proliferation in the Mouse. *Cell*.
- RABINOWITZ, J. D. & ENERBACK, S. 2020. Lactate: the ugly duckling of energy metabolism. *Nat Metab*, 2, 566-571.

- RAJORIA, B., ZHANG, X. & YEE, D. 2023. IGF-1 Stimulates Glycolytic ATP Production in MCF-7L Cells. *Int J Mol Sci*, 24.
- RALPH J. DEBERARDINIS, A. M., EVGUENI DAIKHIN, ILANA NISSIM, MARC YUDKOFF, SUZANNE WEHRLI, & THOMPSON, A. C. B. 2007. Beyond aerobic glycolysis: Transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *PNAS*.
- RALPH MEUWISSEN, S. C. L., R. ILONA LINNOILA, JOHN ZEVENHOVEN, WOLTER J. MOOI, AND ANTON BERNIS 2003. Induction of small cell lung cancer by somatic inactivation of both Trp53 and Rb1 in a conditional mouse model. *Cancer Cell*, 4.
- RAMI-PORTA, R., CALL, S., DOOMS, C., OBIOLS, C., SANCHEZ, M., TRAVIS, W. D. & VOLLMER, I. 2018. Lung cancer staging: a concise update. *Eur Respir J*, 51.
- RAWLINS, E. L., OKUBO, T., XUE, Y., BRASS, D. M., AUTEN, R. L., HASEGAWA, H., WANG, F. & HOGAN, B. L. 2009. The role of Scgb1a1+ Clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. *Cell Stem Cell*, 4, 525-34.
- REINFELD, B. I., MADDEN, M. Z., WOLF, M. M., CHYTIL, A., BADER, J. E., PATTERSON, A. R., SUGIURA, A., COHEN, A. S., ALI, A., DO, B. T., MUIR, A., LEWIS, C. A., HONGO, R. A., YOUNG, K. L., BROWN, R. E., TODD, V. M., HUFFSTATER, T., ABRAHAM, A., O'NEIL, R. T., WILSON, M. H., XIN, F., TANTAWY, M. N., MERRYMAN, W. D., JOHNSON, R. W., WILLIAMS, C. S., MASON, E. F., MASON, F. M., BECKERMANN, K. E., VANDER HEIDEN, M. G., MANNING, H. C., RATHMELL, J. C. & RATHMELL, W. K. 2021. Cell-programmed nutrient partitioning in the tumour microenvironment. *Nature*, 593, 282-288.
- REKHTMAN, N., ANG, D. C., SIMA, C. S., TRAVIS, W. D. & MOREIRA, A. L. 2011. Immunohistochemical algorithm for differentiation of lung adenocarcinoma and squamous cell carcinoma based on large series of whole-tissue sections with validation in small specimens. *Mod Pathol*, 24, 1348-59.
- RITA COSTA, A., ELISA RODRIGUES, M., HENRIQUES, M., AZEREDO, J. & OLIVEIRA, R. 2010. Guidelines to cell engineering for monoclonal antibody production. *Eur J Pharm Biopharm*, 74, 127-38.
- ROCK, J. R., BARKAUSKAS, C. E., CRONCE, M. J., XUE, Y., HARRIS, J. R., LIANG, J., NOBLE, P. W. & HOGAN, B. L. 2011a. Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. *Proc Natl Acad Sci U S A*, 108, E1475-83.
- ROCK, J. R., GAO, X., XUE, Y., RANDELL, S. H., KONG, Y. Y. & HOGAN, B. L. 2011b. Notch-dependent differentiation of adult airway basal stem cells. *Cell Stem Cell*, 8, 639-48.

- ROCK, J. R. & HOGAN, B. L. 2011. Epithelial progenitor cells in lung development, maintenance, repair, and disease. *Annu Rev Cell Dev Biol*, 27, 493-512.
- ROMERO, R., SAYIN, V. I., DAVIDSON, S. M., BAUER, M. R., SINGH, S. X., LEBOEUF, S. E., KARAKOUSI, T. R., ELLIS, D. C., BHUTKAR, A., SANCHEZ-RIVERA, F. J., SUBBARAJ, L., MARTINEZ, B., BRONSON, R. T., PRIGGE, J. R., SCHMIDT, E. E., THOMAS, C. J., GOPARAJU, C., DAVIES, A., DOLGALEV, I., HEGUY, A., ALLAJ, V., POIRIER, J. T., MOREIRA, A. L., RUDIN, C. M., PASS, H. I., VANDER HEIDEN, M. G., JACKS, T. & PAPAGIANNAKOPOULOS, T. 2017. Keap1 loss promotes Kras-driven lung cancer and results in dependence on glutaminolysis. *Nat Med*, 23, 1362-1368.
- RUDIN, C. M., BRAMBILLA, E., FAIVRE-FINN, C. & SAGE, J. 2021. Small-cell lung cancer. *Nat Rev Dis Primers*, 7, 3.
- SCHAFFER, B. E., PARK, K. S., YIU, G., CONKLIN, J. F., LIN, C., BURKHART, D. L., KARNEZIS, A. N., SWEET-CORDERO, E. A. & SAGE, J. 2010. Loss of p130 accelerates tumor development in a mouse model for human small-cell lung carcinoma. *Cancer Res*, 70, 3877-83.
- SCHLABACH, M. R., HU, J. K., LI, M. & ELLEDGE, S. J. 2010. Synthetic design of strong promoters. *Proc Natl Acad Sci U S A*, 107, 2538-43.
- SCHÖNTHAL, A. H. 2004. *Checkpoint Controls and Cancer Volume 1 Reviews and Model Systems*, Humana Press.
- SCOTLAND, P. H. 2023. *Public Health Scotland - Lung cancer* [Online]. [Accessed 26th September 2023].
- SEMENZA, G. L. 2013. HIF-1 mediates metabolic responses to intratumoral hypoxia and oncogenic mutations. *J Clin Invest*, 123, 3664-71.
- SHAHBAZI, M., CUNDIFF, P., ZHOU, W., LEE, P., PATEL, A., D'SOUZA, S. L., ABBASI, F., QUERTERMOUS, T. & KNOWLES, J. W. 2019. The role of insulin as a key regulator of seeding, proliferation, and mRNA transcription of human pluripotent stem cells. *Stem Cell Res Ther*, 10, 228.
- SHEN, G. M., ZHANG, F. L., LIU, X. L. & ZHANG, J. W. 2010. Hypoxia-inducible factor 1-mediated regulation of PPP1R3C promotes glycogen accumulation in human MCF-7 cells under hypoxia. *FEBS Lett*, 584, 4366-72.
- SHREYA RAGHAVAN , P. M., ERIC N. HORST 2 , MARIA R. WARD 1 , KATELYN R., ROWLEY , GEETA MEHTA 2016. Comparative analysis of tumor spheroid generation techniques for differential in vitro drug toxicity. *Oncotarget*.
- SILVIA FERNA;NDEZ DE MATTOS, E. W. F. L. A. A. T. 2002. An E2F-binding site mediates the activation of the proliferative isoform of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase by phosphatidylinositol 3-kinase. *Biochemical Journal*.

- SITLAULA, S. & BURRIS, T. P. 2016. Cholesterol and Other Steroids. *Encyclopedia of Cell Biology*.
- SKERRETT-BYRNE ANTHONY, D., JIANG CHEN, C., NIXON, B. & HONDERMARCK, H. 2023. Transcriptomics. *Encyclopedia of Cell Biology*.
- SUTHERLAND, K. D., PROOST, N., BROUNS, I., ADRIAENSEN, D., SONG, J. Y. & BERNIS, A. 2011. Cell of origin of small cell lung cancer: inactivation of Trp53 and Rb1 in distinct cell types of adult mouse lung. *Cancer Cell*, 19, 754-64.
- SUTHERLAND, K. D., SONG, J. Y., KWON, M. C., PROOST, N., ZEVENHOVEN, J. & BERNIS, A. 2014. Multiple cells-of-origin of mutant K-Ras-induced mouse lung adenocarcinoma. *Proc Natl Acad Sci U S A*, 111, 4952-7.
- SWANTON, C. & GOVINDAN, R. 2016. Clinical Implications of Genomic Discoveries in Lung Cancer. *N Engl J Med*, 374, 1864-73.
- TAMOUTOUNOUR, S., HAN, S. J., DECKERS, J., CONSTANTINIDES, M. G., HURABIELLE, C., HARRISON, O. J., BOULADOUX, N., LINEHAN, J. L., LINK, V. M., VUJKOVIC-CVIJIN, I., PEREZ-CHAPARRO, P. J., ROSSHART, S. P., REHERMANN, B., LAZAREVIC, V. & BELKAID, Y. 2019. Keratinocyte-intrinsic MHCII expression controls microbiota-induced Th1 cell responses. *Proc Natl Acad Sci U S A*, 116, 23643-23652.
- TELO, S., CALDERONI, L., VICHI, S., ZAGNI, F., CASTELLUCCI, P. & FANTI, S. 2020. Alternative and New Radiopharmaceutical Agents for Lung Cancer. *Curr Radiopharm*, 13, 185-194.
- TEMPLE, J., VELLIU, E., SHEHATA, M. & LEVY, R. 2022. Current strategies with implementation of three-dimensional cell culture: the challenge of quantification. *Interface Focus*, 12, 20220019.
- THAI, A. A., SOLOMON, B. J., SEQUIST, L. V., GAINOR, J. F. & HEIST, R. S. 2021. Lung cancer. *Lancet*, 398, 535-554.
- TIAN, T., LI, X. & ZHANG, J. 2019. mTOR Signaling in Cancer and mTOR Inhibitors in Solid Tumor Targeting Therapy. *Int J Mol Sci*, 20.
- TIMOTHY A. STEWART, P. K. P., AND PHILIP LEDER 1984. Spontaneous Mammary Adenocarcinomas in Transgenic Mice That Carry and Express MTV/myc Fusion Genes. *Cell*, 38.
- TRAVIS, W. D., BRAMBILLA, E., NICHOLSON, A. G., YATABE, Y., AUSTIN, J. H. M., BEASLEY, M. B., CHIRIEAC, L. R., DACIC, S., DUHIG, E., FLIEDER, D. B., GEISINGER, K., HIRSCH, F. R., ISHIKAWA, Y., KERR, K. M., NOGUCHI, M., PELOSI, G., POWELL, C. A., TSAO, M. S., WISTUBA, I. & PANEL, W. H. O. 2015. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol*, 10, 1243-1260.

- TRZPIS, M., MCLAUGHLIN, P. M., DE LEIJ, L. M. & HARMSSEN, M. C. 2007. Epithelial cell adhesion molecule: more than a carcinoma marker and adhesion molecule. *Am J Pathol*, 171, 386-95.
- TSAO, P. N., VASCONCELOS, M., IZVOLSKY, K. I., QIAN, J., LU, J. & CARDOSO, W. V. 2009. Notch signaling controls the balance of ciliated and secretory cell fates in developing airways. *Development*, 136, 2297-307.
- TYLER JACKS, A. F., EARLENE M. SCHMITT, RODERICK T. BRONSON, MARGARET A. GOODELL & ROBERT A. WEINBERG 1992. Effects of an Rb mutation in the mouse. *Nature*.
- VANDER HEIDEN, M. G. & DEBERARDINIS, R. J. 2017. Understanding the Intersections between Metabolism and Cancer Biology. *Cell*, 168, 657-669.
- VISVADER, J. E. 2011. Cells of origin in cancer. *Nature*, 469, 314-22.
- WAGNER, V. & GIL, J. 2020. Senescence as a therapeutically relevant response to CDK4/6 inhibitors. *Oncogene*, 39, 5165-5176.
- WALKER, L. S. & SANSOM, D. M. 2011. The emerging role of CTLA4 as a cell-extrinsic regulator of T cell responses. *Nat Rev Immunol*, 11, 852-63.
- WANSLEEBEN, C., BARKAUSKAS, C. E., ROCK, J. R. & HOGAN, B. L. 2013. Stem cells of the adult lung: their development and role in homeostasis, regeneration, and disease. *Wiley Interdiscip Rev Dev Biol*, 2, 131-48.
- WARBURG, O. 1956. On the Origin of Cancer Cells. *Science*, 123.
- WEINBERG, A. R. A. R. A. 2003. Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nature*.
- WERNER, C., DOENST, T. & SCHWARZER, M. 2016. Metabolic Pathways and Cycles. *The Scientist's Guide to Cardiac Metabolism*.
- WEST, J. A., PAKEHHAM, G., MORIN, D., FLESCHNER, C. A., BUCKPITT, A. R. & PLOPPER, C. G. 2001. Inhaled naphthalene causes dose dependent Clara cell cytotoxicity in mice but not in rats. *Toxicol Appl Pharmacol*, 173, 114-9.
- WHITSETT, J. A., WERT, S. E. & WEAVER, T. E. 2010. Alveolar Surfactant Homeostasis and the Pathogenesis of Pulmonary Disease. *Annual Review of Medicine*, 61, 105-119.
- WHO. *International Agency for Research on Cancer. Global Cancer Observatory: cancer today*. [Online]. Available: <https://gco.iarc.fr/today> [Accessed 2nd September 2023].
- WILLIAM D. TRAVIS, E. B., H. KONRAD MÜLLER-HERMELINK AND CURTIS C. HARRIS 2004. *Pathology & Genetics. Tumours of the Lung, Pleura, Thymus and Heart*, World Health Organization (WHO).

- WILLIAMS, M. C. 2003. Alveolar type I cells: molecular phenotype and development. *Annu Rev Physiol*, 65, 669-95.
- WINSLOW, M. M., DAYTON, T. L., VERHAAK, R. G., KIM-KISELAK, C., SNYDER, E. L., FELDSER, D. M., HUBBARD, D. D., DUPAGE, M. J., WHITTAKER, C. A., HOERSCH, S., YOON, S., CROWLEY, D., BRONSON, R. T., CHIANG, D. Y., MEYERSON, M. & JACKS, T. 2011. Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature*, 473, 101-4.
- WOOLBRIGHT, B. L. & HARRIS, R. A. 2021. PDK2: An Underappreciated Regulator of Liver Metabolism. *Livers*, 1, 82-97.
- WRIGHT, V. C. 2008. Cervical Glandular Disease. *Colposcopy*.
- XIA, J., PSYCHOGIOS, N., YOUNG, N. & WISHART, D. S. 2009. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res*, 37, W652-60.
- XIAOMING LIU, R. R. D. A. J. F. E. 2004. Airway Glandular Development and Stem Cells. *Current Topics in Developmental Biology*, 64, 33-56.
- XING, D. & ORSULIC, S. 2006. A mouse model for the molecular characterization of brca1-associated ovarian carcinoma. *Cancer Res*, 66, 8949-53.
- XU, C., FILLMORE, C. M., KOYAMA, S., WU, H., ZHAO, Y., CHEN, Z., HERTER-SPRIE, G. S., AKBAY, E. A., TCHAICHA, J. H., ALTABEF, A., REIBEL, J. B., WALTON, Z., JI, H., WATANABE, H., JANNE, P. A., CASTRILLON, D. H., RUSTGI, A. K., BASS, A. J., FREEMAN, G. J., PADERA, R. F., DRANOFF, G., HAMMERMAN, P. S., KIM, C. F. & WONG, K. K. 2014. Loss of Lkb1 and Pten leads to lung squamous cell carcinoma with elevated PD-L1 expression. *Cancer Cell*, 25, 590-604.
- XU, H., YU, S., LIU, Q., YUAN, X., MANI, S., PESTELL, R. G. & WU, K. 2017. Recent advances of highly selective CDK4/6 inhibitors in breast cancer. *J Hematol Oncol*, 10, 97.
- XU, Z. Z., LI, Z. J., DU, L. X., LI, J. & WANG, L. Y. 2013. Using bovine pituitary extract to increase proliferation of keratocytes and maintain their phenotype in vitro. *Int J Ophthalmol*, 6, 758-65.
- YANG, C., GONZALEZ, I., ZHANG, L. & CAO, D. 2019. Neuroendocrine Markers Insulinoma-Associated Protein 1, Chromogranin, Synaptophysin, and CD56 Show Rare Positivity in Adenocarcinoma Ex-Goblet Cell Carcinoids. *Gastroenterology Res*, 12, 120-127.
- YATES, L. R. & CAMPBELL, P. J. 2012. Evolution of the cancer genome. *Nat Rev Genet*, 13, 795-806.
- YIN, Z., GONZALES, L., KOLLA, V., RATH, N., ZHANG, Y., LU, M. M., KIMURA, S., BALLARD, P. L., BEERS, M. F., EPSTEIN, J. A. & MORRISEY, E. E. 2006. Hop functions downstream of Nkx2.1 and GATA6 to mediate HDAC-dependent negative regulation of pulmonary gene expression. *Am J Physiol Lung Cell Mol Physiol*, 291, L191-9.

- YOO, H. C., YU, Y. C., SUNG, Y. & HAN, J. M. 2020. Glutamine reliance in cell metabolism. *Exp Mol Med*, 52, 1496-1516.
- YU, M., BOJIC, S., FIGUEIREDO, G. S., ROONEY, P., DE HAVILLAND, J., DICKINSON, A., FIGUEIREDO, F. C. & LAKO, M. 2016. An important role for adenine, cholera toxin, hydrocortisone and triiodothyronine in the proliferation, self-renewal and differentiation of limbal stem cells in vitro. *Exp Eye Res*, 152, 113-122.
- YUNEVA, M. O., FAN, T. W., ALLEN, T. D., HIGASHI, R. M., FERRARIS, D. V., TSUKAMOTO, T., MATES, J. M., ALONSO, F. J., WANG, C., SEO, Y., CHEN, X. & BISHOP, J. M. 2012. The metabolic profile of tumors depends on both the responsible genetic lesion and tissue type. *Cell Metab*, 15, 157-70.
- ZECHA, J., SATPATHY, S., KANASHOVA, T., AVANESSIAN, S. C., KANE, M. H., CLAUSER, K. R., MERTINS, P., CARR, S. A. & KUSTER, B. 2019. TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. *Mol Cell Proteomics*, 18, 1468-1478.
- ZHANG, H., FILLMORE BRAINSON, C., KOYAMA, S., REDIG, A. J., CHEN, T., LI, S., GUPTA, M., GARCIA-DE-ALBA, C., PASCHINI, M., HERTER-SPRIE, G. S., LU, G., ZHANG, X., MARSH, B. P., TUMINELLO, S. J., XU, C., CHEN, Z., WANG, X., AKBAY, E. A., ZHENG, M., PALAKURTHI, S., SHOLL, L. M., RUSTGI, A. K., KWIATKOWSKI, D. J., DIEHL, J. A., BASS, A. J., SHARPLESS, N. E., DRANOFF, G., HAMMERMAN, P. S., JI, H., BARDEESY, N., SAUR, D., WATANABE, H., KIM, C. F. & WONG, K. K. 2017. Lkb1 inactivation drives lung cancer lineage switching governed by Polycomb Repressive Complex 2. *Nat Commun*, 8, 14922.
- ZHANG, L., YU, H., YUAN, Y., YU, J. S., LOU, Z., XUE, Y. & LIU, Y. 2020. The necessity for standardization of glioma stem cell culture: a systematic review. *Stem Cell Res Ther*, 11, 84.
- ZHANG, W. C., SHYH-CHANG, N., YANG, H., RAI, A., UMASHANKAR, S., MA, S., SOH, B. S., SUN, L. L., TAI, B. C., NGA, M. E., BHAKOO, K. K., JAYAPAL, S. R., NICHANE, M., YU, Q., AHMED, D. A., TAN, C., SING, W. P., TAM, J., THIRUGANANAM, A., NOGHABI, M. S., PANG, Y. H., ANG, H. S., MITCHELL, W., ROBSON, P., KALDIS, P., SOO, R. A., SWARUP, S., LIM, E. H. & LIM, B. 2012. Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. *Cell*, 148, 259-72.
- ZHANG, Y., GOSS, A. M., COHEN, E. D., KADZIK, R., LEPORE, J. J., MUTHUKUMARASWAMY, K., YANG, J., DEMAYO, F. J., WHITSETT, J. A., PARMACEK, M. S. & MORRISEY, E. E. 2008. A Gata6-Wnt pathway required for epithelial stem cell development and airway regeneration. *Nat Genet*, 40, 862-70.
- ZHAO, J. Y., FENG, K. R., WANG, F., ZHANG, J. W., CHENG, J. F., LIN, G. Q., GAO, D. & TIAN, P. 2021a. A retrospective overview of PHGDH and its inhibitors for regulating cancer metabolism. *Eur J Med Chem*, 217, 113379.

- ZHAO, Q., CHENG, Y. & XIONG, Y. 2021b. LTF Regulates the Immune Microenvironment of Prostate Cancer Through JAK/STAT3 Pathway. *Front Oncol*, 11, 692117.
- ZHAOHUI HE, F. T., ZECHAO LU, YUCONG HUANG, HANQI LEI, ZHIBIAO LI, GUOHUA ZENG 2018. Analysis of differentially expressed genes, clinical value and biological pathways in prostate cancer. *American Journal of Translational Research*
- ZHONG WANG, M. G. A. M. S. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *nature genetics*.
- ZHOU, Y., YANG, Y., GUO, L., QIAN, J., GE, J., SINNER, D., DING, H., CALIFANO, A. & CARDOSO, W. V. 2022. Airway basal cells show regionally distinct potential to undergo metaplastic differentiation. *Elife*, 11.
- ZOIS, C. E., FAVARO, E. & HARRIS, A. L. 2014. Glycogen metabolism in cancer. *Biochem Pharmacol*, 92, 3-11.