UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Elucidating the roles and therapeutic targeting of molecular drivers of chronic obstructive pulmonary disease (COPD)/emphysema

by Tayyaba Sadaf

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the primary supervision of Prof Philip M. Hansbro
co-supervision of Dr Alen Faiz
Dr Nicole Gower Hansbro
Prof Garry Stewart Anthony Myers
A/Prof Kamal Dua

University of Technology Sydney
Faculty of Science

February 2023

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Tayyaba Sadaf, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Life Sciences, Faculty of Science at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed prior to publication.

Signature: *Tayyaba Sadaf*

Date: 5th Feb 2023

Acknowledgment

First and foremost, I extend my deepest gratitude to my supervisor, ***Professor Philip M. Hansbro***, for accepting me as a Ph.D. student and affording me the chance to advance my career in Australia. I am profoundly grateful for his unwavering support and guidance, without which the successful completion of my candidature would not have been attainable. I am indebted to him for facilitating the expansion of my scientific knowledge and for welcoming me into his esteemed research group, which stands as one of the preeminent Respiratory research groups in Australia.

I extend my sincere appreciation to my co-supervisor, ***Dr. Alen Faiz,*** for his consistent guidance, support, and mentorship throughout my Ph.D. endeavors. ***Dr. Faiz*** has exemplified exceptional mentorship, providing invaluable encouragement and imparting the skills necessary to navigate and address the most daunting challenges. I am profoundly grateful for the privilege of being under his supervision and guidance and consider myself exceedingly fortunate to have benefitted from his expertise.

I would like to acknowledge ***Dr. Nicole Hansbro*** for her invaluable assistance with administrative matters, ensuring the seamless operation of my Ph.D. endeavors. Her diligent support in all facets of logistical management, from procurement to the coordination of experimental procedures, has been instrumental in facilitating the smooth progression of my research activities.

I extend my sincere appreciation to our collaborators Dr. Andrew Deane, Dr. Emma L. Beckett, Andrew Jarnicki, and Dr. Richard Y. Kim from Hunter Medical Research Institute and The University of Newcastle, New South Wales, Australia for their contributions to generating and sharing the cigarette smoke-induced COPD time course gene expression and miRNA

expression array datasets. Additionally, I am deeply grateful to Dr. David A. Skerrett-Byrne for providing the proteomics data related to cigarette smoke-induced experimental chronic obstructive pulmonary disease for inclusion in this study. Furthermore, my gratitude extends to Dr. Caitlin Gillis for sharing samples from the time course experimental COPD model, which proved instrumental in validating our findings.

I would also like to thank our collaborators from the Charles Perkin Centre (CPC) at the University of Sydney, specifically Prof. Greg Neely, Felicity Chung, and Zina Hamoudi, for their invaluable assistance, guidance, and provision of necessary reagents to facilitate the execution of the CRISPR screen.

Furthermore, I wish to express gratitude to all co-supervisors, collaborators, as well as past and present members of the Centre for Inflammation (CFI) and the Respiratory Bioinformatics and Molecular Biology (RBMB) group.

I wish to extend my heartfelt gratitude to Nisha Panth, my esteemed colleague and dear friend, and Dr. Keshav Raj Paudel, for their exceptional level of support, particularly during the concluding stages of my Ph.D. program. Words cannot adequately convey the depth of appreciation I hold for the support and encouragement they both have generously provided.

I am also grateful to the University of Technology Sydney and the Australian Government for their support through the Australian Government Research Training Program (RTP) Scholarship, the Faculty of Science Scholarship - Science Associate Dean Research Scholarship, as well as the University of Technology Sydney - International Research Scholarship (IRS), to pursue my Ph.D. at one of the world's leading universities. Aditionally, I am also thankful to the School of Life Sciences and the Faculty of Science for their invaluable administrative support throughout my academic journey.

Personal

I am extremely thankful to my beloved parents, ***Mr. Muhammad Sadiq Bhatti***, whom I hold in the highest esteem, and ***Mrs. Shamim Sadiq***, for their unwavering encouragement and support throughout my educational journey. I am thankful for their consistent motivation and belief in my potential to achieve higher aspirations. Additionally, I extend my appreciation to my elder brothers, Bilal Sssadiq and Muhammad Afzaal Sadiq, for their exceptional understanding, unconditional love, and guidance. I am extremely thankful to them for their endless love and interminable support throughout my life.

A very special thanks to the most significant and cherished individuals in my life, who hold a special place in my heart: my beloved nieces, Tanzeel Sadaf and Irha Bilal. Their boundless love and heartfelt prayers have been a source of immense comfort and joy.

I wish to express my sincere appreciation to my dear friends, Sidrah Anjum and Amna Umer Khayam, for their wonderful companionship. I consider myself fortunate to have such loyal and dependable friends.

Oral presentations based on thesis work

1. **Sadaf, T.,** et al. Pooled CRISPR screening reveals protective factors against cigarette smoke exposure. TSANZ NSW/ACT Branch Annual Scientific Meeting & RNIG Thoracic Clinical Update 2023, Sydney, Australia, November 2023.
2. **Sadaf, T.,** et al. Whole-genome CRISPR loss-of-function screen to identify protective factors against cigarette smoke. NSW Airway Meeting 17 (NAMe17) meeting, Sydney, Australia, October 2023.
3. **Sadaf, T.,** et al. Genome-wide CRISPR/Cas9 Knock-out library screening against cigarette smoke. NAMe16: The 16th Annual NSW Asthma meeting, Sydney, Australia, November 2022.
4. **Sadaf, T.,** et al. Genome-wide CRISPR/Cas9 Knock-out library screening against cigarette smoke. TSANZ NSW/ACT Branch Annual Scientific Meeting & RNIG Thoracic Clinical Update 2022, Sydney, Australia, November 2022.
5. **Sadaf, T., et al.** mRNA expression profiling in lung tissue of Cigarette Smoke induced COPD in mice. TSANZ NSW/ACT Branch Annual Scientific Meeting & RNIG Thoracic Clinical Update 2019, Sydney, Australia, October 2019.

Poster presentations based on thesis work

1. **Sadaf, T.,** et al. Integration of miRNA and gene expression profiles from lung tissues of CS-induced COPD. Sydney Bioinformatics Research Symposium (SBRS) 2023, Sydney, Australia, September 2023.
2. **Sadaf, T.,** et al. Genome-wide CRISPR/Cas9 Knock-out library screening against cigarette smoke. TSANZSRS 2022 The Australia & New Zealand Society of Respiratory Science and The Thoracic Society of Australia and New Zealand

(ANZSRS/TSANZ), Annual Scientific Meeting for Leaders in Lung Health & Respiratory Science 2023, March 2023, Christchurch, New Zealand.

3. **Sadaf, T.,** et al. Gene expression profiling differences in CS-induced experimental COPD. TSANZSRS 2022 The Australia & New Zealand Society of Respiratory Science and The Thoracic Society of Australia and New Zealand (ANZSRS/TSANZ), Annual Scientific Meeting for Leaders in Lung Health & Respiratory Science 2022, March-April, 2022, virtual.
4. **Sadaf, T.,** et al. mRNA expression profiling in lung tissue of Cigarette Smoke induced COPD in mice.36th Combined health science conference Sydney New Horizons 2019, Sydney, Australia, 2019, November 2019.
5. **Sadaf, T.,** et al. mRNA expression profiling in lung tissue of Cigarette Smoke induced COPD in mice. NAMe15 Annual NSW Asthma meeting 2019, Sydney, Australia November 2019.

Awards received as a PhD student

1. TSANZ NSW/ACT Branch Annual Scientific Meeting & RNIG Thoracic Clinical Update, 2023 - **(Best Oral presentation Award- 2nd place).**
2. NSW Airway Meeting 17 (NAMe17) meeting, 2023 – **(Best PhD student presentation - 1st place)**.
3. TSANZ ASM Travel Grant for TSANZSRS 2023: Annual Scientific Meeting from Sydney Australia to Christchurch, New Zealand, **(Travel Grant).**
4. NAMe16: The 16th Annual NSW Asthma meeting, 2022 - **(Best PhD student presentation - 2nd place).**
5. TSANZ NSW/ACT Branch Annual Scientific Meeting & RNIG Thoracic Clinical Update, 2022 - **(Best Oral presentation Award).**

6. NAMe15 Annual NSW Asthma meeting, 2019. **(Best Poster Award).**
7. University of Queensland (UQ) – Institute of Molecular Bioscience 2019 winter school student travel award, 2019. **(Travel Award).**
8. **Australian Government Research Training Program (RTP) Scholarship,** University of Technology Sydney (UTS), 2019-2023.
9. **Faculty of Science Scholarship - Science Associate Dean Research Scholarship** Issued by University of Technology Sydney (UTS), 2019-2022.

   University of Technology Sydney - International Research Scholarship (IRS) Issued by University of Technology Sydney (UTS), 2019.

Table of Content

CERTIFICATE OF ORIGINAL AUTHORSHIP i
Acknowledgment ii
Oral presentations based on thesis work v
Poster presentations based on thesis work v
Awards received as a PhD student vi
Table of Content viii
List of Abbreviations xiii
List of Figures xvi
List of Tables xviii
Abstract xx
Chapter 1: Introduction 1
1.1 COPD epidemiology and Global burden 1
1.2 COPD aetiology 3
1.3 COPD prevention and treatment 5
1.4 Molecular mechanisms of COPD development 6
1.5 Importance of mouse models in COPD research 8
1.6 Transcriptomics and gene expression profiling 11
1.6.1 Microarrays 11
1.6.2 Investigating microRNAs using microarrays 13
1.6.3 RNA sequencing 15
1.7 Integrative omics and its role in understanding COPD 17
1.8 Functional studies through CRISPR/Cas systems 20
1.8.1 CRISPR/Cas system 22
1.8.2 Genome-wide loss of function studies 22
1.8.3 CRISPR knockout libraries 27
1.8.4 Positive and negative screens 28
1.8.5 High-throughput CRISPR screening in the respiratory system 29
1.9 Aims 30
Chapter 2: Longitudinal analysis of genome-wide expression profiling of cigarette smoke-induced experimental COPD 32
2.1 Abstract 32
2.2 Introduction 34
2.3 Methodology 36
2.3.1 *In-vivo* elucidation of gene changes 37

2.3.2 Induction of cigarette smoke-induced COPD model....37
2.3.3 Data generation and processing....37
2.3.4 Correlation with protein profiling....38
2.3.5 Characterization of functionally enriched DE genes....39
2.3.6 Generation of knock-out (KO) cells by CRISPR-Cas9....39
2.3.7 Long-term exposure of *ARNT2* KO cells with CS extract (CSE)....40
2.3.8 Impact of CSE on Cell viability and apoptosis....40
2.3.9 Library preparation and RNA sequencing....40
2.3.10 qPCR validation....41
2.3.11 Statistical analysis....41
2.4 Results....41
2.4.1 Sub-chronic CS exposure alters the gene expression profile....41
2.4.2 Association of gene expression with disease development in CS-induced experimental COPD....42
2.4.3 Altered longitudinal gene signatures define the disease progression phase....43
2.4.4 Gene expression changes in *Cyp1a1* are replicated at the protein level....45
2.4.5 Smoke-protective genes modulate the disease progression....45
2.4.6 Functional effects of *ARNT2* in response to CS....48
2.4.7 Deletion of *ARNT2* in KO bronchial epithelial cells exposed to CSE protects against cell death....49
2.4.8 RNA sequencing of KO cells....49
2.4.9 Functional enrichment analysis of *ARNT2* KO cells....50
2.4.10 qPCR validation of potential hits....52
2.5 Discussion....56
2.5.1 Proposed mechanism....59
Chapter 3: Longitudinal evaluation of whole lung miRNAs and their interactions with gene expression in experimental COPD....62
3.1 Abstract....62
3.2 Introduction....64

3.3 Methodology 66
3.3.1 Short-term CS-induced experimental COPD 67
3.3.2 miRNA and gene expression profiling 67
3.3.3 Data processing and statistical analysis 67
3.3.4 Association of miRNAs with clinical features in a human cohort 68
3.3.5 Canonical correlation of miRNA expression level with gene expression levels 69
3.3.6 Characterization of functionally enriched miRNAs 70
3.3.7 Investigation of significant correlated miRNA-gene expression in an independent experimental COPD model 70
3.3.8 Quantification of miRNA and mRNA expression by real-time qPCR 71
3.3.9 Statistical analysis 71
3.4 Results 72
3.4.1 Longitudinal effects of smoke exposure on miRNA expression 72
3.4.2 The association of differentially expressed miRNAs shows a significant correlation with lung function parameters in human COPD 74
3.4.3 Correlation analysis showed significant correlations between differentially expressed miRNAs and their target genes 74
3.4.4 Identification of direct miRNA targets highlights validated and predicted targets of dysregulated miRNAs in COPD 77
3.4.5 Correlation analyses showed strong correlations between the expression of miRNAs and their directs targets in COPD 77
3.4.6 Functional enrichment analysis of negatively correlated direct targets of miRNAs in experimental COPD 78
3.4.7 Correlation shows *Arnt2*, *Bbox1,* and *Ttc25* may be direct targets of dysregulated miRNA in experimental COPD 82
3.5 Discussion 85
3.6 Proposed Mechanism 89
Chapter 4: Genome-wide CRISPR-Cas9 Knock-out library screening to identify protective factors against cigarette smoke 91
4.1 Abstract 91

4.2 Introduction 93
4.3 Methodology 94
4.3.1 Cell lines 94
4.3.2 Human Toronto knockout CRISPR library version 3 (TKOv3) 95
4.3.3 Genome-scale lentiviral TKOv3 gRNA library construction 95
4.3.4 Determination of lentivirus multiplicity of infection (MOI) for transducing the cell lines 96
4.3.5 Working concentration of puromycin for mammalian cell lines 97
4.3.6 Determination of cell death curve in cell lines against CSE 98
4.3.7 Large-scale pooled CRISPR depletion screening against CSE 99
4.3.8 Genomic DNA extraction and precipitation 101
4.3.9 Library preparation for next-generation sequencing (NGS) 101
4.3.10 High-throughput DNA sequencing and computational analysis 101
4.3.11 Functional analysis of differentially enriched guides 102
4.3.12 Statistical Analysis 102
4.4 Results 103
4.4.1 Identification of sensitive genes to CSE in A549 103
4.4.2 The biological relevance of potential functional hits from the A549 screen 104
4.4.3 Selection of biologically relevant enriched sgRNA guides targeting genes from the A549 screen 108
4.4.4 Identification of genes sensitive to CSE in BCi NS1.1 cells 110
4.4.5 Similarities between the two screens 110
4.5 Discussion 112
4.6 Summary of CRISPR library screen outcome 117
Chapter 5: General Discussion 118
5.1 Key findings from this research 119
5.2 Limitations of the current studies 121
5.3 Future directions 122
5.4 Conclusion 123
References 124

Appendix 147
Chapter 2: Longitudinal analysis of genome-wide expression profiling of cigarette smoke-induced experimental COPD 147
Chapter 3: Longitudinal evaluation of whole lung miRNAs and their interactions with gene expression in experimental COPD 160
Chapter 4: Genome-wide CRISPR-Cas9 Knock-out library screening to identify protective factors against cigarette smoke 162
Supplementary tables 163

List of Abbreviations

Abbreviation	Details
7AAD	7-aminoactinomycin D
A549	Adenocarcinoma human alveolar basal epithelial cells
Ago2	Argonaute
AhR	Aryl hydrocarbon receptor
Ahrr	Aryl hydrocarbon receptor repressor
ANOVA	Analysis of variance
ARNT	Aryl hydrocarbon receptor nuclear translocator
ASM	Airway Smooth Muscle
ATCC	American Type Culture Collection
BCi-NS1.1	Human airway epithelium-derived basal cell line
Beas-2B	Human bronchial epithelial cell line
BH	Benjamini–Hochberg
BP	Biological process
BTNL9	Butyrophilin-like 9
Cas9	CRISPR-associated 9
CC	Cellular component
cDNA	complementary DNA
CMH	Chronic Mucus Hypersecretion
COPD	Chronic obstructive pulmonary disease
CRISPR	Clustered regularly interspaced short palindromic repeats
CRISPRa	CRISPR activation
CRISPRi	CRISPR inhibition
crRNA	CRISPR RNA
CS	Cigarette smoke
CSE	Cigarette smoke extract
CSR1	Cellular stress response 1
Ct	Cycle threshold
DEGs	Differentially expressed genes
DLco	Diffusing lung capacity for carbon monoxide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSBs	Double stranded breaks
ECM	Extracellular Matrix
EGFP	Enhanced green fluorescent protein
eIF2α	Eukaryotic translation initiation factor 2 subunit 1
EIF3G	Eukaryotic translation initiation factor 3
ER	Endoplasmic reticulum
ES	Enrichment score
FBS	Fetal bovine serum
FC	Fold change

FDR	False Discovery Rate
FEF	Forced Expiratory Flow
FEV1	Forced expiration volume in one second
FITC	Fluorescein isothiocyanate
FRC	Functional Residual Capacity
FVC	Forced vital capacity
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GeCKO	Genome-wide CRISPR knockout library
GO	Gene ontology
GOLD	The Global Initiative for Chronic Obstructive Lung Disease
gRNA	guide RNA
GSVA	Gene set variation analysis
HBE	Human bronchial epithelial
HEK293T	Human embryonic kidney line expresses a mutant version of the SV40 large T antigen
HIF	Hypoxia inducible factor
HPRT	Hypoxanthine-guanine phosphoribosyl transferase
ICE	Inference of CRISPR Edits
ICS	Inhaled corticosteroids
IDT	Integrated DNA Technologies
ILs	Interleukins
INDELs	Insertions and deletions
IP10	Interferon-γ-induced protein 10
IPF	Idiopathic Pulmonary Fibrosis
IVC	Individually Ventilated Cage
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	Knock out
LABA	Long-acting ß-agonist
LAMA	Long-acting muscarinic antagonist
LC	Liquid chromatography
LCV2	lentiCRISPR v2
LOF	Loss of function
log	Logarithm
LPS	Lipopolysaccharide
LTOT	Long-term oxygen therapy
MAGeCK	Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout
MF	Molecular function
miRNA	microRNA
MMPs	Matrix Metalloproteinases
MOI	Multiplicity of infection
mRNA	messenger RNA
MRPs	Mitochondrial ribosomal proteins
MS	Mass spectrometry

MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
ncRNAs	non-coding RNAs
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NSCLC	Non-Small Cell Lung Cancer
PAC	Puromycin-N-acetyltransferase
PAM	Protospacer adjacent motif
PCA	Principal component analysis
PEG	Polyethylene glycol
PES	Polyethersulfone
PLLP	Plasmolipin
PPI	Protein-protein interaction
pre-miRNA	premature miRNA
pri-miRNA	primary miRNA
qPCR	Quantitative polymerase chain reaction
REAC	Reactome database
RIN	RNA integrity number
RISC	RNA-induced silencing complex
RMA	Robust Multi-Array Average
RNA	Ribonucleic acid
ROS	Reactive oxygen specie
RPs	Ribosomal proteins
RRA	Robust ranking aggregation
rRNA	ribosomal RNA
RT	Room temperature
RT-PCR	Real time-PCR
SEM	Standard error of mean
SF3A3	Splicing factor 3A3
snoRNA	small nucleolar RNA
SNPs	Single nucleotide polymorphisms
snRNA	small nuclear RNA
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TIDE	Tracking of Indels by Decomposition
TKOv3	Toronto Knockout version 3
TLC	Total lung capacity
TNPO1	Transportin-1
tracrRNA	trans-activating CRISPR RNA
tRNA	transfer RNA
UPR	Unfolded protein response
UTR	untranslated region
WT	Wild type
XREs	Xenobiotic response elements

List of Figures

Figure 1.1: Chronic obstructive pulmonary disease (COPD) 2
Figure 1.2: Risk factors for the development and progression of COPD 4
Figure 1.3: Experimental workflow of microarrays 12
Figure 1.4: miRNA biogenesis 14
Figure 1.5: Workflow for RNA-Seq 16
Figure 1.6: Gene editing through CRISPR-Cas9 and DNA repair 23
Figure 1.7: Main steps of a CRISPR library screening 24
Figure 1.8: Workflow of genome editing library screening against a phenotype 26

Figure 2.1: Workflow of expression profiling in CS-induced experimental COPD and *in-vitro* functional analysis using CRISPR-Cas9 36
Figure 2.2: Gene expression profile in CS-induced experimental COPD. 44
Figure 2.3: Functional enrichment analysis and expression pattern of DEGs over time in experimental COPD 48
Figure 2.4: *In-vitro* functional analysis of CRISPR-Cas9-generated *ARNT2* knockout (KO) in the BEAS-2B cell line. 51
Figure 2.5: Functional analysis and validation of identified dysregulated genes in *ARNT2* knockout (KO) cells exposed to CSE 54
Figure 2.6: Proposed mechanism of *Arnt2* in regulating COPD development 61

Figure 3.1: Data integration approach 66
Figure 3.2*:* Expression profiling of miRNAs in lung tissues in CS-induced experimental COPD 73
Figure 3.3: Correlation of longitudinal differentially expressed miRNAs with disease severity in human COPD 75
Figure 3.4: Integration of miRNAs and gene expression profiles in CS-induced experimental COPD that were differentially expressed during longitudinal smoke exposure 77
Figure 3.5: Integration of differentially expressed miRNAs with validated and predicted direct targets identified using multiMiR in COPD 79
Figure 3.6: Anti-correlated miRNAs and their direct targeted genes identified through longitudinal analysis in CS-induced experimental COPD 80
Figure 3.7: Functional enrichment analysis of negatively correlated direct targets of selected miRNAs in experimental COPD (BALB/c mice). 81

Figure 3.8: *Correlation* of negatively associated miRNAs and their target genes in CS-induced experimental COPD in C57BL/6 mice. 83
Figure 3.9: Correlation of longitudinal differentially expressed miRNAs with lung function parameters in experimental COPD in C57BL/6 mice. 84
Figure 3.10: Proposed mechanism of miRNAs regulating the targeted gene *Arnt2* 90

Figure 4.1: Estimation of percentage of viable A549 and BCi NS1.1 cell line 99
Figure 4.2: Pooled genome-wide CRISPR library screen workflow 100
Figure 4.3: Distribution of significantly enriched guides targeting genes in CSE exposed knock-out A549 cells 106
Figure 4.4: Functional enrichment analysis for identified hits from the A549 screen. 107
Figure 4.5: Gene level and sgRNA level analysis of top identified hits from T1 and T3 of A549 screen. 109
Figure 4.6: Distribution of significantly enriched and depleted sgRNA guides targeting genes in CSE exposed knock-out BCi NS1 cells 111
Figure 4.7: Hits from the A549 CRISPR library screen that might protect against CS-induced cell death and COPD 117

Figure S2.1: Gene set variation analysis of smoke-affected and disease-related genes through GSVA 154
Figure S2.2: STRING interaction network and functional analysis of dysregulated genes identified during longitudinal analysis in CS-induced COPD mouse model 155
Figure S2.3: Human *ARNT2* isoforms extracted from ENSEMBL to ensure complete ablation of the gene 156
Figure S2.4: Distribution of *observed genes* during RNA-sequencing of ARNT2 KO CSE exposed cells 157
Figure S2.5: Using STRING, the cluster analysis of identified dysregulated genes from RNA sequencing of *ARNT2* KO CSE exposed cells 158
Figure S2.6: Functional enrichment analysis potential candidates from *ARNT2* KO CSE exposed cells through g:GOSt module of g:profiler. 159

Figure S3.1: The relative expression of genes in whole lungs of C57BL/6 mice 160
Figure S3.2: The relative expression of miRNAs in whole lungs of C57BL/6 mice 161

Figure S4.1: Percentage of mapped and unmapped reads after alignment of guides sequences to the TKOv3 reference library. 162

List of Tables

Table S2.1: Primers used for quantitative PCR. 163
Table S2.2: Differentially expressed genes identified during initial cigarette smoke (CS) exposure of 4 weeks in mice. 164
Table S2.3: Differentially expressed genes identified during longitudinal cigarette smoke (CS) exposure in mice from 4 to 12 weeks. 173
Table S2.4: List of enriched pathways and gene ontology (GO) terms identified by STRING analysis. 177
Table S2.5: List of functionally enriched disease related differentially expressed genes identified by using g:GOSt module of g:Profiler. 178
Table S2.6: List of enriched gene ontology (GO) terms of cluster 4 and cluster 5 identified by STRING analysis. 185
Table S2.7: List of functionally enriched top dysregulated genes identified by using g:GOSt module of g:Profiler. 199

Table S3.1: Primers used for quantification of miRNAs through qPCR. 203
Table S3.2: Primers used for quantification of genes through qPCR. 204
Table S3.3: Differentially expressed miRNAs identified during longitudinal cigarette smoke (CS) exposure in mice from 4 to 12 weeks. 204
Table S3.4: Functional enrichment analysis of negatively correlated direct targets of disease related miRNAs identified by using g:GOSt module of g:Profiler. 206

Table S4.1: Hits identified during first round of CSE exposure (Time point T1) in A549 KO cells using MAGeCK. 208
Table S4.2: Hits identified during second round of CSE exposure (Time point T2) in A549 KO cells using MAGeCK. 210
Table S4.3: Hits identified during third round of CSE exposure (Time point T3) in A549 KO cells using MAGeCK. 210
Table S4.4: Functional enrichment analysis of hits identified during first round of CSE exposure (Time point T1) in A549 KO cells using enrichr GO:BP and REACTOME database. 219

Table S4.5: Functional enrichment analysis of hits identified during second round of CSE exposure (Time point T2) in A549 KO cells using enrichr GO:BP and REACTOME database. 233
Table S4.6: Functional enrichment analysis of hits identified during second third of CSE exposure (Time point T3) in A549 KO cells using enrichr GO:BP and REACTOME database. 235
Table S4.7: Hits identified in BCi NS1.1 KO cells during CSE exposure using MAGeCK. 237

Abstract

Chronic obstructive pulmonary disease (COPD) is a multifactorial irreversible complex chronic respiratory disease characterised by airway inflammation, mucus overproduction, and parenchymal tissue destruction. Cigarette smoke (CS) is one of the fundamental risk factors driving COPD pathogenesis. Due to the complex nature of COPD that remains incompletely understood, there are no treatments that stop the progression or reverse the condition, and the focus is solely on symptom management. Thus, there is an urgent need to elucidate the roles of molecular drivers of COPD and therapeutically target them. Keeping that in mind, the primary goal of this thesis was to identify potential candidates that can be targeted to abate the damage caused by abnormal processes in response to toxic inhalants, especially CS, that leads to COPD.

We first sought to address the fundamental knowledge gap through a comprehensive longitudinal gene expression profiling study of our well-characterized mouse model of CS-induced COPD. This highlighted that the dynamic interplay of smoke protective genes between COPD onset and progression becomes exhausted with constant CS exposure. Furthermore, aryl hydrocarbon receptor nuclear translocator 2 (*ARNT2)* was identified as a potential repressor of smoke-protective genes in long-term CS-induced COPD that may be a novel target to protect against the detrimental effect of smoke-induced cell death.

Our next aim was to understand disease pathogenesis and regulation at the post-transcriptional level in response to CS that induced COPD. To do this, we leveraged/combined high throughput profiling methods by integrating transcriptomics with miRNA expression profiles. The dynamic model of gene expression of CS-induced COPD coupled with miRNA provided insights into the disease at the molecular level. Such advances are important in defining the

miRNA regulation of direct target genes. We found that let-7b-5p, miR-181a-5p, and miR-181c-5p could be potential regulators of *Arnt2*.

Our final aim was to screen the whole genome against CS using a CRISPR-based genome-wide loss-of-function approach to investigate the targets such as *MMP15*, and *MRPL11* whose silencing/loss-of-function may mitigate the damage caused by CS. This study identified a critical set of novel potential genes that are responsive to CS damage that are mainly involved in RNA and protein biogenesis such as *RPS3*, *RPS25*, *MRPL11*, *ANK2,* and *RPL11*, and apoptosis such as apoptosis linked gene-2 *(ALG-2)*.

Overall, this thesis presents a longitudinal investigation of the pathogenesis of CS-induced experimental COPD at the transcriptional and post-transcriptional level, along with the genome-wide functional screening, to define potential candidates that can be targeted in the future to mitigate the progression of CS-induced lung damage especially in COPD.

Chapter 1: Introduction

Chronic obstructive pulmonary disease (COPD) is a chronic, inflammatory disease of the airways and/or the lung parenchyma that can destroy parenchyma, narrow the airways, and change the pulmonary vasculature according to the global strategy for prevention, diagnosis, and management of COPD (GOLD) (1). Chronic bronchitis and emphysema are the most prevalent conditions that make up COPD (Figure 1.1). COPD is also characterized by airway, lung tissue, and pulmonary blood vessel inflammation, and mucus overproduction, in which tissue destruction can cause obstructed exhalation, dyspnoea, coughing, fatigue, chest pain, sputum production, and other symptoms associated with COPD. Additional endogenous variables, such as persistent infection or autoimmunity, may also be the driving factor in COPD progression (2, 3).

1.1 COPD epidemiology and Global burden

A set ratio between the forced expiratory volume in one second (FEV_1) and the forced vital capacity (FVC), as recommended by GOLD, is used to diagnose COPD (4). The overall severity of this disease worsens by comorbidities and exacerbations. According to the Australian Bureau of Statistics, about 5.1% of Australians over 45 have COPD (5), while around 328 million people worldwide are believed to have COPD (6). The burden of COPD on society is significant; with more than 3 million fatalities annually. COPD is currently the world's third most common cause of death (7). Recent research has shown high mortality and morbidity of COPD, and this trend is rising each year (8, 9). The continued risks of COPD seriously affect the quality of life, endangering patients' health (7). Middle-aged people over 45 years are at higher risk of developing COPD, and this risk increases in heavy smokers (10).

Low-income countries often have more COPD cases due to indoor pollution and cigarette smoking as risk factors (11).

The diagnosis of COPD has historically been more common in men, but in recent years, the gender gap has narrowed. The reason could be women smoking more cigarettes. Still, the link is more complicated than that, involving varying susceptibilities between the sexes and the possibility that hormonal and genetic factors are involved (12, 13). Awareness of the adverse effects of smoking on health has drastically declined tobacco use in many countries. For instance, COPD morbidity decreased in Norway from 2001 to 2017 (14).

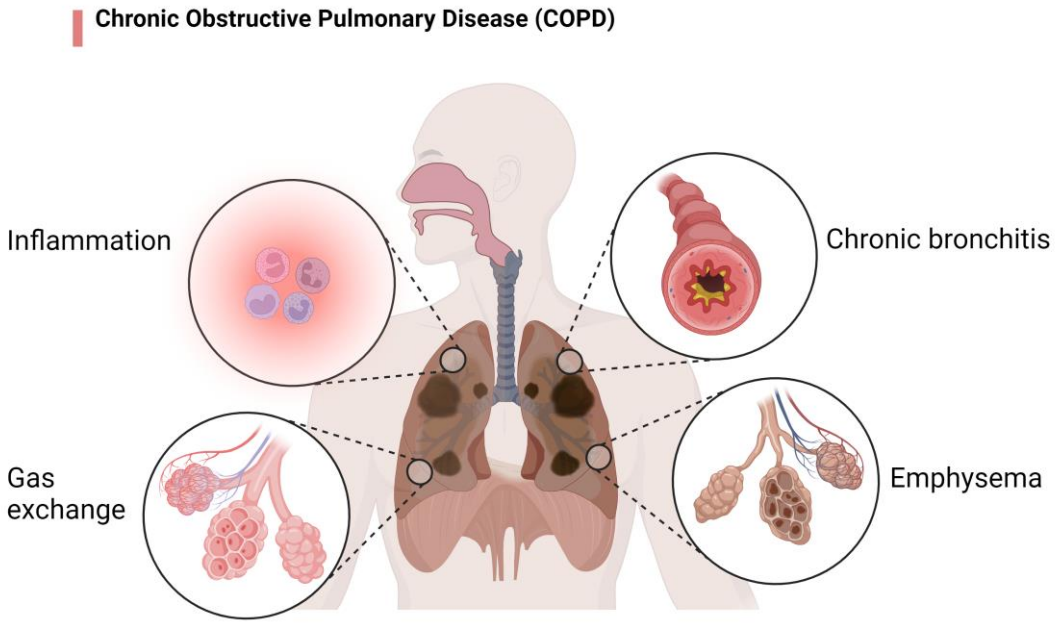

Figure 1.1: Chronic obstructive pulmonary disease (COPD): *COPD is a group of lung diseases characterized by airflow limitation, inflammation, impaired lung function, and gas exchange. Chronic bronchitis leads to swelling of the airways, which in turn increases mucus production, while emphysema results in stiffness and damage to alveolar air sacs. (Created with BioRender.com)*

1.2 COPD aetiology

The aetiology of COPD typically involves prolonged exposure to toxic chemicals or particles. In the past, smoking cigarettes and exposure to biomass fuels were thought to be the main contributing factors to developing COPD. However, recently air pollution has been found to increase the prevalence of COPD (15). Genetics has also been found to play a role in the susceptibility of developing COPD with a small subset of COPD patients having alpha-1-antitrypsin deficiency, which can cause emphysema at a young age (16). Even though several significant risk factors, including smoking, genetic factors, environmental factors, other lung infections, and prolonged exposure to chemicals, have been discovered (Figure 1.2), not everyone exposed to these risk factors develops COPD (17). Some patients appear to have more severe COPD, even at lower exposure levels. Several pathogenic processes underlying COPD and its progression appear to be driven by oxidative stress (18, 19). Exacerbations are a significant cause of hospitalization (20). Patients with COPD who experience three or more exacerbations during a 5-year follow-up have significantly lower survival odds (30% versus 80%) than those who do not (20). Even though smoking is the leading risk factor, 25–45% of never-smokers can also develop COPD. Studies have shown that COPD results from a more intricate interaction between noxious gas and particle exposures over time (from smoking, air pollution, and/or occupational exposure) and several host factors, such as epigenetic factors, poor lung development, age, and airway hyperresponsiveness.

Alterations in respiratory systems during COPD due to long-term exposure to harmful substances can broadly be divided into three distinct processes. Many patients have changes in their airways, where long-term inflammation brought on by immune cells invading the mucosa causes smooth muscle hypertrophy and mucus gland hyperplasia. Inhalation of harmful substances triggers the release of several chemotactic factors including IL-4, 5, 6, 8, and 13,

interferon-γ-induced protein 10 (IP10), monokine induced by interferon-γ, and leukotriene B4 by airway epithelium and macrophages. These factors attract and activate crucial inflammatory cells lining the airway mucosa (21, 22). Prolonged activation leads to the accumulation of neutrophils, CD8 T cells, and/or CD4 helper T cells in the airways (23, 24). This persistent immune response contributes to the degradation of elastin, especially in the alveoli, ultimately leading to emphysema. Additionally, the accumulated and hyperactivated neutrophils contribute to increased mucus secretion from goblet cells (25).This results in the thickening of the airway walls, airway obstruction, and sputum overproduction (26). Emphysema, or lung parenchymal destruction, is a second pathological alteration in COPD. Emphysema enlarges the distal airways by destroying the airway walls of the alveoli beyond the terminal bronchioles. Thirdly, there may be modifications to the pulmonary vasculature, such as the expansion of the intima and hypertrophy of the smooth muscles (27).

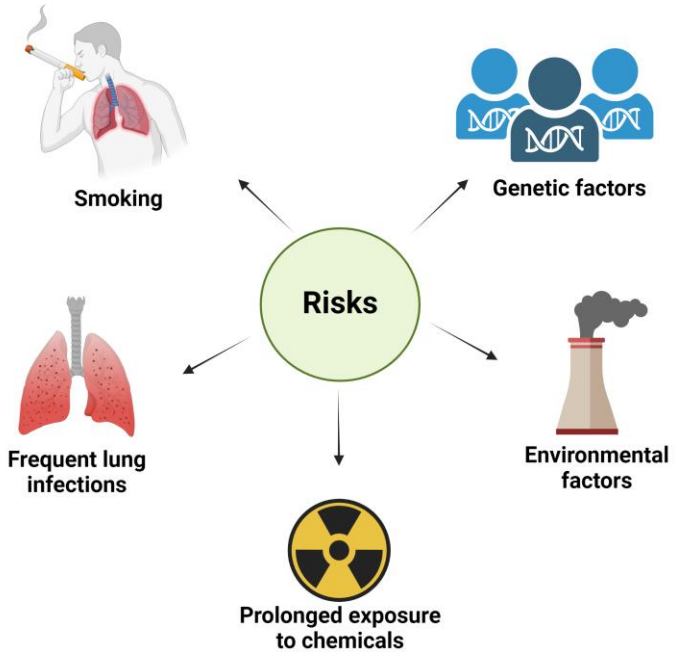

Figure 1.2: Risk factors for the development and progression of COPD: *Different risk factors including smoking, frequent lung infections, prolonged exposure to chemicals, and environmental and genetic factors can lead to COPD development and progression. (Created with BioRender.com)*

1.3 COPD prevention and treatment

Currently, there are only preventive measures and no definitive cure for COPD. First and foremost, quitting smoking is the best method to maintain lung health and prevent COPD (28). One of the best ways to protect lung health on a population-wide level is to impose restrictive tobacco laws (28). Tobacco-control programs have increased smoking cessation rates and lowered smoking initiation rates. The most effective intervention against COPD could be higher cigarette taxes (28, 29). Patients with COPD should be advised to stop smoking to maintain a better lifestyle (1, 30). The Lung Health Study found that sustained quitters experienced a slower deterioration in lung function, fewer hospitalizations, and a reduced overall death rate (31). COPD patients with fragile health can significantly benefit from pulmonary rehabilitation, which can help with dyspnea, physical activity, and general health (32). Vaccination against pneumococcal illness and seasonal influenza is another critical strategy. Influenza vaccination was strongly related to a decreased exacerbation in two distinct Cochrane evaluations (33, 34).

Most of the medications on the market are inhaled products that aim to lessen symptoms and, to some extent, lower the frequency of exacerbations (35). The beta-adrenergic agonist salbutamol is the most used drug for rescue therapy, while ipratropium bromide, a short-acting anticholinergic drug, is also used. Long-acting drugs, such as long-acting ß-agonists (LABA), long-acting muscarinic antagonists (LAMA), or a combination of the two, are used as maintenance therapies. Inhaled corticosteroids (ICS) are the most widely used treatment but are linked with an increased risk of pneumonia (36). Thus, ICS should be maintained for patients with eosinophilia or those experiencing frequent exacerbations. Some patients with end-stage COPD who experience respiratory failure, hypoxemia, or hypercapnia may benefit from long-term oxygen therapy (LTOT). Patients with severe COPD symptoms frequently

require systemic corticosteroids and, in many instances, antibiotics (35, 37). Increasing endogenous antioxidants or reducing oxidative stress can be a promising treatment strategy against COPD. However, due to the extremely high levels of oxidative stress in the lungs, finding safe and efficient antioxidants for COPD has been challenging. The goal of treatment is to return the lungs' redox balance to normal while maintaining the advantages of oxidant signaling. Effective biomarkers are lacking to determine which patients will benefit from antioxidant therapy the most or what doses can help restore the redox balance in COPD patients (18).

1.4 Molecular mechanisms of COPD development

It has always been challenging to understand the underlying mechanisms behind COPD due to its complex nature (38-41).

Molecular mechanisms for COPD progression include inflammatory pathways, oxidative stress, and protease/anti-protease imbalances. Airway epithelium and macrophages, upon exposure to cigarette smoke release interleukins ((IL)-4, 5, 6, 8, and 13), cytokine factors, interferon-γ-induced protein 10 (IP10), leukotriene B4 and monokine induced by interferon-γ (42, 43). Activation of inflammatory pathways leads to accumulating CD8/CD4 and neutrophils in the airways. This causes sustained immune responses and elastin degradation, causing emphysema (44, 45). The hyperactivation and accumulation of neutrophils increase mucus hypersecretion from goblet cells (46). Macrophage and epithelial cells also secrete transforming growth factor-β, causing fibroblast proliferation (47). Additionally, chemokine increasing epithelial senescence, known as a connective tissue growth factor, is also secreted by fibroblasts which causes a decrease in cellular regeneration and leads to emphysema (48). A higher prevalence of age-associated cellular senescence has also been observed in elderly patients with COPD (49).

Another key driver in COPD pathogenesis is oxidative stress, where reactive oxygen species (ROS) are produced by neutrophils and macrophages (18). ROS increases protease activity by activating p38 mitogen-activated protein kinase and nuclear factor-kappa B, which triggers the inflammatory genes and decreases the anti-protease activity process (50, 51). For protection against smoke-induced oxidative stress, maintenance of lung cells' antioxidant responses is essential. The inactivation of several antioxidant enzymes and transcription factors that control multiple antioxidant genes, such as forkhead box O3 (*FOXO3*) and factor-erythroid 2-related factor (*Nrf2*), has also been linked to a decrease in antioxidant defenses (52, 53). Many genetic investigations have identified changes in the mitochondrial malfunctioning route and the nuclear Nrf2-mediated oxidative stress pathway (54). Oxidative stress increases lung aging, damages DNA, activates autoantibodies, induces fibrosis and emphysema, and promotes chronic inflammation (55). Similarly, in COPD, oxidative stress regulates the TGF-β1 signaling pathway, the PI3K/mTOR signaling pathway, the NF-κB signaling pathway, and the expression and activation of sirtuin family members, including sirtuins 1–6, which in turn alter the inflammatory response, autophagy, or apoptosis process (56).

Another important element in the development and progression of COPD is an imbalance between proteases and antiproteases that results in lung parenchymal damage (57, 58). Proteases and antiproteases balance each other out in the healthy lung. Upon smoke or other pollutants exposure, cells such as neutrophils and macrophages migrate in and trigger the increased production of proteases that destroy healthy lung parenchyma (59). These proteases such as matrix metalloproteinases (MMPs) and elastase secreted by activated macrophages, neutrophils, and epithelial cells, break down the collagen and elastin fibers in the extracellular matrix (ECM), producing collagen-derived peptides or elastin fragments that facilitate the emphysema. Furthermore, the broken-down extracellular matrix produces chemotactic peptide fragments that support neutrophil and macrophage accumulation and fuel the inflammatory

processes associated with COPD (60). Pulmonary fibrosis and COPD can also develop as a result of airway epithelial and smooth muscle cell hyperplasia, which results in excessive ECM deposits in the interstitial lung (61, 62). Also, abnormalities in protease-antiprotease protection in the lung cause alpha-1 antitrypsin deficiency (AATD), a hereditary component that induces COPD (63).

1.5 Importance of mouse models in COPD research

Animal models are crucial for elucidating the mechanisms of disease pathogenesis, identifying new therapeutic targets, and developing and testing new therapies as well as validating *in-vitro* findings in a whole lung environment before clinical trials (64). Several small animal models have been developed to recapitulate COPD features and develop and test new therapeutic regimens against COPD (64, 65). Different methods are used to simulate COPD in animal models (66). These strategies involve genetic alteration, inflammatory stimuli (like LPS), proteolytic enzymes (like elastase), and CS (the main etiological cause for COPD) exposure in experimental animals (67-69). The mentioned COPD inducers have been administered to a variety of species, including dogs, guinea pigs, rats, mice, and hamsters. Among all the COPD inducers CS is the most important etiological cause for COPD (70, 71) and is the most prevalent inducer used in in-vivo research (72). Furthermore, cigarette smoke from the environment can exacerbate respiratory symptoms and COPD (71). Nevertheless, one of the limitations of using CS as an in vivo COPD-inducer is that there isn't currently a defined procedure or approach for animal exposure. Accordingly, the kind of smoke produced (research vs. commercial, filter or not), the components of the CS used for exposures, the delivery methods (whole-body vs. nose-only), and—most importantly—the amount of smoke given to the animals are all significant determinants (67, 73). Despite these drawbacks, CS has been demonstrated to cause

emphysema, pulmonary infiltration of neutrophils and macrophages, and airway fibrosis in animals with COPD (74-81).

Appropriate species that are frequently employed in COPD research are guinea pigs (72, 82-84). Due to the similarities in the morphology and physiology of their lungs with those of humans, these animals have numerous advantages (83, 85). There are also several physiological resemblances between humans and guinea pigs, particularly in the areas of airway autonomic regulation and allergen response (76, 83). Non-human primates are another suitable animal model for studying the processes underlying allergic airway disorders and COPD. Research has demonstrated that monkeys exposed to CS have various airway abnormalities and chronic respiratory bronchiolitis (86). Dogs' histology and pathophysiology of developing chronic bronchitis and emphysema after being exposed to CS are also comparable to those of humans (87). Like other models of COPD, the canine model has also been used to evaluate novel therapies before human trials (87). Nevertheless, there are drawbacks, including the expense of acquiring and maintaining the animals as well as the absence of molecular tools, and the requirement to test a large number of chemicals for pharmacological investigations (66).

Among the small animal models, the most popular species to be exposed to CS to induce an animal model of COPD in mice (67, 88). Mice are the ideal choice for animal models of COPD when it comes to immunological processes. The mouse genome has also been extensively sequenced and has revealed similarities to the human genome (89). Additionally, suggestions have been made (66, 90) on the possibilities of manipulating gene expression. Mouse models have many advantages, including short breeding time, low breeding and housing costs, easy genetic modification, extensive knowledge of the genome, and the availability of a plethora of immunological tools (64, 65, 91-93). Nevertheless, several investigations have demonstrated that distinct mouse strains exhibit varying degrees of sensitivity to CS challenge (67). In many investigations, different exposure protocols were applied, and mice were exposed to CS in a

smoking apparatus as whole-body exposure (94-99) or nose-only exposure (100-103), once or twice a day, many times a week, for varying days, weeks, or months. Rats are also utilized as animal models of COPD (66), however due to their apparent resistance to the development of COPD (66, 76) they are regarded as poor models (76, 104).

A well-defined short-term CS-induced model that manifests COPD features after only 8 weeks of CS exposure in mice (102, 103, 105-110) was used in this thesis. In this model, a specially designed and purpose-built nose-only, directed flow inhalation and smoke-exposure system housed in a fume and laminar flow hood is used to expose BALB/c or C57BL/6 (B6) to cigarette smoke (twelve 3R4F reference cigarettes [University of Kentucky, Lexington, Ky] twice per day and five times per week for one to twelve weeks). Every exposure lasts for 75 minutes. Up to 12 weeks, cigarette smoke was delivered into the nares of the experimental mice. During the first three weeks of exposure, the animals lost 10% of their starting weight and recovered back 5% of it. After four days of exposure to cigarette smoke, increased neutrophil and macrophage counts in the BALF indicate acute inflammation of the airways. After eight weeks, the inflammation still exists and gets worse due to the participation of lymphocytes, especially CD8+T cells. Smoking also leads to increased chronic parenchymal inflammation. Increased mucus-secreting goblet cells in the airways starting in week six and thicker airway epithelium starting in week eight are the indicators of airway remodeling. After eight weeks, alveolar enlargement indicates emphysematous tissue destruction, that gets severe by 12 weeks. Exposure increased functional residual capacity (FRC), total lung capacity (TLC), and decreased hysteresis, transpulmonary and airway-specific resistance, tissue damping, and forced expiratory volume in 100 ms (FEV100)/forced vital capacity (FVC) ratio (representative of FEV1/FVC ratio in human subjects). These parameters indicate mice's impaired lung function after smoke exposure is comparable to what is seen in COPD-affected human beings. Furthermore, losing skeletal muscle mass, systemic inflammation, and cardiovascular illness

are common in people with COPD, indicating that the lung is not the only organ negatively impacted in humans. In this model, the proportion of leukocytes in the blood significantly changed after 8 weeks of exposure indicating systemic changes as well. Additionally, there was a decrease in the mass of skeletal muscle, specifically the quadriceps. In addition, the mice exposed to smoke had larger hearts with over 25% more fatty tissue surrounding them. Moreover, experimental COPD induction also makes respiratory tract infections worse. Furthermore, in our model, 8 weeks of smoke exposure followed by 4 weeks of cessation did not improve lung function, emphysema, airway inflammation, or abnormalities in circulating leukocytes. This is comparable to what has been observed in patients with COPD, whose clinical conditions typically do not significantly improve after quitting smoking, and in many cases, lung function worsens.

1.6 Transcriptomics and gene expression profiling

One strategy to uncover the mechanisms that drive COPD is to measure changes in transcriptome, which refers to the quantitative measurement of RNA transcripts, including mRNA, rRNA, tRNA, and other non-coding RNA molecules made in a cell (111).

1.6.1 Microarrays

Since its introduction in the middle of the 1990s, microarray technology swiftly emerged as the most popular for transcriptome profiling (112, 113). A microarray is a set of microchips and microbeads that carry DNA probes (nucleotide oligomers). The RNA extracted from target and control samples is put through reverse transcription, labeling, and cDNA hybridization before being applied to a microarray. Fluorescently marked probes are used to quantify the amount of hybridization, and the data are then used to calculate expression levels (114) (Figure 1.3). DNA microarray is a flexible instrument that may be used to investigate various cell

features, including DNA methylation, single nucleotide polymorphisms (SNPs), DNA copy number variation, and gene and miRNA expression (115). A DNA microarray is often a glass slide with thousands of microscopic DNA probe spots organized in rows and columns on a flat surface where each spot represents a different gene. There are different microarray technology types, but two of the most often used are spotted cDNA microarrays and oligonucleotide arrays. Spotted cDNA microarrays use cDNA probes that are spotted onto glass slide robotic arrays (116). DNA microarrays have offered crucial insights for understanding the process underlying genetics and gene expression regulation. DNA microarrays are a well-established and reasonably priced technique nowadays. However, because this method depends on nucleic acid hybridization, it has intrinsic limitations. It generates a lot of background noise because of cross-hybridization artifacts, which reduces the precision of the expression measurements.

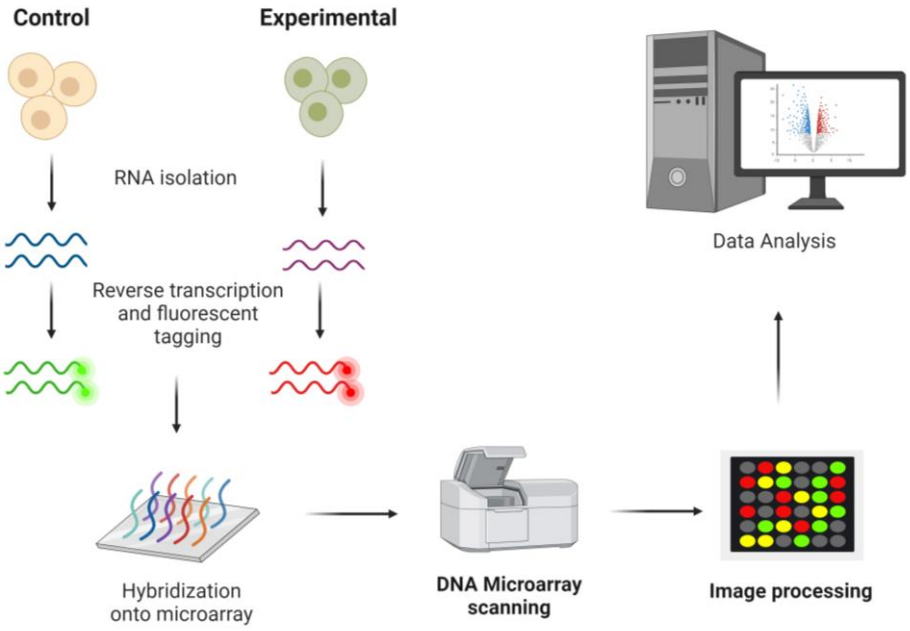

Figure 1.3: Experimental workflow of microarrays: *The first step in microarrays is to extract RNA from control and experimental samples which are then reverse transcribed and fluorescently tagged before hybridizing onto microarray chips. The data is then scanned using microarray scanning machines followed by Bioinformatics analysis. (Created with BioRender.com)*

The number of transcripts examined using microarrays is likewise constrained, especially for transcripts with low abundance (111, 117). Thus, quantifying absolute expression levels is not always a good fit for DNA microarrays. Microarrays have been extensively used in the context of understanding various respiratory diseases, including COPD, through the identification of differentially expressed genes linked to critical molecular pathways (113, 118-121). Different studies have compared smokers without COPD and those with COPD and found a significant difference between their expression profiles and observed an increase in expression levels of *TGFB1, EGR1, CTGF*, and *CYR61* (122-124). Additionally, emphysema patients' primary lung fibroblasts were discovered to have these genes upregulated compared to regular donors (125). New studies are now identifying novel COPD gene signatures by exploring public datasets using computational approaches. For example, one recent study investigated seven publicly available COPD datasets and identified new genes involved in several immune and cell cycle-related pathways (126).

1.6.2 Investigating microRNAs using microarrays

As was indicated earlier, nucleic acid characteristics like DNA methylation, single nucleotide polymorphisms, and copy number variations can all be measured using microarrays. It is feasible to use probes to measure the expression of non-coding RNAs in transcription and probes for detecting microRNA (miRNA) expression. One type of non-coding RNA that has recently attracted a lot of attention and is a frequent topic of investigation using microarray-based platforms is microRNA. Short RNA transcripts known as miRNAs (20–23 nucleotides) attach to the 3′UTR of mRNA targets and can alter the expression levels or translation rates of those targets (127). miRNAs play a crucial function as regulators of gene expression (Figure 1.4).

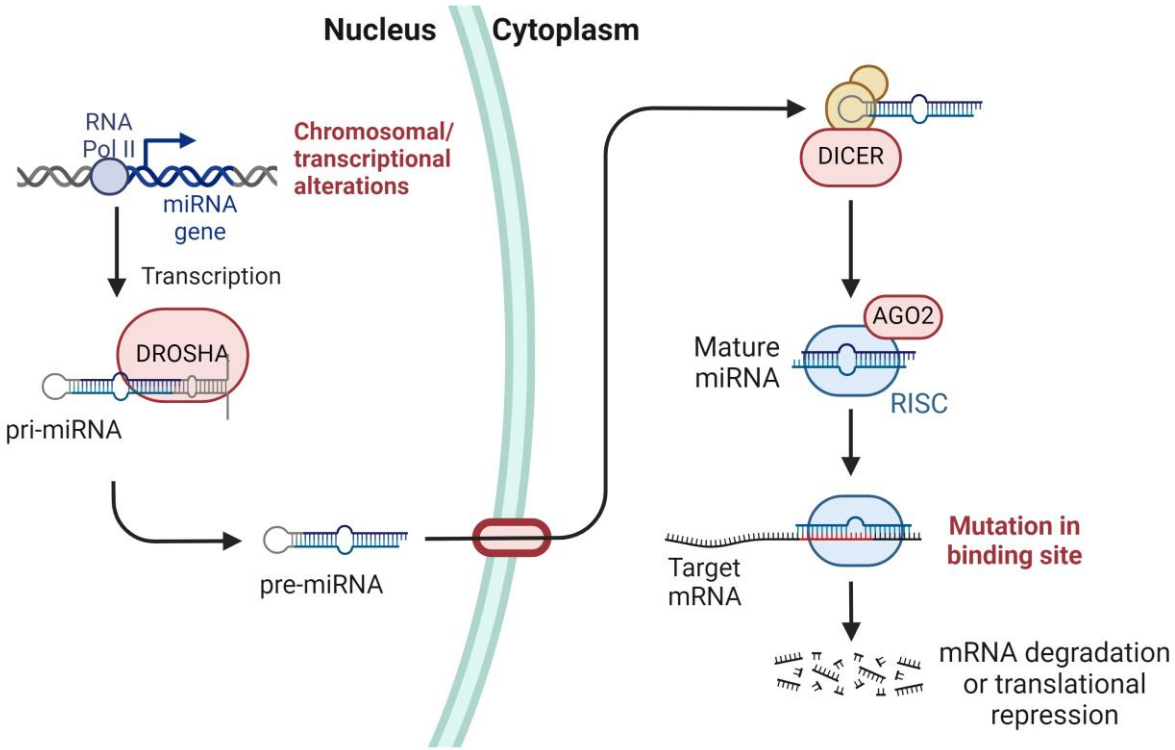

Figure 1.4: miRNA biogenesis: *RNA polymerase II or III helps miRNA maturation through the production of primary miRNA transcript (pri-miRNA), which is then cleaved inside the nucleus by a complex called Drosha. After cleavage, pre-miRNA exports into the cytoplasm through an exportin protein. The hairpin of exported pre-miRNA is then cleaved by RNase dicer and transformed into a mature miRNA. Mature miRNA along with Argonaute (Ago2), guides the RNA-induced silencing complex (RISC) to target the complementary mRNA sequence and induce silencing either through mRNA cleavage, deadenylation, or translational repression. (Created with BioRender.com)*

Similarly, disruption in the expression patterns of miRNAs may aid in the aetiology of disease. MiRNAs might be effective biomarkers as well. Several commercial systems have been created to simultaneously measure the expression level for hundreds of miRNAs (128). To better understand the role of microRNAs in pathogenesis and the development of biomarkers, some research utilizing microarrays has explored the expression of miRNAs in the physiologic

response to tobacco smoke and lung disease, including COPD, asthma, and lung cancer (129-131). Multiple studies have revealed the association of COPD pathogenesis with the up/down-regulation of different miRNAs, including miR-145, miR-1, miR-144, miR-125, and miR-101 (132, 133). Various studies have found an association between the deregulation of miRNAs with COPD progression. In one study, downregulation of miRNAs including miR-483-5p, miR-29b, miR-629, miR-106b, miR-101, miR-133b, and miR-532-5p was observed in plasma from COPD patients (134). Upregulation of miR-126 and miR-29c was observed in stage III COPD patients, suggesting their potential role as diagnostic biomarkers for COPD patients (133).

1.6.3 RNA sequencing

Now that sequencing technology has advanced, researchers can examine the entire transcriptome down to the single base level, thanks to RNA sequencing (RNA-Seq) technology. Deep sequencing, or massively parallel sequencing, can assess the levels of mRNA or miRNA expression, just like microarrays. However, in contrast to microarrays, RNA-Seq does not rely on the principle of hybridization to assess expression, which requires prior knowledge of the gene sequence to construct appropriate probes. RNA-Seq is majorly being used for the investigation of differential gene expression, where up-and-down-regulated genes can be identified (135). This technique is used to specify alternative splicing in addition to gene fusion events, new transcripts, and splice junctions (136, 137). RNA sequencing's primary goal is to learn more about the cellular components, biological functions, and molecular mechanisms of potential gene products. The metabolic pathways within the transcriptome can be predicted and visualized using different bioinformatics methods (138, 139). The extraction of RNA from biological samples is the first step in RNA sequencing which is then reverse transcribed and converted into double-stranded cDNA. The DNA barcodes are appended to

one or both ends of each fragment, and the cDNA readings are coupled to universal adapter sequences. Using a next-generation sequencing (NGS) platform, the sequencing procedure results in millions of so-called reads (like Illumina HiSeq). However, the resulting raw datasets are enormous and intricate. Reads are collected and mapped to a reference genome or a transcriptome. The main objective is to identify the genomic region where each short read most closely resembles the reference genome (Figure 1.5).

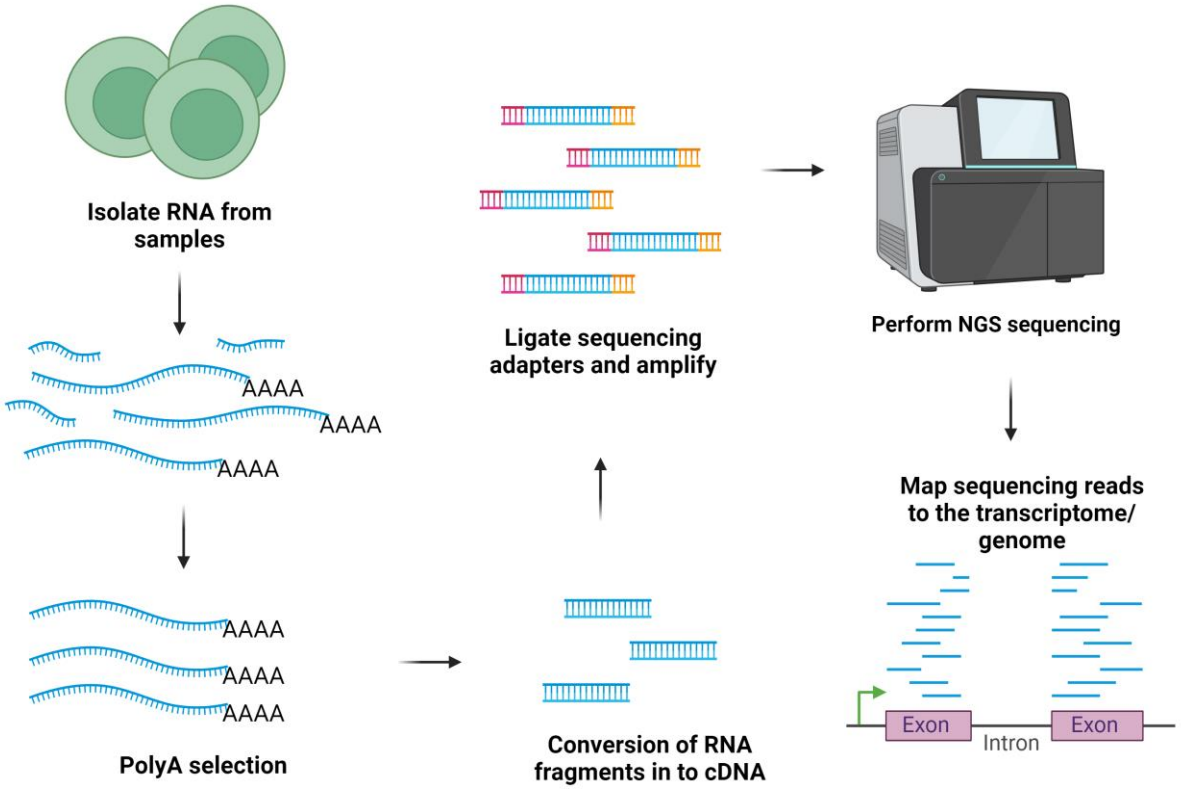

Figure 1.5: Workflow for RNA-Seq: *Firstly, RNA is extracted from a biological sample and is selected based on different protocols to enrich polyadenylated transcripts like poly-A selection. This method removes rRNA or tRNAs from the sample. Next, RNA is reverse transcribed to cDNA, followed by RNA-Seq library preparation to attach sequence adapters to the ends of the cDNA strands. cDNA strands were then subjected to NGS. Subsequently, the raw reads from sequences were then mapped to the reference genome for analysis. (Created with BioRender.com)*

The readings are then compiled and summarised at the transcript, exon, and gene levels (140). Several studies have been published to understand COPD pathogenesis through RNA-Seq profiling (119, 120, 141-143). Most of these studies have been conducted in peripheral blood (144-146). Macrophages have been found to have a putative role in lung diseases; therefore, studies are now investigating the gene expression profile of these cells to understand COPD (147). A recent study explored gene expression profiles across three tissues including alveolar macrophage, large airway epithelium, and peripheral blood samples from the same COPD and control subjects. A significant increase in known smoking genes, including *CYP1B1* and *AHRR*, was observed in large airway epithelium, while an increase in B-cell-related genes was seen in peripheral blood. There was an overlap of immune and haemostasis-related genes across all three tissues for emphysema (139). Ghosh et al. recently performed RNA-Seq to explore the gene expression profile of patients with COPD and healthy individuals. A significant increase in myeloid genes including *SFTPD, PSMD3, HAL, HHIP,* and *RIN3 was* observed in COPD, and gene signature (*GPR4, MFHAS1, HHIP-AS1, HEYL, SFTPD, PSMD3, HAL, NR4A2, CHI3L2, IRF2BP1, HHIP, RIN3,* and *BCAR1*) upregulated in COPD was associated with lower lung function *(148).*

1.7 Integrative omics and its role in understanding COPD

None of the available omics techniques can provide complete information about the biological system, as every single method has its advantages and disadvantages. The concept of integrative omics has emerged to cope with the limitations of different omics techniques, but can we use integrative omics as a solution to deal with the current omics limitations? The answer to that question is a yes, as during recent years, the integration of various omics technologies, including genomics, proteomics, and metabolomics, has led to significant discoveries in respiratory research (149-151). For example, a miRNA-mRNA-protein

dysregulated network of lung-derived primary immune cells associated with COPD in females elucidated new molecular insights highlighting dysregulated interactions (152). MiR-134-5p, miR-146a-5p, and let-7 family, along with their potential target genes, such as *KRAS* and *EDN1*, were discovered as potentially important miRNA-mRNA networks regulating chronic mucus hypersecretion (CMH) in COPD during an integrative regulatory network investigation (153). Integrating coding and noncoding transcriptomics, methylomes, and proteomes revealed novel molecular relationships in idiopathic pulmonary fibrosis (IPF), including 10 lncRNAs that were likely important in regulating the gene matrix metalloproteinase 7 (MMP-7), which has a known role in IPF, and 18 differentially methylated regions (154). Another integrated analysis of transcriptomics and proteome profiles of IPF patients discovered possible biomarkers such as butyrophilin-like 9 (BTNL9) and plasmolipin (PLLP) and their relevance in IPF (155). Briefly, computational biology has played a substantial role in integrating omics data helping diagnose and treating respiratory diseases (156, 157). Genomics and transcriptomics help measure an organism's gene expression, while genomics and proteomics help understand complex biological pathways. Integration of these technologies can help develop a better picture of cellular responses from transcription to translation. Despite their importance in understanding diseases, both technologies (i.e., genomics and proteomics) have their limitations (158). For example, only 5-10% genetic variance occurs in a population, and drug toxicity and disease progression can significantly vary between people. More data gives power and helps develop a better understanding at the systems level (158). Thus, the integration of omics technologies together can help find the answers that single omics techniques might not provide. It is possible to understand how aberrant miRNA expression may contribute to the mechanisms behind disease-related gene expression patterns by combining gene and miRNA expression measurements in the same clinical samples. (115). Several studies have been conducted to understand COPD by integrating mRNA and miRNA datasets (132, 159, 160).

The study of miRNAs in respiratory diseases including COPD has been rapidly extended. Upregulation of miR-15b was found to be associated with the downregulation of *SMAD7* and *SMURF2* in bronchial epithelial cell lines (132). O'Leary et al. investigated the role of miR-145 in COPD progression by targeting the *SMAD3* gene, a critical downstream signaling mediator within the TGF-β pathway (161). They elucidated that the regulation of miR-145 expression is under negative control by pathways involving MAP kinases, specifically MEK-1/2 and p38 MAPK. Moreover, through experimental manipulation involving the overexpression of miR-145 using synthetic mimics in airway smooth muscle (ASM) cells derived from COPD patients, they observed a significant suppression of interleukin-6 (*IL-6*) and *CXCL8* release. This suppression effectively reduced cytokine levels to a magnitude comparable to those observed in non-smoker controls. These findings underscore the role of miR-145 in negatively modulating the release of pro-inflammatory cytokines from ASM cells in COPD through its targeting of the *SMAD3* gene (161). Another study elucidated the role of miR-96 in regulating genes within the p53/hypoxia pathway, by observing the overexpression of *HIF1A*, *MDM2*, and *NFKBIB*, members of this pathway, in tissues obtained from patients with emphysema (162). Subsequently, integrated parallel miRNA and mRNA-Seq from the same samples highlighted miR-96 as a pivotal regulatory hub within the p53/hypoxia gene-expression network, revealing the involvement of miR-96 in regulating p53/hypoxia pathway genes (162). Several studies have confirmed the influence of smoking in the alteration of miRNA expression that can drive COPD progression. For example, miR-181c has been identified as a regulatory element targeting *CCN1*, resulting in a reduction of inflammation, reactive oxygen species (ROS) levels, and neutrophil infiltration (163). Remarkably, *CCN1* serves as a pivotal signaling mediator governing the progression of lung pathogenesis, orchestrating pivotal processes such as inflammation, apoptosis, and fibrosis. Studies have emphasized the crucial roles of *CCN1* in lung epithelial cell apoptosis following exposure to

oxidative stress (164). It has been observed that CSE induces an upregulation of *CCN1* expression in lung epithelial cells by triggering ROS and endoplasmic reticulum stress. Subsequently, this induction leads to an augmented IL-8 release through the activation of the Wnt pathway (165). Moreover, in another study, miR-126 deregulation was found to be associated with ATM kinase activation (166).

1.8 Functional studies through CRISPR/Cas systems

CRISPR stands for clustered, regularly interspaced short palindromic repeats, while Cas stands for CRISPR-associated protein. CRISPRs are repeating DNA sequences found in prokaryotes, including bacteria and archaea. In 1987, Yoshizumi Ishino and his colleagues discovered CRISPRs in E. coli after inadvertently cloning an odd sequence of repeats interspersed with spacer sequences while examining a gene that converts alkaline phosphatase (167). However, the biological function of these arrays remained unknown because insufficient DNA sequence data was available, which was defined as the CRISPR locus later. In the early 2000s, Mojica and colleagues observed that the spacer sequences resembled sequences from viruses, plasmids, and bacteriophages (168). They found that bacteriophages whose sequences were discovered in CRISPR spacers were unable to infect a spacer host cell but could infect closely related strains that did not have this spacer, indicating that these sequences are involved in prokaryotes' adaptive immune system (168). Mojica was the first to correctly surmise that the locus itself functions as a "compartment for storing DNA chunks of invaders" and that the purpose of the CRISPR system is to develop protection against foreign DNA (168, 169). Another group of researchers discovered protein-coding genes close to repeat loci, which they termed CRISPR-associated (Cas) proteins (170). Cas9 was demonstrated to be a crucial protein needed for the CRISPR system to inactivate the invasive phage (171). According to research conducted in 2008, it has been demonstrated that short RNA (CRISPR RNA, cRNA) generated

from the CRISPR locus attaches to Cas proteins and directs them to the target DNA to provide immunological protection (172).

The CRISPR/Cas adaptive immune response comprises three main phases: adaptation, expression, and interference (173). During the adaptation phase, the Cas proteins connect to the target region close to a protospacer-adjacent motif (PAM) after recognizing the foreign DNA. The target region is then removed to produce a protospacer and inserted into the CRISPR array to create a spacer. During the expression, the CRISPR array with the proper spacer is transcribed into pre-CRISPR RNA, which then becomes the mature CRISPR RNA (crRNA) through Cas proteins. The crRNA identifies the protospacer in the bacteriophage or virus encroaching during the interference phase. It removes the desired sequence from the genome using a Cas nuclease or nucleases, such as Cas9 (174).

In recent years, the CRISPR/Cas9 system has been extensively used to understand the function of different genes. The CRISPR/Cas9 system comprises two main parts: a guide RNA and a Cas9 endonuclease. Originally, the CRISPR/Cas system was based on two RNAs named crRNA and trans-activating CRISPR RNA (tracrRNA). The role of crRNA was to recognize the target region, and tracrRNA was to link the crRNA and Cas9 and mediate the processing of pre-crRNA into mature crRNA (175). In recent years, a single guide RNA (sgRNA) has been developed through the synthetic fusion of the crRNA and tracrRNA (176). Different variants of Cas9 protein are being developed, and the most commonly used variant is SpCas9, derived from Streptococcus pyogenes (177). A recent variant with enhanced fidelity known as eSpCas9 is also being developed (178).

1.8.1 CRISPR/Cas system

As it has been mentioned before, CRISPR/Cas systems have two main components: guided RNA (gRNA) and Cas9 endonucleases. It was possible to guide the Cas9 endonuclease activity by designing a specific chimeric single-guided RNA (sgRNA) sequence to cleave specifically the targeted DNA region. This discovery led to the development of a specific and efficient way of genome editing via the CRISPR/Cas9 system that revolutionized the era of functional genetics/genomics (179, 180).

With regards to gene editing in eukaryotic cells, the sgRNA binds to the specific site in the genome that guides the Cas9 to introduce double-stranded breaks (DSBs) at specified regions. These DSBs provoke the DNA repair mechanism called the non-homologous end joining method (NHEJ) repair mechanism. During the repair mechanism, small insertions or deletions (INDELS) incorporated into the DNA sequences disrupt the reading frame resulting in nonsense translation, leading to the premature stop codon. This premature translation ends up effectively knocking out the targeted genes (179, 181, 182) (Figure 1.6).

1.8.2 Genome-wide loss of function studies

The simplicity of the CRISPR-Cas9 system has made it possible to multiplex thousands of sgRNAs in pooled libraries to perform efficient and accurate gene knockout screening, even at the genome scale (183-187). CRISPR genomics screens are used to study the role of individual genes at scale by multiplexing many sgRNAs in a library for efficient and accurate genome-wide screening. The fundamental principle of CRISPR screening is to delete all potentially significant genes, but only one gene per cell. The intention behind this is to generate a mixed population of cells with each cell having a unique deleted gene. This mixed population of cells with different genetically encoded alterations is pooled together for pooled screening

procedures. The majority of CRISPR screening investigations aim to identify genes that are enabled to survive under particular circumstances such as drug treatment or other physiologically relevant circumstances. A successful experiment generates a list of potential genes or genetic sequences that seem to contribute to the physiological effect under study.

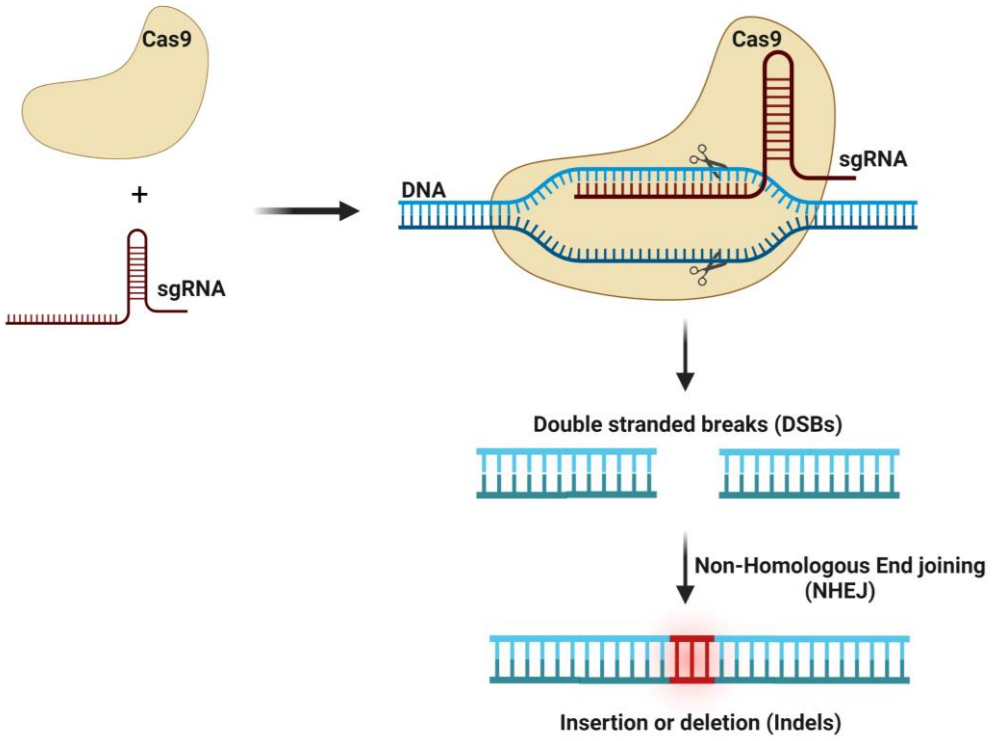

Figure 1.6: Gene editing through CRISPR-Cas9 and DNA repair. *The double-stranded DNA is cut to make a double-stranded break (DSB) through the Cas9 protein guided by a single-strand guide RNA (gRNA). This then leads to DNA repair through a repair mechanism called nonhomologous end joining (NHEJ). (Created with BioRender.com)*

Predominantly, the CRISPR screen has three main steps following by amplification of plasmids, packaging of plasmids, and transduction of cell line (Figure 1.7). Libraries developed by the Zhang lab have been the most extensively utilized resources for genome-wide CRISPR knockout screens (188). These lentiviral libraries target exons in humans and mice and are

offered as one-vector systems, in which the sgRNAs and Cas9 are located on the same plasmid, or as two-vector systems, in which they are located on different plasmids. Generally, the CRISPR libraries are supplied at very low concentrations, thus the first thing to do is to amplify the library to a concentration high enough to produce enough lentivirus for the experiment. sgRNA and Cas9 must be introduced into the cell via a delivery mechanism to perform targeted gene deletion using CRISPR/Cas9. Although, there are many delivery strategies for CRISPR, third-generation lentiviral vectors are the most commonly used for genome-wide loss-of-function screening (186, 188-191). These lentiviral vectors can integrate into the genomes of a wide variety of cell types and efficiently transduce these cells (192, 193).

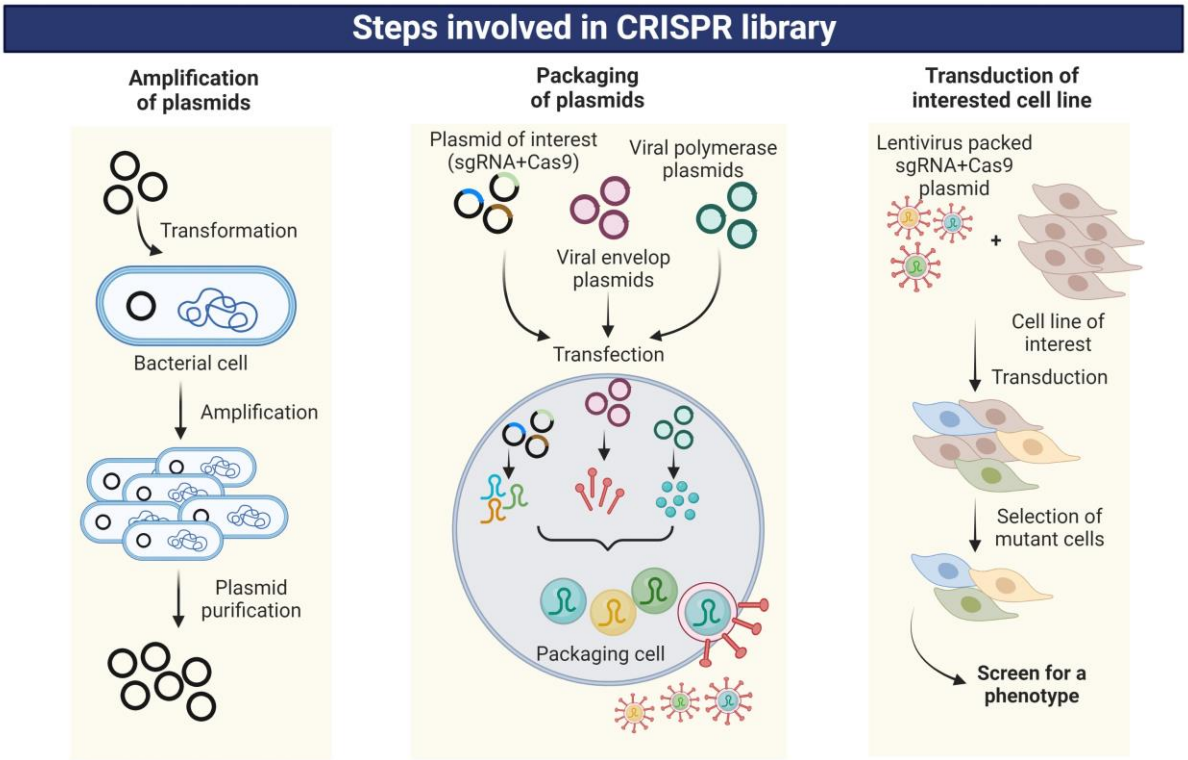

Figure 1.7: Main steps of a CRISPR library screening: *Firstly, to generate a CRISPR library the plasmids were first transformed into bacteria for amplification. The amplified plasmids were then packed into lentivirus by transfecting the packaging cell line with plasmids of interest along with plasmids containing viral components including viral polymerase and viral envelop plasmids.*

Subsequently, in a pooled screen, viral vectors carrying the complete sgRNA CRISPR library are bulk transduced into the cultured cells with low MOI to ensure each cell is infected by only one sgRNA-conta9ining particle. Following transduction, the cells are usually subjected to a selection procedure in which the non-transduced cells are isolated from transduced cells using selective markers associated with the CRISPR library (such as, only the transduced cells survive when cells with a resistance gene are exposed to the corresponding antibiotic). (Created with BioRender.com)

The production of third-generation lentiviral particles involves co-transfecting 293T human embryonic kidney (HEK) cells with an interchangeable envelope plasmid that encodes for an envelope glycoprotein of another virus (usually the G protein of the vesicular stomatitis virus; one or two transfer plasmids, depending on the applied library, encoding for Cas9 and sgRNA, two packaging plasmids, one encoding Rev and the other Gag and Pol, as well as selection markers (188). The supernatant containing the lentiviral particles is collected, concentrated, and then applied to the target cells for infection (194). In a pooled screen, viral vectors carrying the complete sgRNA library are bulk transduced into the cultured cells. Generally, a low multiplicity of infection (MOI) of 0.3-0.6 is utilized to ensure that each cell is infected by one sgRNA-containing particle (195, 196). To find the integrated sgRNAs, transduced cells will need to be selected for the desired phenotype either by positive or negative selection. Genetic sequencing will then be required (195). After phenotypic selection, a control cell population and the chosen clones' genomic DNA are extracted (195, 197, 198). The most widely used genome-wide knockout procedures involve a two-step polymerase chain reaction (PCR) to build a Next-generation sequencing (NGS) library (195, 197). Using primers unique to the lentiviral integration sequence, the sgRNA region is amplified in the first phase of amplification, and Illumina i5 and i7 sequences are then added in the second (197). The recovered sgRNAs can be identified using NGS of the PCR products, and the relative abundance of each sgRNA can be quantified (197).

The last stage in the screen is to discover which genes and pathways may be responsible for the observed phenotype involves computationally evaluating the significantly enriched or depleted sgRNAs and linking them to their corresponding genes (Figure 1.8). The Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout (MAGeCK) approach is the most widely used of the several methods available for this purpose (199). Additionally, the data, analysis, and quality controls can be interactively explored by users through an integrated interactive web-based visualization tool called VISPR (200).

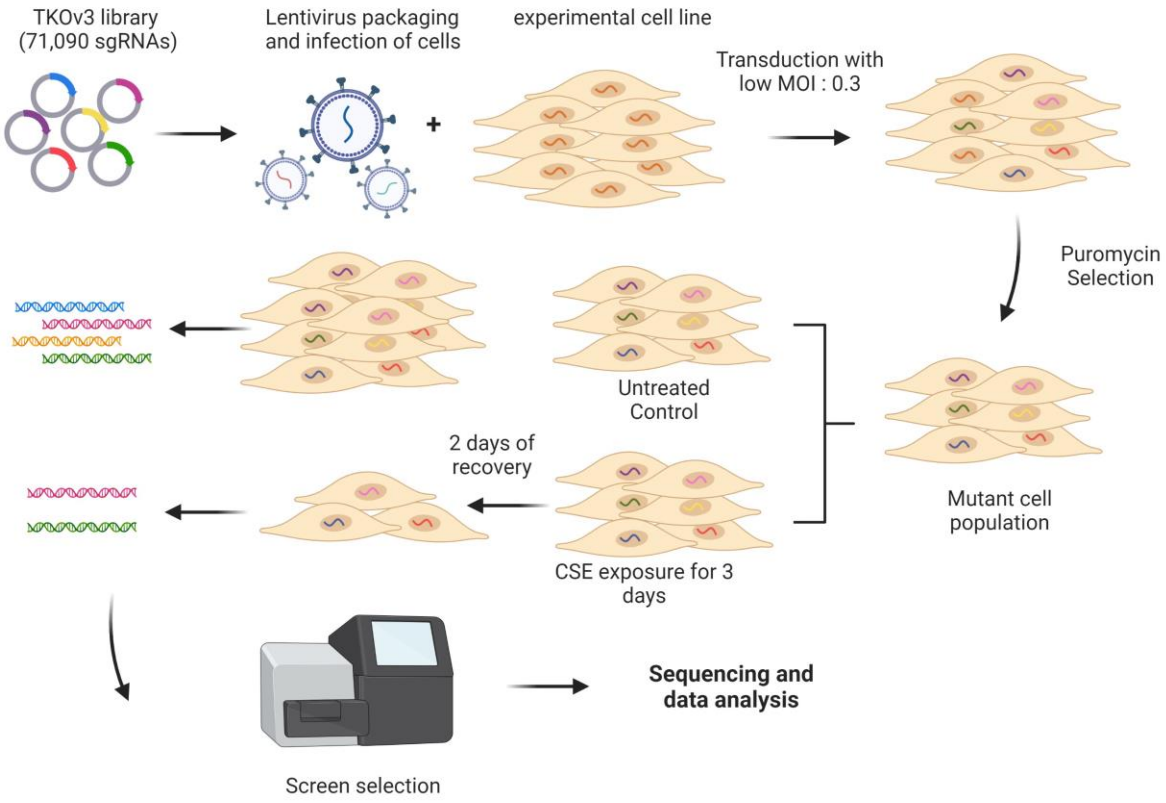

Figure 1.8: Workflow of genome editing library screening against a phenotype: *The schema represents the general workflow of the genome-wide CRISPR/Cas9 knockout library. Thousands of sgRNAs containing human genome-wide CRISPR/Cas9 knockout library are loaded in a lentiviral particle, and transduction into the cell line is done at low multiplicity of infection (MOI). Selection of cells with sgRNA transduction was made using puromycin to create a mutant cell pool characterized as control and treatment groups. Genomic DNA extraction from both control and treated cells was then*

done, followed by sgRNA fragment amplification using PCR. The high-throughput screening was used to determine the sgRNA copy numbers, and analysis was done using the MAGeCK algorithm. (Created with BioRender.com)

1.8.3 CRISPR knockout libraries

Currently, some libraries are designed to cause knockout of gene expression modification at the genome level. With the advent of technology, various types of gene expression modifications can be introduced these days including activation (CRISPRa) and repression (CRISPRi). Based on the biological question, one can choose the right type of library for investigation. For knock-out (KO) or loss-of-function (LOF) studies, it is highly recommended to use the early and constitutively expressed exons that guarantee the result of the production of non-functional protein.

For the *in-vitro* screening, the Zhang Lab and Sabatini/Lander Lab created the most commonly used genome-wide CRISPR knockout libraries (188, 201). More than 18,000 protein-coding genes have been targeted in the human genome through initial versions of knockout libraries (186, 188, 191). The Zhang lab also has the genome-wide CRISPR knockout library known as GeCKO, which contains additional sgRNAs that target more than 1,000 miRNAs (186). Each group developed methods to identify the genome-wide sgRNA target sequences for every gene, with a particular emphasis on the 5' constitutively expressed exons (201). The sgRNA design rules incorporated by the Sabatini/Lander lab were to reduce the disease's off-target effects and to enhance cleavage efficiency (191, 202).

Recently, Hart et al. described a 2nd generation CRISPR KO library, Toronto Knockout (TKO) library (153) that utilizes only effective sgRNAs with 0-1 genomic off-target sites (203). This library is available in two formats: 1) as a one-component library that contains Cas9, alleviating the need to include a separate Cas9, and 2) as a two-component without Cas9, which can be

utilized in cells that stably express Cas9. We used the TKO version 3 one-component library in our screen in this thesis for a genome-wide knock-out study.

The CRISPR libraries are also available as predesigned sub-pool libraries that target only genes with similar functions (202, 203). Such libraries are more focused and cost-effective and have more advantages in terms of use.

1.8.4 Positive and negative screens

There are two broad categories of pooled screening assays: negative and positive selection screens.

A negative selection screen aims to detect perturbations that impair cell survival or proliferation, leading to the depletion of the perturbed cells during the selection process. Here, two sets of cell populations are transduced, with one set being selected and the other acting as a non-selected control. Next, an analysis of the gRNA abundance in both groups is performed to determine which gRNAs have been depleted as a result of selection. To find the genes needed for cell proliferation, one of the simplest types of negative selection screens involves keeping cells growing for a long time. This method has been used to identify genes that are crucial for particular cell types as well as to profile genes that are generally necessary for cell viability (202-208). When compared to knockdown procedures that use both CRISPRi and short-hairpin RNA, the CRISPR-KO approach has been shown to perform better with low noise, limited off-target effects, and experimental consistency, notably in lethality-based critical gene screens (209). The fact that a perturbation can only be reduced to the degree that it was present in the initial library is a constraint on negative selection tests by definition. This indicates that negative selection screens may have more signal when the library size is modest with a corresponding larger relative abundance of each disturbance.

However, positive selection screens depend on cell enrichment. Positive selection screens apply a high level of selective pressure, making it unlikely that cells will be chosen in the absence of genetic disturbance. These screens have been used to find perturbations that confer resistance to pathogen infections (210-217), drugs (186, 218), and toxins (219). Positive selection screens, in contrast to negative selection screens, typically have a strong signal because resistant cells can be easily identified since the abundance of relevant gRNAs in these screens is higher than that of the other gRNAs. As a result, results from positive selection screens are frequently easier to understand than those from negative selection screens.

Both positive and negative screens can identify true positives. Still, a negative screen is much more technically challenging as it requires deep sequencing and rigorous control samples for identifying significant alterations in sgRNA representation when the survival of the cells is higher in the screen.

1.8.5 High-throughput CRISPR screening in the respiratory system

High-throughput screening for the identification of new genes linked to respiratory diseases is made easier by CRISPR (188, 220, 221). Genes that influence lung epithelial cells' susceptibility to influenza and picornavirus infection have also been identified through CRISPR knockout screens, along with elements linked to lung tumorigenesis and epidermal growth factor receptor (EGFR) dependence in Non-small cell lung cancer (NSCLC) (222-225). Genes found in such preclinical screens can be validated to uncover mechanisms underlying respiratory illnesses and facilitate novel treatment approaches. Because CRISPR-based screens are versatile, they can be used for transcriptional activation, repression, and epigenetic editing using dCas9, in addition to gene knockout or loss-of-function screening (226, 227). The sophistication of readouts from CRISPR screens has been significantly increased by several recent techniques that combine droplet-based single-cell RNA sequencing with pooled sgRNA

libraries (228, 229). In the future, the study of respirology is likely to benefit greatly from the use of this unique technique, particularly in the identification of genes responsible for complex features like immunological responses linked to asthma or respiratory distress.

Cumulatively, the pathogenesis of COPD is incompletely understood, hindering the development of new treatments. Despite extensive research efforts, no therapies currently exist to halt or reverse COPD progression. In spite of many studies highlighting the harmful effects of smoking, it remains the primary cause of COPD. While researchers have investigated COPD's pathogenic mechanisms at various molecular levels, the identities and dynamics of key molecular players remain unclear. Thus, urgent identification of therapeutic targets for COPD's molecular drivers is needed.

1.9 Aims

Overall, the entire purpose of conducting our studies was to identify critical therapeutic targets that have the potential to help abate CS-induced damage in COPD utilizing current molecular techniques and knowledge. Thus, to develop a research framework to elucidate the roles and therapeutic targeting of molecular drivers of COPD/emphysema, our aims were to:

1. *Investigate the changes in gene expression during initial smoke exposure and disease development in experimental CS-induced COPD.*

Most COPD transcriptional studies rely on cross-sectional analysis, offering valuable insights into disease status. However, such studies lack information on disease progression. Longitudinal studies, which track changes over time, are deemed crucial for understanding COPD dynamics. Implementing longitudinal studies in humans is challenging due to ethical and legal considerations. Thus, to enhance our understanding of COPD pathogenesis, we examined the transcriptional responses of lungs over time in a cigarette smoke (CS)-induced

experimental COPD model. We employed microarrays for gene expression profiling and bioinformatics for data analysis. Subsequently, we investigated the identified target's role using CRISPR/Cas9, followed by RNA-sequencing of knockout cells to profile expression changes in response to CS.

2. *Integrate the gene and miRNA expression patterns to define mechanisms that control disease pathways in experimental COPD.*

miRNAs are small non-coding endogenous RNAs that play a crucial role in regulating gene expression. However, the interactions of miRNAs with their target genes that contribute to COPD development remain incompletely elucidated. It is commonly recommended that multi-omics research offers more comprehensive, and reliable results compared to single-omics studies. For this reason, we investigated the longitudinal effects of CS to the COPD development and progression at the transcriptional level in combination with post-transcriptional changes in CS-induced experimental COPD. Thus, we integrated the microarray-based miRNA expression profile from CS-induced experimental COPD model with gene expression profile through bioinformatics analysis.

3. *Define potential novel and crucial targets for resistance against challenges with cigarette smoke extract (CSE).*

Both alveolar and endothelial apoptotic cells were found to be increased in the lungs of COPD patients, causing lung tissue destruction that leads to the development of emphysema. The potential regulatory mechanisms of CS-induced lung structural cell death are complicated and need to be fully understood to maximize the resistance against CS -induced cell death. Also, CS is a good experimental stimulus as research-grade cigarettes are highly quality-controlled. Thus, we hypothesized that genome-wide screen against CS exposure will help determine the genes that are protective against CS. To achieve this, we performed Genome-wide CRISPR-Cas9 knockout screens in human epithelial cell lines (A549 and BCi NS1.1).

Chapter 2: Longitudinal analysis of genome-wide expression profiling of cigarette smoke-induced experimental COPD

2.1 Abstract

Introduction: Chronic obstructive pulmonary disease (COPD) is a multifactorial complex inflammatory disease of the lungs induced by chronic exposure to inhaled irritants that have no effective treatment. Its mechanisms of pathogenesis remain incompletely understood and this has hampered the development of new treatments. Investigation of the transcriptional responses of the lungs over time in cigarette smoke (CS)-induced experimental COPD can provide novel insights into the molecular mechanisms that underpin the pathogenesis of the disease.

Methods: We employed our well-established short-term mouse model of experimental COPD induced by 8 weeks of CS exposure that replicates the characteristic features of the human disease. Gene expression profiling was performed at 4, 6, 8, and 12 weeks of exposure using microarrays. We first examined the initial smoke effect on gene expression at 4 weeks and then compared the changes in expression with those over time to define the essential genes and pathways involved in the development of COPD. CRISPR-Cas9 directed knock-out (KO) of an identified target candidate i.e. aryl hydrocarbon receptor nuclear translocator 2 (*ARNT2*) in the BEAS-2B cell line, was exposed to 2.5% CS extract for 96 hours. RNA sequencing was performed to investigate the impact of *ARNT2* KO on genome-wide gene expression. Subsequently, the potential downstream factors involved in the effects were validated using qPCR. Cell viability and the apoptotic study were performed at different concentrations of CS extract (CSE), including 1-, 2.5-, 5-, 7.5-, and 10%.

Results: We identified 282 genes and 115 genes that were altered by 4 weeks of CS exposure and over time with fold change (FC) >|2| and false discovery rate (FDR) < 0.05, respectively.

The protective genes against CS, including *Cyp1a1* that were high during initial smoke exposure, reduced over time. However, the expression level of regulatory genes of smoke protective mechanism including aryl hydrocarbon receptor nuclear translocator (*Arnt*) and aryl hydrocarbon receptor (*Ahr*) remained unchanged. Interestingly, *Arnt2*, a homolog of *Arnt*; was significantly elevated over time with CS exposure (FDR<0.05). An *ARNT2* KO cell line exposed to CSE had increased *CYP1A1* expression compared to wild-type cells ($p<0.05$) after 96 hours of exposure. A cell viability assay showed that the KO cells were significantly protected against death after 24 hours of exposure to CSE.

Conclusion: We defined the dynamic interplay of smoke protective genes between two phases of COPD onset and progression becomes impeded with chronic 6-8 weeks of CS exposure. We identified *ARNT2* as a potential repressor of smoke-protective genes. *ARNT2* may be a novel therapeutic target and its inhibition may protect against the detrimental effect of CS-induced cell death in COPD.

Keywords: COPD, expression profiling, CRISPR-Ca9, xenobiotic response elements

2.2 Introduction

Chronic obstructive pulmonary disease (COPD) is a progressive inflammatory disease of the lungs characterized by airflow limitation that is not fully reversible and severe breathing difficulties (108). It is the 3rd leading cause of morbidity and mortality worldwide (230-232). It is a complex multifactorial disease, and cigarette smoke (CS) is a common risk factor (233-235). Currently, no treatment stops the progression or reverses COPD, and the focus of therapy is symptom management. COPD is commonly defined by impaired lung function with a forced expiratory volume in 1 second (FEV_1) less than 80%, the FEV_1/FVC ratio is less than 70% and is also assessed through the diffusing lung capacity for carbon monoxide (DL_{CO}) (108).

The pathogenesis of COPD is incompletely understood, as it is a condition with many interlinked pathogenic mechanisms, including cell death, oxidative stress, and inflammation (236). Oxidative stress has previously been linked with COPD pathogenesis by initiating and mediating different gene expression and redox-sensitive signal transduction pathways (236, 237). Studies have also reported the involvement of various other cellular processes in COPD, including the aryl hydrocarbon receptor (AhR) pathway (238). *AhR* is a ligand-activated transcription factor that uses several external ligands, including tobacco smoke, to activate detoxification pathways (239). Tobacco exposure causes chronic AhR activity, leading to mitochondrial impairment, atrophy, and neuromuscular junction degeneration (239). Upon *AhR* ligand binding, the activated *AhR* dissociates from the AhR complex of molecular chaperones, including *XAP2 (AIP)*, *Hsp90*, *Src*, and *P23*, and subsequently translocates to the nucleus, where it binds with aryl hydrocarbon receptor nuclear translocator (*ARNT)* to form a functional *AhR/ARNT* complex. The binding of the AhR/ARNT complex to xenobiotic response elements (XREs) regulates the expression of downstream xenobiotic response elements or oxidative stress-related genes, including *CYP1A1* and *CYP1B1*. *AhR* ablation is

associated with diminished endogenous antioxidants necessary to combat oxidative stress (240).

Most COPD transcriptional studies are based on cross-sectional analysis that provides essential information regarding disease status. However, such studies lack information on disease progression over time; therefore, longitudinal studies are more helpful in understanding the early clinical signs of COPD and bring forward further knowledge for the developmental and progressive phases of the disease. Designing longitudinal studies in humans is complicated due to ethical concerns and legal ramifications. Animal models have been proven helpful in addressing various scientific questions. Thus, to improve our understanding of the pathogenic mechanism of COPD, we investigated a unique short-term mouse model of CS-induced experimental COPD that we developed and recapitulates the characteristics features of human disease after 8 weeks of nose-only CS exposure (102, 105-107, 241-243). Exposure is representative of a pack-a-day smoker (108). The ability to define the differentially expressed genes over time course in experimental COPD can be used to understand better how chronic CS exposure alters the expression of interacting molecules that contribute to disease development (244).

Thus, in the current study, we elucidated changes in gene expression in CS-induced experimental COPD during initial smoke exposure and disease development and progression. We focused on *AhR* signaling and its downstream physiological processes to understand the biological systems affected, including xenobiotic metabolism and oxidative stress. We observed that the expression of detoxification and oxidative stress-related genes increased during initial CS exposure followed by downregulation as the disease developed. This identifies new potential therapeutic targets and biomarkers of COPD, which can be modulated

to comprehensively elucidate the regulatory changes associated with the development and progression of COPD and their potential for therapeutic modulation.

2.3 Methodology

An overview of the experimental design is shown in Figure 2.1. Detailed procedures of mouse exposure to cigarette smoke (CS), data generation, and processing and procedure for the *in-vitro* functional study are described in the supplementary method section.

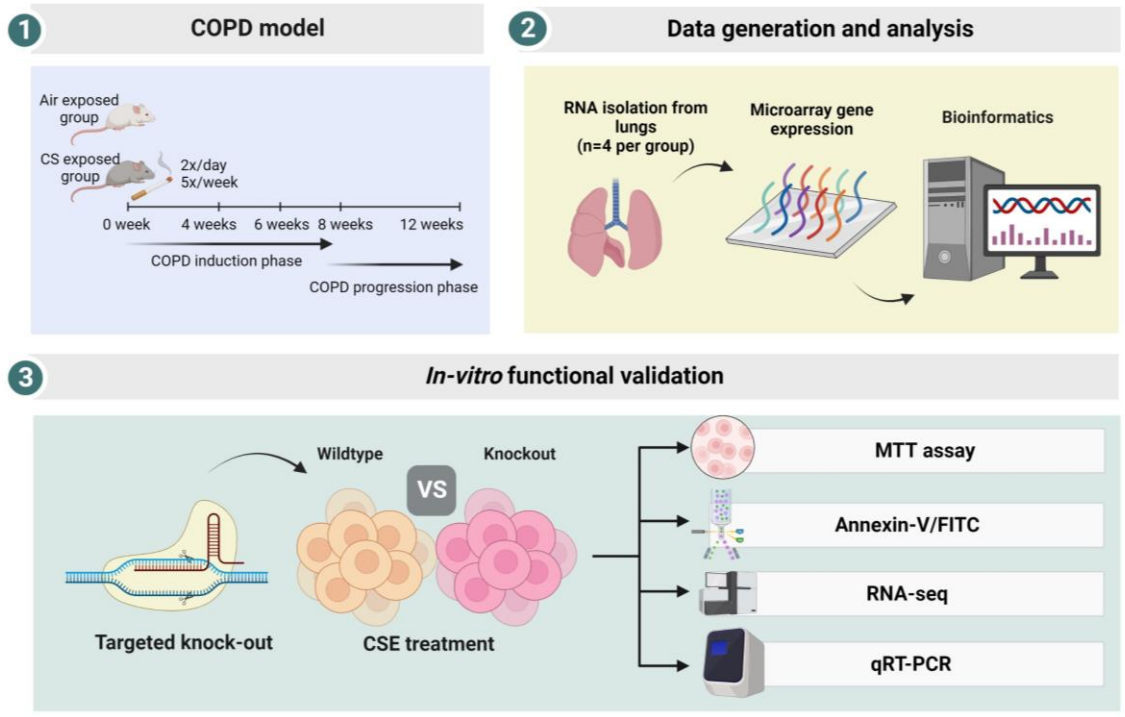

Figure 2.1: Workflow of expression profiling in CS-induced experimental COPD and *in-vitro* functional analysis using CRISPR-Cas9: ***1)*** *Tightly controlled amounts of CS were delivered to mice for 1 to 12 weeks via nose-only exposure, of which up to 8 weeks of exposure is the COPD induction phase and 8 to 12 weeks are termed as the COPD progression phase.* ***2)*** *Total RNA was collected from lung tissues of air vs. CS-exposed mice at each time point for expression profiling.* ***3)*** *In-vitro functional analysis begins with the creation of ARNT2 knockouts via CRISPR-Cas9 and is exposed to cigarette smoke extract (CSE) that was further investigated through various molecular techniques. (Created with BioRender.com)*

2.3.1 *In-vivo* elucidation of gene changes

Age- and sex-matched female BALB/c and C57BL/6 mice (n = 4-8/group), 6–8 weeks old were randomly assigned to be exposed to either normal air or CS. In the Hunter Medical Research Institute's (NSW, Australia) or Centenary Institute's (NSW, Australia) BioResources Facility, all animals were bred and kept in a particular pathogen-free barrier at 18–23°C and 40–60% humidity with a 12-hour light and dark cycle. The experimental groups and the control animals were co-housed in the same room. All mice were caged in individually ventilated cage (IVC) systems. All experimental procedures were approved by the institutional animal ethics committee under the National Health and Medical Research Council of Australia's Australian Code of Practice for the care and use of animals for research purposes.

2.3.2 Induction of cigarette smoke-induced COPD model

Mice were exposed to nose-only directed flow inhalation of twelve 3R4F reference cigarettes (University of Kentucky, Lexington, KY) for 12 weeks to induce the development (8 weeks) and progression (12 weeks) of experimental COPD (102, 103, 105-108, 245, 246), as explained in the appendix. Normal air-exposed mice were used as controls.

2.3.3 Data generation and processing

Lung tissue from BALB/c mice was collected from biological replicates of each group over a time course of 4-,6-,8- and 12 weeks of CS exposure. Total RNA was extracted and samples with RNA integrity number (RIN) >8 were profiled using the Affymetrix Mouse 430 plate array platform as explained in the appendix. Subsequently, the longitudinally generated expression data were analyzed with R statistical software (v3.5.2) (247, 248) with RStudio version 1.3.1 (249) and Bioconductor (250). Detail information on data analysis is in the appendix, briefly, differential expressions of RMA (Robust Multi-Array Average) normalized

genes after batch effect correction with the ComBat (251) function of sva package (252) (https://bioconductor.org/packages/sva) were analyzed using the "limma" package (doi:10.1093/nar/gkv007) (253, 254) of R. The data were corrected for multiple tests using the Benjamini–Hochberg (BH) false discovery rate (FDR). FDR < 0.05 and fold change (FC) >|2| were defined as cut-off values for identifying significantly differentially expressed genes (DEGs). Cross-sectional study at initial smoke exposure of 4 weeks and longitudinal changes in expression from 4 to 12 weeks was measured using the following equations.

$$Initial\ smoke\ exposure\ effect\ (cross-sectional) = Smoke04weeks - Air04weeks$$

(eq.1)

$$Longitudinal\ effect\ of\ smoke\ = (Smoke12weeks - Smoke04weeks) -$$

$$(Air12weeks - Air04weeks)$$ (eq.2)

Gene expression signatures in each sample were analyzed by gene set variation analysis (GSVA) using the GSVA R package (https://doi.org/10.1186/1471-2105-14-7) in terms of enrichment score (ES) (255).

2.3.4 Correlation with protein profiling

Fractions of proteins from whole lung homogenates taken from experimental COPD were quantified by high-resolution nano liquid chromatography-tandem mass spectrometry for proteome profiling (nLC-MS/MS) by David et al., (256). David et al., employed quantitative label-based proteomics to map the changes in the lung tissue proteome of our COPD model that has been published in 2021 (256). For detailed explanation of protein quantification methodology please see appendix. Using that proteomic profile, the relationships between top

DEGs with their translation into the protein over time were also investigated. For this, FC expression of genes was compared with normalized peptide abundance levels of proteins.

2.3.5 Characterization of functionally enriched DE genes

Potential interactions among longitudinal DEGs were identified by using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) v11.0 (https://string-db.org/) (257), as described in the appendix. Functional enrichment analysis was performed by mapping DEGs to known functional sources through g:profiler (https://biit.cs.ut.ee/gprofiler/gost) (258, 259).

2.3.6 Generation of knock-out (KO) cells by CRISPR-Cas9

For CRISPR-cas9 gene modification in a normal human bronchial epithelium cell line (BEAS-2B), paired gRNAs for *ARNT2* were designed using benching software (https://www.benchling.com/) (260). The CRISPR gRNA was ligated into a human codon-optimized pSpCas9 and chimeric gRNA expression plasmid containing a 2AEGFP (pSpCas9(BB)-2A-GFP (PX458)) insert (plasmid 48138, Addgene). Cells were seeded and cultured for 24 hours (to $1x10^4$/ml) and were transfected with an *ARNT2*-targeting EGFP-CRISPR-cas9 constructed plasmid. The $EGFP^+$ cells were sorted using a FACS AriaII flow cytometer. DNA was extracted from the sorted cells and sequenced to confirm biallelic KO clones. The allelic profile, as well as indel patterns of each clone, were analyzed using tracking of indels by decomposition (TIDE) (https://tide.nki.nl/) (261, 262) and inference of CRISPR edit (ICE) (https://www.synthego.com/products/bioinformatics/crispr-analysis) (263, 264) software programs. Further details of cell culturing and generation of knock-outs are in the appendix.

2.3.7 Long-term exposure of *ARNT2* KO cells with CS extract (CSE)

Smoke from one filter-less 3R4F reference cigarette (12mg tar and 1.0mg nicotine/cigarette) was continuously drawn through 10 ml of pre-warmed serum-free culture media. The resulting suspension was then filtered through a 0.22um pore-size filter. The obtained solution was considered 100% CSE and used within 20 minutes of preparation. The resulting 100% CSE was diluted to the desired concentration with 1% fetal bovine serum (FBS) culture medium. WT and *ARNT2* KO Beas-2B cells at 80% confluency were incubated in serum-free culture media for 24 hours before CSE exposure. Cells were exposed to 2.5% CSE in 1%FBS culture media for about 96 hours. CSE media was refreshed after every 24 hours.

2.3.8 Impact of CSE on Cell viability and apoptosis

Cell viability and apoptotic behavior of the KO cells in response to different concentrations of CSE exposure after 24 hours were assessed by 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay and an Annexin V apoptotic assay using Recombinant Annexin V conjugated to FITC (BD Bioscience cat# 556420) and 7AAD fluorescent stains, respectively as described in the appendix.

2.3.9 Library preparation and RNA sequencing

For RNA sequencing total RNA was extracted and purified using RNeasy Mini Kit (50) (Qiagen Cat. No. ID: 74104). The extracted RNA was DNase-treated to digest the remains of DNA from the samples. The quality and integrity of RNA were assessed using the NanoDropTM, ND1000 spectrophotometer, Agilent 2100 Bioanalyzer, and RNA Nano 6000 Assay Kit. Biological replicates of each group were pooled together with the same concentration of 60ng/ul for each replicate. Samples were prepared by Illumina® Stranded mRNA Prep, ligation kit and sequenced with NextSeq500 High Output 1x75bp flow cell. Delta

analysis was performed by comparing the KO CSE exposed cell with WT CSE cells while removing the WT CSE effect ((KO.CSE - KO.CTL) - (WT.CSE-WT.CTL)) as detailed in the appendix section.

2.3.10 qPCR validation

For validation, extracted RNA from BEAS-2B cell lines was reverse transcribed to cDNA and quantified using real-time qPCR for identified targets (see Table S2.1).

2.3.11 Statistical analysis

The data was assumed to be normally distributed. Statistical testing methods were performed using GraphPad Prism (v8) (265). Comparison between different groups was performed by using one-way or two-way ANOVA followed by multiple comparison tests, as appropriate.

2.4 Results

A high level of concordance for each group (n=4 mice per group) (i.e., air-exposed control groups, 4-6 weeks of smoke exposure (COPD induction phase), and 8-12 weeks of smoke exposure (COPD progression phase) was revealed through principal component analysis (PCA) (Figure 2.2A).

2.4.1 Sub-chronic CS exposure alters the gene expression profile

Changes in gene expression during initial exposure of 4 weeks comparing smoke (n=4) to air-exposed mice were (n=4) assessed to investigate the smoke effect on gene expression before the development of chronic disease features. We identified 282 genes (termed the "smoke-affected genes") that were significantly affected by initial smoke exposure (FC >|2| and FDR <0.05, (Figure 2.2B). Among them, expression of 218 and 64 genes were increased and

decreased, respectively. Among the identified list of smoke signatures, we found *Cyp1a1* as the most significantly highly expressed gene (FC = 7.4 and p = 1.69^{e-24}, (Figure 2.2B-C). The top 50 genes differentially expressed in this analysis are shown in a heatmap (Figure 2.2C). The complete list of smoke signatures is in Table S2.2.

2.4.2 Association of gene expression with disease development in CS-induced experimental COPD

Next, we investigated the transcriptional changes during disease development by performing a longitudinal gene expression analysis in CS-induced experimental COPD. The influence of disease development on gene expression was obtained by comparing the expression profile from week 4 to week 12 in the CS-exposed group (n=4 at each time point) relative to the air-exposed mice group (n=4 at each time point). We found 115 genes (termed: "disease-related genes") were differentially expressed over time from 4 to 12 weeks relative to the air-exposed group, including 79 upregulated and 36 downregulated genes (FC>|2| and FDR <0.05, (Figure 2.2D-E). Interestingly, in this analysis, we identified *Cyp1a1* as one of the most significant genes with downregulation (FC = -5.91 and p = 1.16^{e-17}). The complete list of dysregulated disease-related genes is in Table S2.3.

We further explored the shared and unique DEGs identified during our two analyses i.e., initial smoke exposure and longitudinal smoke exposure. We found 31 genes in common and 84 unique genes that were expressed differentially only during disease development and progression in our model (Figure 2.2F). Among 31 common genes; one is *Cyp1a1*, the most significant DEG in both analyses. We plotted a cross-sectional and longitudinal expression profile to visualize the change in the expression pattern of *Cyp1a1* over time during CS exposure (Figure 2.2G).

2.4.3 Altered longitudinal gene signatures define the disease progression phase

To better understand the differential gene expression pattern set at each time point, we transformed the expression profile into an ES using GSVA. We observed a significant decrease in the expression of the downregulated smoke signature at the COPD induction phase (4- and 6- weeks as well as COPD development (8 weeks) and progression (12 weeks) phases. Similarly, the upregulated smoke signature was significantly high at all time points, as shown in Figure S2.1A. Interestingly, GSVA of the disease signature identified during longitudinal analysis showed a noticeable difference in expression profiles between the COPD induction phase (4-6 weeks) and the COPD progression phase (8-12 weeks). The upregulated disease signature prominently indicates high expression levels during the phases of disease development and progression (8 and 12 weeks), contrasting with the lower expression levels observed during the induction phase of COPD (4 and 6 weeks) (Figure S2.1B). Conversely, the downregulated disease signature exhibits decreased expression levels, notably during the phases of disease development and progression (8 and 12 weeks), contrasting with the elevated expression preceding disease onset (4-6 weeks) compared to air-exposed mice. These temporal patterns underscore the significance of the identified disease signature, particularly its association with disease development, rather than merely reflecting the effects of smoke exposure Figure S2.1B).

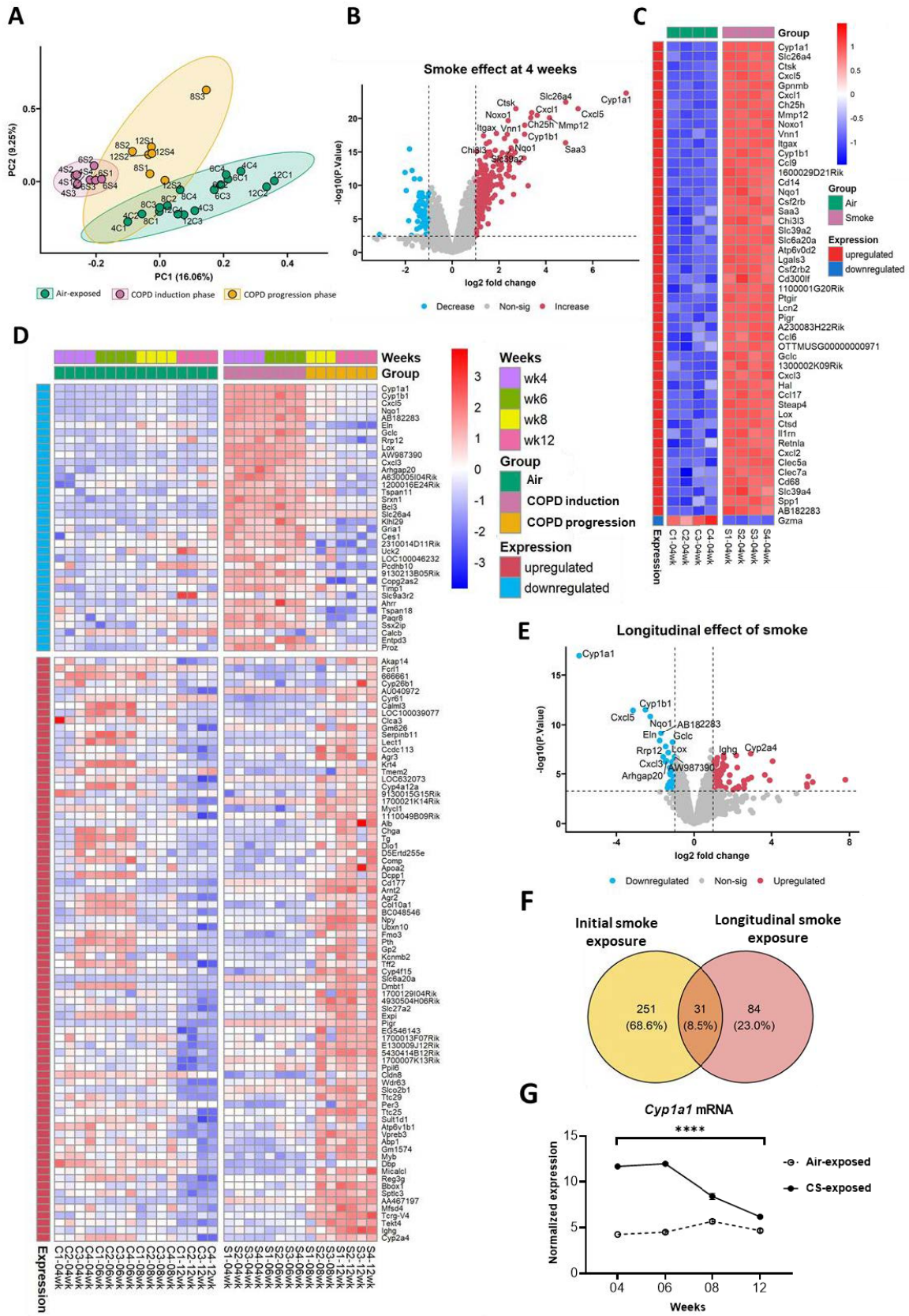

Figure 2.2: Gene expression profile in CS-induced experimental COPD: ***A)*** *Principal component analysis (PCA) plot for quantile plus combat normalized data showed that air-exposed control group*

(n=16; 4 mice at each timepoint 4, 6, 8 and 12 weeks), COPD induction phase group (n=8; 4 mice at timepoint 4 and 6 weeks), and COPD development and progression phase group (n=7-8; 4 mice at timepoint 8 and 12 weeks) are clustered separately, ***B)*** *Volcano plot,* ***C)*** *Heatmap for DEGs identified during the smoke effect of 4 weeks.* ***D)*** *Heatmap,* ***E)*** *Volcano plot for DEGs identified due to longitudinal effect of smoke from 4 to 12 weeks that leads to COPD development and progression. Plot representing the significance ($-\log_{10}(P)$) vs $\log_2(FC)$ on the y-axis and x-axis, respectively. The horizontal dotted line indicates the FDR cut-off of 0.05 and the vertical lines indicate the $\log_2(FC) > |2|$. The heatmap represents the expression of DEGs in each sample. Each row in the heatmap represents a DEG, where each column indicates a sample in the study. The level of gene expression was scaled from high (red) to low (blue).* ***F)*** *The Venn diagram represents the number of common factors identified during initial smoke exposure and the longitudinal smoke exposure effect.* ***G)*** *Normalized expression of Cyp1a1 over time from 4 to 12 weeks of exposure. **** $p < 0.0001$.*

2.4.4 Gene expression changes in *Cyp1a1* are replicated at the protein level

We also observed the expression pattern of Cyp1a1 at the protein level over time and likely observed a significant reduction in expression during COPD progression (Figure 2.3A). For this, we compared the fold change protein abundance of Cyp1a1 protein over time. We were also able to quantify Nqo1 and Gclc, the two other most significant DEGs identified during longitudinal analysis (Figure 2.3B). Interestingly, we found a similar protein expression pattern matching gene expression for these genes due to the smoke and disease effects.

2.4.5 Smoke-protective genes modulate the disease progression

Protein-protein interaction (PPI) network of longitudinal DEGs through STRING interaction network analysis shows a significant protein-protein enrichment $p < 1.0^{e-16}$ with a medium confidence score of 0.4. The total number of nodes was 99, and the total number of edges (interactions) was 77, with an average local clustering coefficient of 0.324. STRING interaction

map was divided into 3 k-means clusters with a distinct set of DEGs (Figure S2.2A). Cluster 1 genes were enriched in response to xenobiotic stimulus and response to toxic substances pathways (Figure 2.3C). Details about all significantly enriched pathways and GO terms from the STRING database are given in supplementary Figure S2.2B-C and Table S2.4.

Functional enrichment analysis of longitudinal DEGs by g:GOSt module of g:profiler identifies the associated critical over-represented GO terms and pathways (Figure 2.3D-E). DEGs were mapped to 65 GO terms with a corrected p-value < 0.05. Approximately 20, 35, and 10 GO terms were grouped into the molecular function (MF), biological process (BP), and cellular component (CC) knowledge, respectively. The disease-associated genes were mapped to 06 KEGG, 06 REAC, and 05 Wiki pathways with adjusted p <0.05 (Figure 2.3D and Table S2.5). These include the gene sets linked to response to xenobiotic stimulus and estrogen metabolism, Phase I functionalization of compounds, and oxidative stress response (Figure 2.3D-E).

Pathway analysis indicates that the significantly enriched DEGs including *Cyp1a1, Cyp1b1,* and *Ahrr* belonging to XREs or regulated during oxidative stress response (Figure 2.3D-E) are the downstream effectors of the AhR pathway (266). This pathway is regulated with regulatory molecules, including *Ahr* and *Arnt*. Surprisingly, the expression of *Ahr* and *Arnt* remains unchanged in our COPD model (Figure 2.3F). However, close homologs such as *Arnt2* and *Ahrr* were significantly dysregulated in our COPD model. Among them, *Arnt2* shows increased expression with disease progression (Figure 2.3F). Whereas, *Ahrr* is downregulated substantially during disease development and progression like *Cyp1a1*, indicating the silencing of a protective response mechanism against smoke exposure leading to disease progression. Thus, *Arnt2* is an important upregulated biological molecule that might lead to a functional effect in CS-induced COPD pathology.

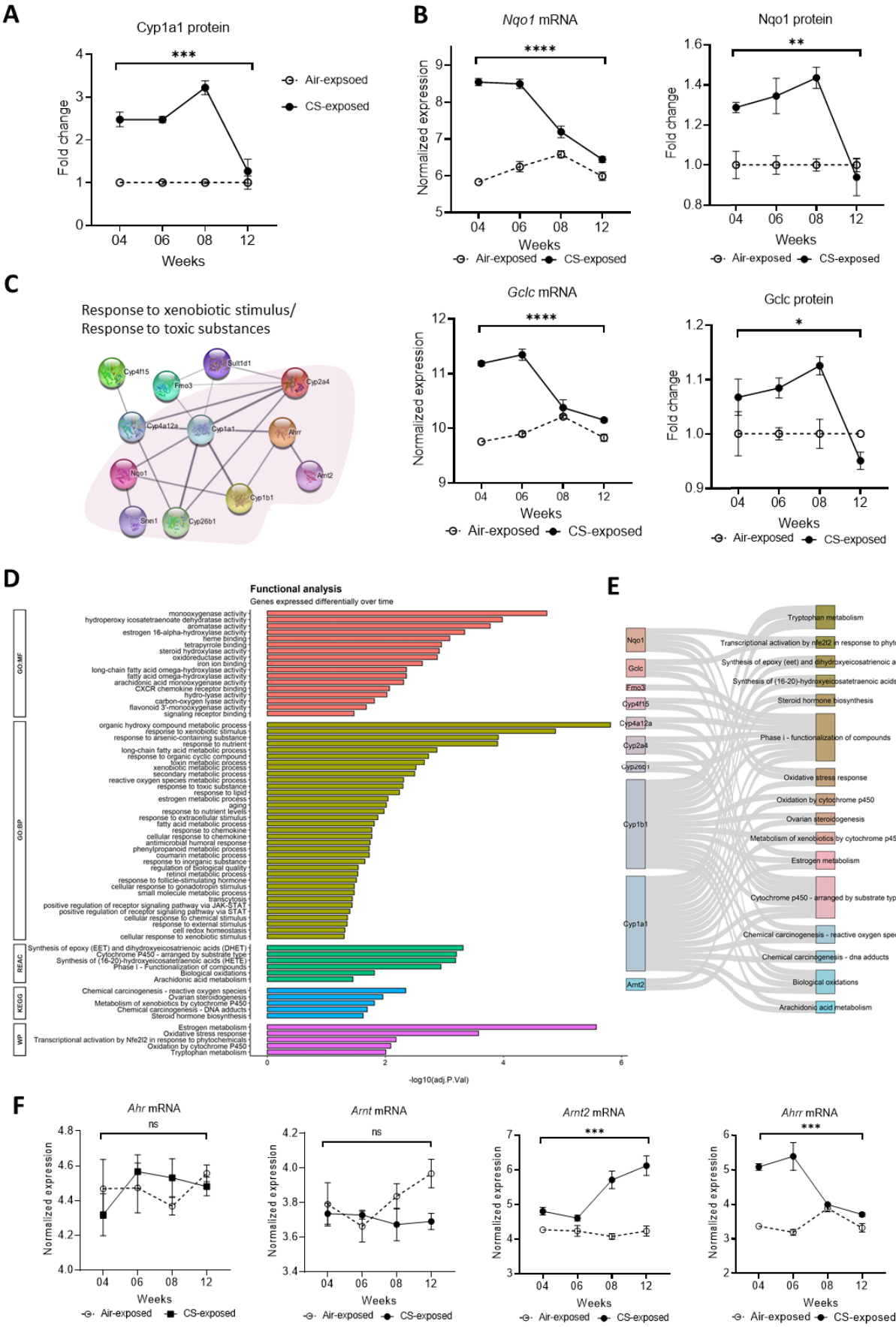
A
Cyp1a1 protein
***
Air-expsoed
CS-exposed
Fold change
Weeks
B
Nqo1 mRNA
****
Normalized expression
Nqo1 protein
**
Fold change
Weeks
Air-exposed
CS-exposed
C
Response to xenobiotic stimulus/
Response to toxic substances
Cyp4f15
Sult1d1
Fmo3
Cyp2a4
Cyp4a12a
Cyp1a1
Ahrr
Nqo1
Arnt2
Cyp1b1
Srxn1
Cyp26b1
Gclc mRNA
****
Normalized expression
Gclc protein
*
Fold change
Weeks
Air-exposed
CS-exposed
D
Functional analysis
Genes expressed differentially over time
GO:MF
monooxygenase activity
hydroperoxy icosatetraenoate dehydratase activity
aromatase activity
estrogen 16-alpha-hydroxylase activity
heme binding
tetrapyrrole binding
steroid hydroxylase activity
oxidoreductase activity
iron ion binding
long-chain fatty acid omega-hydroxylase activity
fatty acid omega-hydroxylase activity
arachidonic acid monooxygenase activity
CXCR chemokine receptor binding
hydro-lyase activity
carbon-oxygen lyase activity
flavonoid 3'-monooxygenase activity
signaling receptor binding
GO:BP
organic hydroxy compound metabolic process
response to xenobiotic stimulus
response to arsenic-containing substance
response to nutrient
long-chain fatty acid metabolic process
response to organic cyclic compound
toxin metabolic process
xenobiotic metabolic process
secondary metabolic process
reactive oxygen species metabolic process
response to toxic substance
response to lipid
estrogen metabolic process
aging
response to nutrient levels
response to extracellular stimulus
fatty acid metabolic process
response to chemokine
cellular response to chemokine
antimicrobial humoral response
phenylpropanoid metabolic process
coumarin metabolic process
response to inorganic substance
regulation of biological quality
retinol metabolic process
response to follicle-stimulating hormone
cellular response to gonadotropin stimulus
small molecule metabolic process
transcytosis
positive regulation of receptor signaling pathway via JAK-STAT
positive regulation of receptor signaling pathway via STAT
cellular response to chemical stimulus
response to external stimulus
cell redox homeostasis
cellular response to xenobiotic stimulus
REAC
Synthesis of epoxy (EET) and dihydroxyeicosatrienoic acids (DHET)
Cytochrome P450 - arranged by substrate type
Synthesis of (16-20)-hydroxyeicosatetraenoic acids (HETE)
Phase I - Functionalization of compounds
Biological oxidations
Arachidonic acid metabolism
KEGG
Chemical carcinogenesis - reactive oxygen species
Ovarian steroidogenesis
Metabolism of xenobiotics by cytochrome P450
Chemical carcinogenesis - DNA adducts
Steroid hormone biosynthesis
WP
Estrogen metabolism
Oxidative stress response
Transcriptional activation by Nfe2l2 in response to phytochemicals
Oxidation by cytochrome P450
Tryptophan metabolism
-log10(adj.P.Val)
E
Nqo1
Gclc
Fmo3
Cyp4f15
Cyp4a12a
Cyp2a4
Cyp2c9q1
Cyp1b1
Cyp1a1
Arnt2
Tryptophan metabolism
Transcriptional activation by Nfe2l2 in response to phytochemicals
Synthesis of epoxy (eet) and dihydroxyeicosatrienoic acids (dhet)
Synthesis of (16-20)-hydroxyeicosatetraenoic acids (hete)
Steroid hormone biosynthesis
Phase i - functionalization of compounds
Oxidative stress response
Oxidation by cytochrome p450
Ovarian steroidogenesis
Metabolism of xenobiotics by cytochrome p450
Estrogen metabolism
Cytochrome p450 - arranged by substrate type
Chemical carcinogenesis - reactive oxygen species
Chemical carcinogenesis - dna adducts
Biological oxidations
Arachidonic acid metabolism
F
Ahr mRNA
ns
Arnt mRNA
ns
Arnt2 mRNA
***
Ahrr mRNA
***
Normalized expression
Weeks
Air-exposed
CS-exposed

Figure 2.3: Functional enrichment analysis and expression pattern of DEGs over time in experimental COPD: ***A)*** *Expression fold change (FC) of Cyp1a1 at protein level over a time course from 4 to 12 weeks of exposure (n=4 per group),* ***B)*** *Expression pattern of Nqo1 and Gclc at gene level as well as at protein level (in terms of fold change) (n=4 per group) at 4, 6, 8 and 12 weeks of exposure. Vertical bars indicate the standard error of the mean (SEM).* ***C)*** *STRING interaction network between DEGs of cluster 1 identified during k-means clustering. Nodes represent the DEGs (protein-coding genes) and lines between nodes refer to edges. The pink circle region includes the proteins/genes belonging to biological processes such as response to a stimulus or toxic substances identified by STRING.* ***D)*** *Functional enrichment analysis of DEGs through g:GOSt module of g:profiler. The bar plot depicts the functional enrichment GO terms with the categories: molecular function, biological process, and associated critical pathways retrieved from Reactome, KEGG, and WikiPathways. The y-axis represents the source of the functional category, and the x-axis represents the enrichment significance in terms of* $-\log_{10}$*(P.value).* ***E)*** *The Sanky diagram highlights the DEGs enriched in identified pathways, particularly from KEGG, Reactome, and WikiPathways databases.* ***F)*** *The longitudinal expression pattern of potential regulators including Arnt, Ahr, Arnt2, and Ahrr at each time-points (4, 6, 8, and 12 weeks) indicates no significant difference in the longitudinal expression pattern of Arnt and Ahr (n=4 per group). Arnt2 expression increased with COPD development at 8 weeks, and Ahrr expression increased during initial smoke exposure but decreased during COPD development and progression. Vertical bars indicate the standard error of the mean (SEM). Significance was calculated using two-way ANOVA Tukey's multiple comparison test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.*

2.4.6 Functional effects of *ARNT2* in response to CS

To investigate the functional effect of *Arnt2*, we generated an *ARNT2* KO cell line in the human bronchial epithelial cells BEAS-2B using CRISPR-Cas9 editing technology (Figure 2.4A). Sequencing of isolated DNA from three different BEAS-2B clones (biological replicates) has

confirmed the successful deletion of the *ARNT2* gene due to the insertions and deletions of random bases in the *ARNT2* exon (Figure 2.4B and Figure S2.3).

To study the effect of smoke exposure on highly significant differentially expressed XRE, *CYP1A1,* in the absence of *ARNT2*, BEAS-2B cells were treated with 2.5% CSE for 96 hours. We observed a significant increase in the expression of *CYP1A1* in treated KO cells compared to treated WT cells during CS exposure after 96 hours (n=10). Also, CYP1A1 expression is significantly higher in treated WT and KO cells compared to untreated cells after 96 hours of CSE exposure indicating higher expression (Figure 2.4C). This suggests that ARNT2 might be essential in regulating the baseline CYP1A1 expression and during CSE response in BEAS-2B cells.

2.4.7 Deletion of *ARNT2* in KO bronchial epithelial cells exposed to CSE protects against cell death

We also investigated the effects on cell survival against the detrimental effect of CS in the absence of *ARNT2*. For this, we exposed the *ARNT2* KO cells to different concentrations of CSE for 24 hrs. Interestingly, the cell viability assay shows that exposing the KO cells to CSE protects against smoke exposure in the cells (n=6), particularly with 5% and 10% CSE (Figure 2.4D). This was also confirmed by comparing the percentage of viable cells via flow cytometry during annexin V-FITC 7-AAD assay (Figure 2.4E). Annexin V-FITC apoptosis staining assay also shows that the cells that underwent early, late apoptosis were significantly lower in KO cells than WT cells when exposed to 7.5% CSE (n=6) (Figure 2.4E-F).

2.4.8 RNA sequencing of KO cells

To investigate the transcriptome-wide gene expression changes due to knocking out the ARNT2, we performed an independent experiment for RNA sequencing of WT and ARNT2

KO cells after 96 hours of CSE exposure (n=10). During the delta analysis, we found 249 genes with logFC > 1 and 203 genes with logFC < -1 (Figure S2.4A). It's worth mentioning here that RNA sequencing of pooled samples from an entirely different experiment also shows the same trend of CYP1A1 expression in KO CSE exposed cells compared to WT CSE exposed cells (Figure S2.4B), which we observed previously in Figure 2.4C.

2.4.9 Functional enrichment analysis of *ARNT2* KO cells

Protein-protein interaction (PPI) network of identified dysregulated genes with logFC > |1| through STRING interaction network analysis shows a significant protein-protein enrichment p< 2.89e-15 with a confidence score of 0.4 (Figure S2.5). The total number of nodes was 310, and the total number of edges (interactions) was 342, with an average local clustering coefficient of 0.351. The interaction network was then further categorized into 5 different clusters using k-mean clustering by the STRING interaction network analysis program to identify the functionally important group of genes (Figure S2.5). Functional enrichment analysis of these clusters highlights the genes of clusters 4 and 5 as enriched genes in different biological processes. Protein-protein interaction enrichment for clusters 4 and 5 was p<4.4e-07 and p<1.0e-16 with 57 and 52 nodes, respectively (Figure 2.5A). We could not find enriched processes for clusters 1, 2, and 3. Cluster 4 and cluster 5 genes were enriched in 17 and 228 biological processes, respectively, with FDR < 0.05 (Table S2.6). The top 10 pathways were selected based on the number of observed enriched genes (Figure 2.5A). Cluster#5 indicates the processes, including cellular response to stimulus and regulation of biological processes (Figure 2.5B).

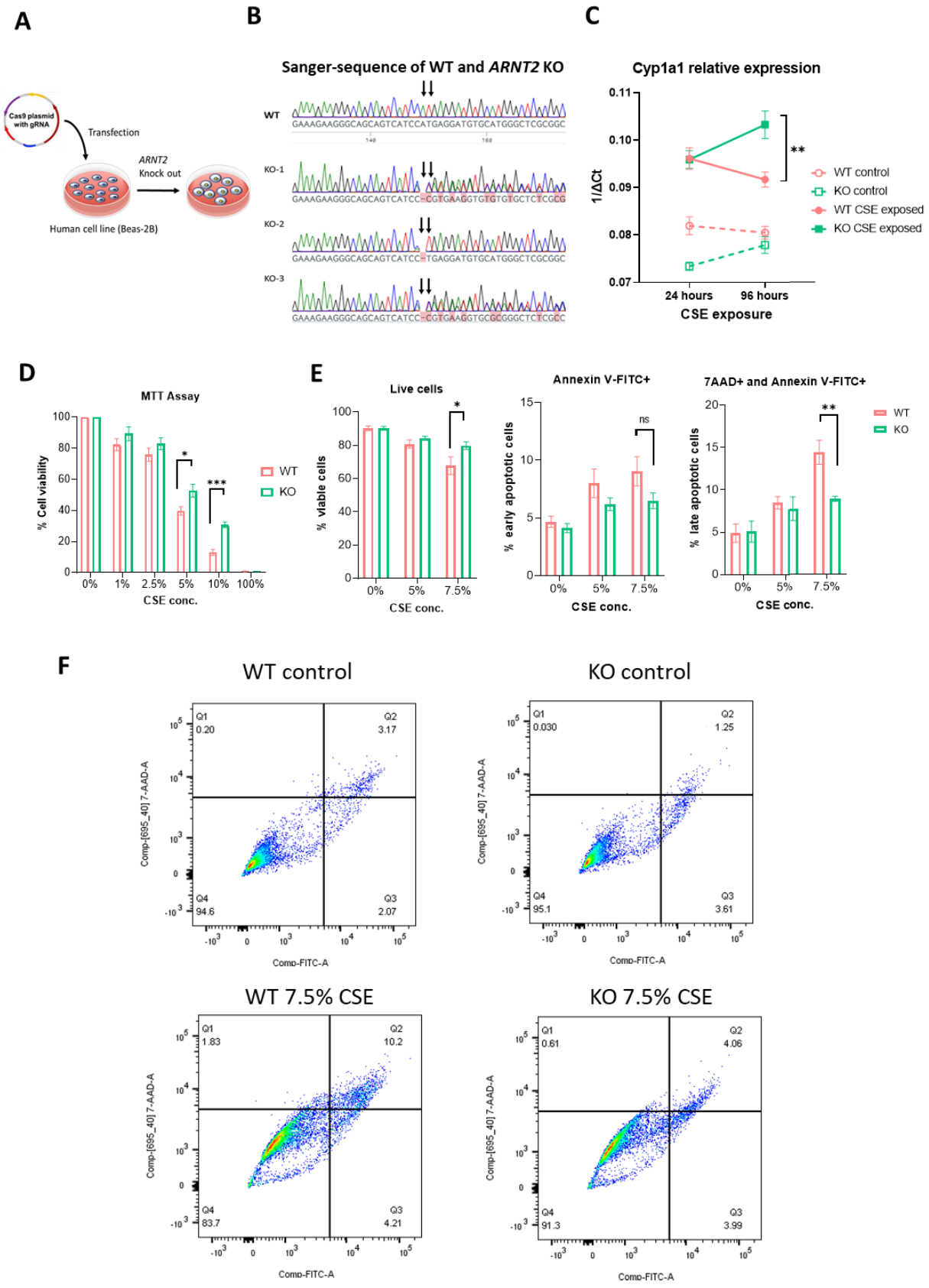

Figure 2.4: In-vitro functional analysis of CRISPR-Cas9-generated *ARNT2* knockout (KO) in the BEAS-2B cell line: ***A)*** *ARNT2 KO cells were generated using CRISPR/cas9 by transfecting the human*

Broncho epithelial cell line with a plasmid containing gRNA for targeting ARNT2. ***B)*** *The generated KOs were assessed by Sanger sequencing. Sequencing peak diagram of WT and mutant ARNT2 highlights the base pairs indels that disrupt the gene's reading frame. All 3 KOs were derived from 3 different single clones.* ***C)*** *Expression of targeted gene CYP1A1, quantified by RT-qPCR relative to the housekeeping gene GAPDH (n=10). The y-axis indicates the 1/delta Ct value. The higher the 1/delta Ct value higher the expression of the targeted gene in the sample. Vertical bars indicate the standard error of the mean (SEM) of three independent experiments each with at least three biological replicates. Significance was calculated using two-way ANOVA followed by Tukey's multiple comparison test.* ***D)*** *Cell viability MTT assay (n=6) shows a comparatively less detrimental effect of CSE on KO cells than WT cells. Vertical bars indicate the standard error of the mean (SEM) of two independent experiments with three biological replicates. Significance was calculated using two-way ANOVA, Sidak's multiple comparison test.* ***E)*** *Annexin V/7AAD assay (biological replicates (n=6)) shows that the percentage of viable cells is significantly higher than KO cells when exposed to 7.5% CSE. The assay also showed that ARNT2 knockout cells have a significantly lower cell apoptosis rate in Beas-2B cells upon 7.5% CSE exposure. Significance was calculated using two-way ANOVA, Sidak's multiple comparison test.* ***F)*** *A quadrant dot plot of the annexin V/7AAD staining assay results obtained by flow cytometry. The early and late apoptotic cells were exhibited as blue dots in the lower right quadrant (Q3) and upper right quadrant (Q2), respectively. The plot indicates the percentage of live and apoptotic cells. The plot indicates the percentage of live and apoptotic cells in WT and KO control and 7.5% CSE exposed cells. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.*

2.4.10 qPCR validation of potential hits

Before validation using qPCR, the results of pathway analysis and the literature mining were carefully scrutinized in addition to clustering to select the most representative clusters. To validate the expression of genes, in a more stringent analysis, we also considered the genes such that to be considered differentially expressed their normalized expression should be greater than 0.5 in each sample from the identified list of genes. In that way, we had only found

24 genes that we further performed the functional impact (Figure S2.6 and Table S2.7) as well the validation of a few candidates, including: (*ID1, ID2, HES1, KCNJ12, SMOC1, IL11RA, and RGS2*). Among them, *ID1, ID2, HES1, KCNJ12, IL11RA*, and *RGS2* show significantly high expression in KO CSE-exposed cells compared to WT CSE-exposed cells where SMOC1 shows significantly lower expression in KO-exposed cells as compared to WT-exposed cells (n=10-12 per group) (Figure 2.5C). We can see that KO and CSE alone have no significant effect on the selected candidates except *IL11RA* and *SMOC1* respectively, but the combined effect of KO and CSE shows a significant effect on the expression regulation of *ID1*, *ID2, HES1, KCNJ12, SMOC1, IL11RA*, and *RGS2* (Figure 2.5C). However, *KCNJ12*, *ID1,* and *ID2* showed low significance (Figure 2.5C).

Our investigations revealed that the knockout of *ARNT2*, coupled with exposure to CSE, led to the dysregulation of pathways associated with the cellular response and the regulation of biological processes. Furthermore, we observed that upon CSE exposure, *ARNT2* KOs resulted in the upregulation of genes that inhibit apoptosis. These findings suggest that targeting *ARNT2* may represent a promising therapeutic strategy for COPD, suggesting that interventions aimed at modulating *ARNT2* expression could hold promise in mitigating the pathological effects associated with CS exposure.

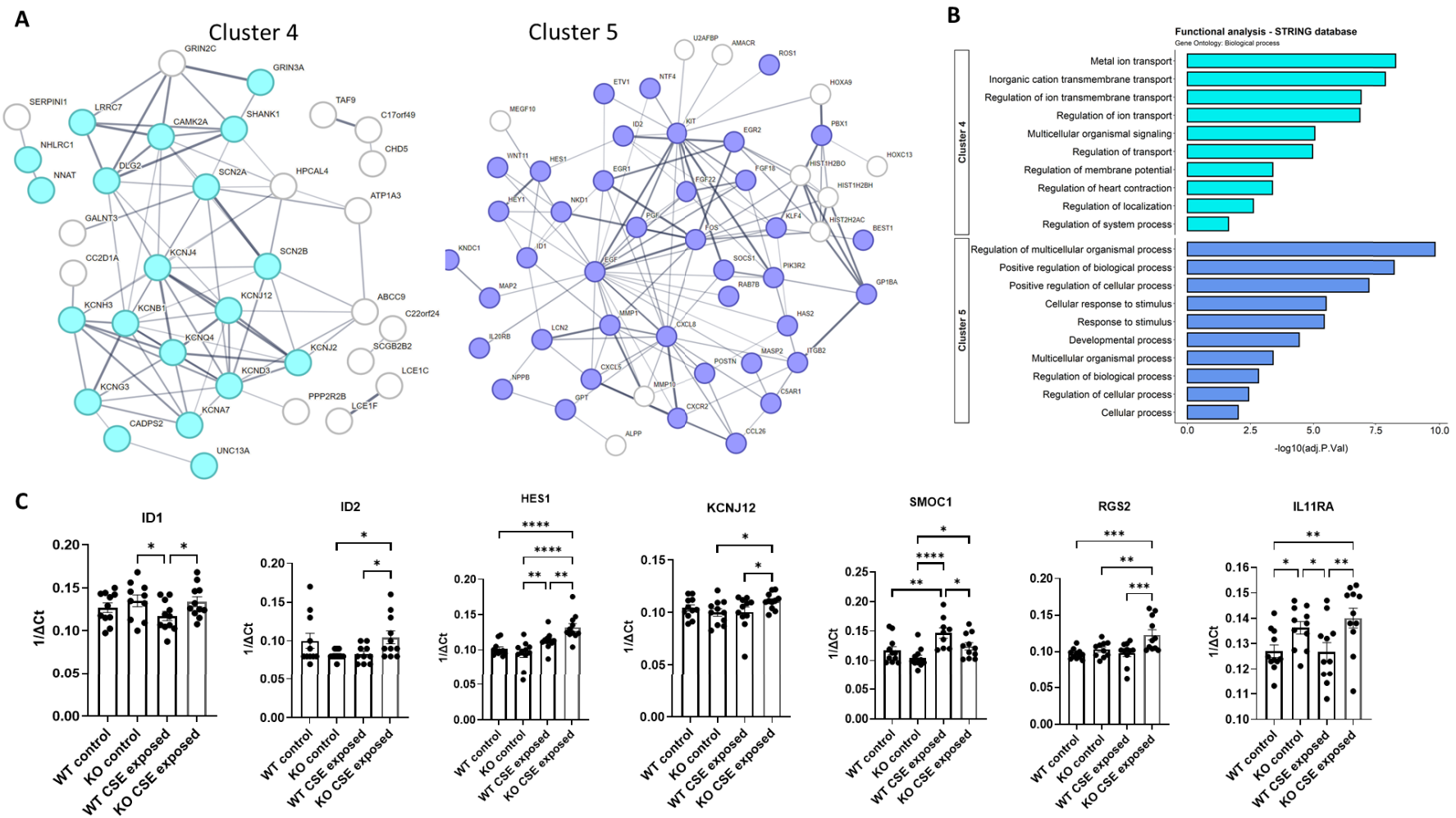

Figure 2.5: Functional analysis and validation of identified dysregulated genes in *ARNT2* knockout (KO) cells exposed to CSE: ***A)*** *Functionally enriched clusters (cyan: Cluster 4 and light purple: Cluster 5) with enriched genes (highlighted) identified through STRING using k-mean clustering. The solid line*

indicates the interaction between the nodes. ***B)*** *Significant enriched biological processes identified from the STRING database in cluster 4 (cyan) and Clsuter5 (light purple).* ***C)*** *Realtime-qPCR validation of identified hits shows a significant difference in the KO CSE exposed group compared to the WT CSE exposed group. Vertical bars indicate the standard error of the mean (SEM) of three independent experiments with at least three replicates. Significance was calculated using ordinary one-way ANOVA untested for multiple comparisons (n=10-12). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.*

2.5 Discussion

In the current study, we investigated a longitudinal gene expression profiling strategy in a mouse CS-induced experimental COPD to examine the smoke effect and the disease effect on the expression level of genes. We simultaneously focused on the changes during the induction and progression phases of CS-induced experimental COPD. It delineated the contribution of genes due to the smoke effect versus disease effect, shedding light on the dynamic interplay between two phases in COPD onset. We found that the gene expressed high in response to CS exposure went down during disease development and progression. Likewise, the downregulated genes increased during disease development and progression due to the smoke effect. The most significant DEGs were XREs and oxidative stress-related genes. Furthermore, the downregulation of the identified potential targeted gene *ARNT2* in bronchial epithelial cells protects against the detrimental effect of smoke-induced cell death.

We used our well-established, unique short-term nose-only CS-induced experimental COPD model that recapitulates human COPD's significant characteristics and features after only 8 weeks of CS exposure in our study (102, 103, 105-110). This timepoint is representative of early GOLD stages 1/2 of COPD. Based on this, we focused on the progression phase, where the 115 genes were altered significantly. This list of genes generated from the transcriptomic scale experiment helps gain insight into the mechanistic study (267). During our longitudinal smoke effect exploratory study, we found the AhR signaling pathway as an enriched biological mechanism for identified DEGs. The induction of AhR signaling pathways is essential for regulating diverse downstream physiological processes, including xenobiotic metabolism. We found that *Cyp1a1*, the most significant dysregulated gene in our study (Figure 2.2), is one of the downstream molecules of the Ahr pathway and is involved in the metabolism of a broad spectrum of xenobiotics (268, 269). Considering the importance of *Cyp1a1* in removing the

toxic elements, it was reported that mice lacking *Cyp1a1* die within 30 days of benzo[a]pyrene treatment, polycyclic aromatic hydrocarbons (PAHs) that are mainly by-products of combustion processes like CS (270). It was also identified as a critical regulator of inflammatory responses (271). The *Ahr* pathway is found to be controlled by regulatory molecules, including *Ahr* and *Arnt,* that remained unchanged at the transcriptional level in our model.

However, the longitudinal expressions of *Ahrr* and *Arnt2*, homologs of *Ahr* and *Arnt*, were found to have significantly changed in our study (Figure 2.3F). Looking at the expression profile, *Arnt2* is upregulated during disease development and progression and might be associated with decreased *Cyp1a1* expression. Conflicting studies exist about ARNT2 *in-vitro* dimerization with AhR in the induction of *CYP1A1* (272, 273). One explanation for these conflicting observations could be due to the use of different ligands that induce slightly different conformations of the AhR (274). However, it has been noticed that the functionality of ARNT2 is as efficient as ARNT in the hypoxic induction of gene expression (272). Thus, ARNT2 could dimerize with the AHR *in-vitro* and is influenced by the activating ligand. However, according to one of the studies in the cell culture, it has a limited ability to influence AhR-mediated signaling (274). Taken together, this led to our interest in investigating the functional role of ARNT2 in the presence of CS exposure. We observed that the expression of *CYP1A1* increases with CSE in both WT and KO cells, with a significantly increased expression in KO cell lines in response to CSE, strengthening our hypothesis that ARNT2 might be acting as a repressor in regulating CYP1A1 expression.

Further, the investigation against CS-exposed cell death in the absence of ARNT2 reveals that ARNT2 deficiency might protect cells against the detrimental effect of CS on lung epithelial cells (Figure 2.4D-F). Given that, the transcriptome-wide study and validation of candidates

including *HES1, ID1, ID2, RGS2, IL11RA, KCNJ12* with higher and *SMOC1* with lower expression in *ARNT2* KO CSE exposed cells support the role of ARNT2 suppression against CS-induced cell death. Reviewing the part of identified candidates, researchers observed a significantly lower expression of *HES1* in COPD patients (275). They also observed decreased gene expression and protein levels of *HES1* in HPMEC (Human Pulmonary Microvascular Endothelial Cells) upon 12 hours of 1% CSE exposure (275). It has been reported that *HES1* silencing promotes endothelial activation and apoptosis and significantly inhibits the repair of endothelial damage (276).

In addition to this, inhibitor of DNA binding proteins (IDs) class of helix-loop-helix (HLH) transcription regulatory factors, particularly ID1 and ID2, are crucial regulators of the adaptive antioxidant-mitochondrial response that promotes cell survival during oxidative stress (277). One study has observed the lower expression of *ID1* in endothelial cells of COPD (278). Where overexpression of *ID2*, a transcriptional repressor found to be essential to attenuate the fibrosis and ameliorate the remodeling via inhibiting the TGF-β1/Smad3/HIF-1α/IL-11 Signalling pathway. The role of *ID2* over-expression has also been reported with the reduction of the apoptosis (279). Interestingly, we found that our *ARNT2* knockout cells expressed more *ID1* and *ID2* when exposed to cigarette smoke (Figure 2.5C), which suggests that knocking out *ARNT2* may have a protective effect against CS-induced apoptosis or cell death. Besides, researchers have highlighted the bronchoprotective role of *RGS2* in acute neutrophilic inflammations (280). It has been mentioned that the RGS2 could potentially attenuate the Th1 responses and increase expression in response to long-acting β_2-adrenoceptor agonists and inhaled corticosteroids improve the lung function in acute neutrophilic inflammation, such as exacerbations of COPD or asthma (281, 282). One study reported that IL-11R agonists ameliorated oxidative stress (283).

In addition, *SMOC1* belongs to the matricellular protein family, which is expressed mainly in the basement of various tissues (284-287). It has been found that SMOC1 silencing reduced the ROS and oxidative stress markers and expression levels of fibrosis-associated proteins (288). Taken together, this indicates that *ARNT2* might be responsible for regulating the oxidative stress in our study. Also, downregulation might provide a protective mechanism against CS cell death in bronchial epithelial cells. These findings demonstrate the unique role of *ARNT2* regarding cellular stress and cell death.

Limitations exist in the current study beginning with the data (gene expression and protein profiling) being generated from different samples. Gene expression and protein profiling from the same samples would be more descriptive and correlational. The longitudinal gene expression profiling through the RNA-Seq method would be more valuable and informative than microarrays for such studies due to its higher sensitivity and accuracy. Further, human COPD longitudinal gene expression data must be collected, and a closer investigation of the identified system in a serially designed human experiment (same individual at different time points) can help understand the biology of findings responsible for disease pathogenesis.

In conclusion, we identified *ARNT2* as a potential repressor against *CYP1A1* regulation during chronic CS exposure that might lead to the development of experimental COPD. Also, targeting the ARNT2 might increase cell survivability against the detrimental effect of smoke-induced cell death.

2.5.1 Proposed mechanism

Upon smoke exposure*, AhR* ligands present in smoke bind to the *Ahr* receptor which in turn activates the protective mechanism by neutralizing the ligands through XRE including *Cyp1a1*. Previous research has shown that primary lung fibroblasts from AhR -/- mice exhibited reduced

viability and proliferation following CSE treatment in comparison to AhR +/+ mice's cells (240). Additionally, according to Bekki et al., the inhibition of apoptosis by the AhR ligand and a strong *CYP1A1* inducer promoted cell growth (289). Therefore, we imagine that the increased expression of *Cyp1a1* and decreased apoptosis are interrelated. *Cyp1a1*, one of the downstream molecules of the AhR pathway, was found to be considerably downregulated during the development of the disease in our COPD model. In addition, it was discovered that throughout the development of the disease, there was a considerable upregulation of *Arnt2*, a close homolog of the *Arnt*; which is a crucial regulator of the AhR pathway. Further investigation on the functional role of Arnt2*,* points to a high level of *Cyp1a1* in the absence of Arnt2 upon CSE exposure, which drew our attention to its potential repressive role in *Cyp1a1* regulation. Besides, *Arnt2* deficient cells were also found to be protective against CS-induced cell death. Thus, we proposed that during COPD development, the expression of Arnt2 increases and acts as a repressor of the smoke-protective mechanism. This leads to the suppression of *Cyp1a1* expression, eventually increasing cell death through apoptosis in response to CS (Figure S2.6). Whereas, investigations into the mechanisms that activate the *Arnt2* expression during COPD development are imperative for deciphering the molecular basis of the disease and identifying potential therapeutic targets. To this end, we integrated gene expression data with miRNA expression data in the subsequent chapter of our study.

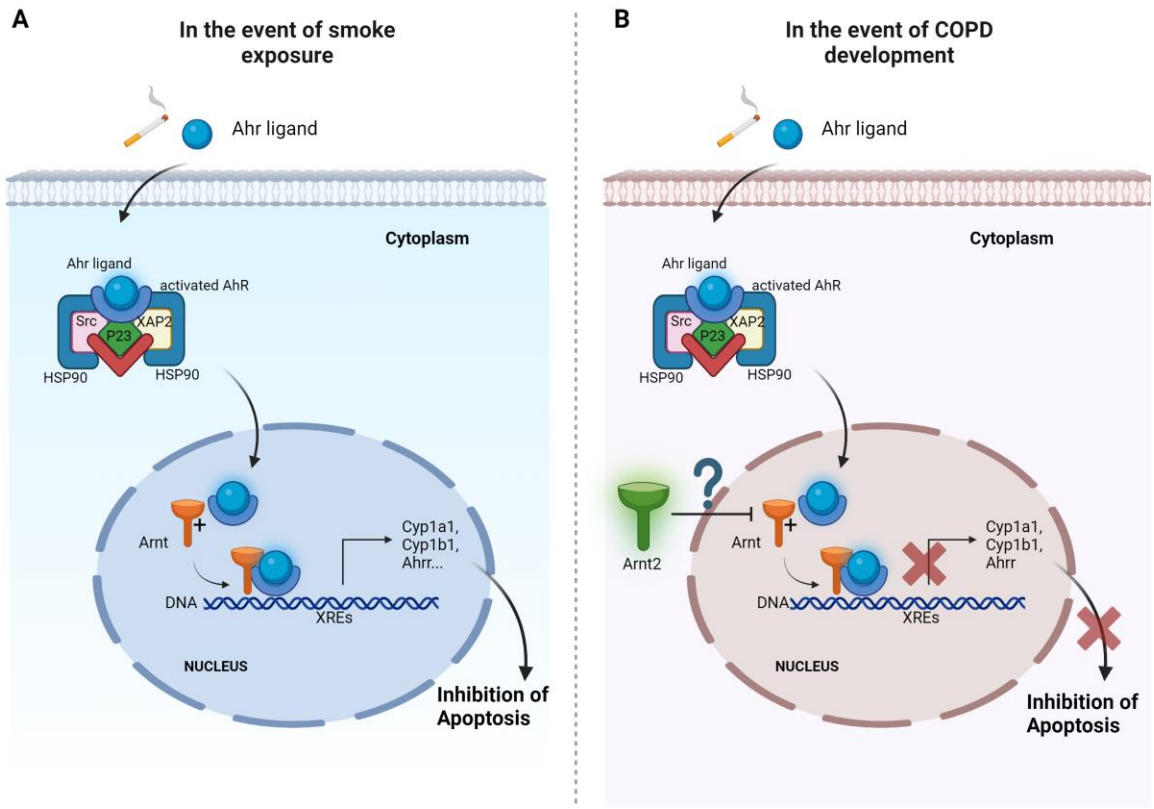

Figure 2.6: Proposed mechanism of *Arnt2* in regulating COPD development: *Under normal circumstances in response to smoke exposure, the Ahr ligand binds to the Ahr receptor that activates the protective response mechanism to neutralize them by promoting the transcription of downstream genes including Cyp1a1. We proposed during the disease development expression of Arnt2 increases and acts as a repressor of the protective response mechanism, leading to the downregulation of Cyp1a1 that is required for the degradation of ligands, ultimately leading to the increased rate of cell death through apoptosis. (Created with BioRender.com)*

Chapter 3: Longitudinal evaluation of whole lung miRNAs and their interactions with gene expression in experimental COPD

3.1 Abstract

Introduction: miRNAs are small non-coding endogenous RNAs involved in regulating the expression of genes that substantially contribute to the development and progression of COPD. However, the interactions of miRNAs with their target genes that contribute to COPD development are incompletely understood. Thus, we investigated the expression of miRNAs and their target genes over a time course of CS-induced experimental COPD.

Methods: Lung miRNA expression was profiled at 4-, 6-, 8-, and 12 weeks in experimental COPD to define the longitudinal smoke exposure effect. Subsequently, the impact of miRNA expression on lung function was determined by correlating specific miRNA expression with post-bronchodilator forced expiration volume in one second (FEV_1) in an independent human cohort. To gain a deeper understanding of the miRNA regulation of transcriptional changes during COPD development, we integrated the miRNA expression profile with the transcriptomic profile in another experimental COPD. Findings were validated using qPCR in a separate experimental COPD model.

Results: We profiled the miRNA changes during the longitudinal smoke exposure and found 48 dysregulated with false discovery rates (FDR) < 0.1. Association of miRNAs with disease severity highlighted 5 miRNAs, let-7b, miR-23b, miR-744, miR-100, and miR-320a, that were positively correlated (p-value < 0.05) with human FEV. Correlation analysis demonstrated that the expression of 7 miRNAs including let-7b-5p, miR-181a-, b-, c-5p, miR-150-5p, miR-151-5p, and miR-30b-5p were significantly altered along with their respective direct targets over the time-course. Functional enrichment analysis showed that the direct targets of miR-181

family, let-7b, and miR-151-5p including *Slco2b1*, *Myb*, *Kcnmb2*, *Bbox1*, and *Arnt2* were enriched as were pathways including cellular protection (xenobiotic response) and cell fate determination (*Wnt* signaling pathway). Further, *Arnt2* negatively correlated with $FEV_{0.1}$/FVC and positively with TLC. Validation in a separate experimental COPD model found that let-7b, miR-181a-5p, and miR-181c-5p expression anti-correlated with *Arnt2* expression, miR-151-5p, and miR-150-5p expression anti-correlated with *Bbox1* and *Ttc25* expression, respectively.

Conclusion: Our study reveals that let-7b, miR-151-5p, miR-181a-5p, and miR-181c-5p, along with their potential targeted genes including *Arnt2, Bbox1*, and *Ttc25*, may have regulatory roles in the development and progression of COPD.

3.2 Introduction

Chronic obstructive pulmonary disease (COPD), is a progressive multifactorial inflammatory disease of the lungs that blocks the airflow and is not fully reversible. Its major risk factor is exposure to cigarette smoke (CS) (108, 290). The pathophysiology of COPD has not been understood completely to date. Currently, there is no treatment available that halts the progression or reverses COPD, and the focus of therapy is on symptom management.

miRNAs are small endogenous non-coding RNAs (ncRNAs) with 19-25 nucleotides that negatively regulate their target genes at the post-transcriptional level through transcript destabilization or protein translational inhibition (291, 292). Approximately 60% of genes may be the targets of miRNAs. It is estimated that miRNAs may target up to one-third of the transcriptome (254). One single miRNA can bind to target sequences in multiple genes, and many miRNAs can cooperatively target a single gene. Thus, they have potential roles in regulating a wide variety of physiological processes and signaling pathways in different types of cells, including cells of the immune system, as well as in controlling inflammatory responses in various tissues (293, 294). Thus, miRNAs may control entire disease pathways.

The study of miRNAs has been rapidly extended to respiratory diseases including COPD, and shows that CS modulates miRNA regulation in samples such as lung tissues (242, 246, 293, 295-298). Studies have established that miRNAs are mediators of inflammation and can regulate the expression of genes, which are responsible for the development and progression of COPD (299). The expression of miRNAs is described to be altered after cigarette smoke (CS) exposure in the lungs of mice and patients with COPD (295, 300-303). miRNAs have also been implicated in the pathophysiology of COPD (304-306). It has been found that the majority of deregulated miRNAs in these studies were downregulated due to the effect of smoke exposure.

To understand miRNA-mediated regulation, the vital thing is to explore the key miRNAs and miRNA-gene interactions before developing miRNA-based therapeutics. Current algorithms that predict miRNA targets depend on evolutionary conservation, sequence complementarity, and thermodynamics stability. Yet, the main issue remains the false interpretations, as most of these interactions may depend on additional cofactors or cell types. However, the inverse correlation between genes and miRNA expression levels from the same model can help improve the target predictions (307). Only a few studies have done the combinatorial analysis of miRNAs and gene regulations (153, 308, 309), those that have been performed show that miRNA and gene pairs can also play important roles in COPD development and pathogenesis.

To fill this gap, in this study, we first investigated if miRNAs were differentially expressed longitudinally over a time course in our well-established mouse model of CS-induced experimental COPD (102, 310-313). The dysregulated miRNAs were also correlated with post-bronchodilator forced expiration volume in one second (FEV_1) in a human cohort. We then assessed the dysregulated miRNA and gene expression links by conducting a canonical correlation analysis on miRNA-gene expression profiles. We also predicted the targeted genes by narrowing down their relationship logically and efficiently. The predicted miRNA and gene expression correlations were then validated in a separate experimental COPD model. Our findings highlight the critical miRNAs and uncover their target genes in the instigation of COPD development, providing avenues for further investigation into potential mechanisms of action and therapeutic targeting.

3.3 Methodology

To investigate miRNAs that are differentially expressed during longitudinal smoke exposure, we investigated the time-course CS-induced COPD model at the post-transcriptional level (Figure 3.1). Here, miRNA profile from an additional COPD model (314) was used, and the model was generated following the same protocol outlined in Chapter 1 and Chapter 2. In summary, lung tissues were collected from biological replicates of each group at 4, 6, 8, and 12 weeks post-exposure (air or smoke).

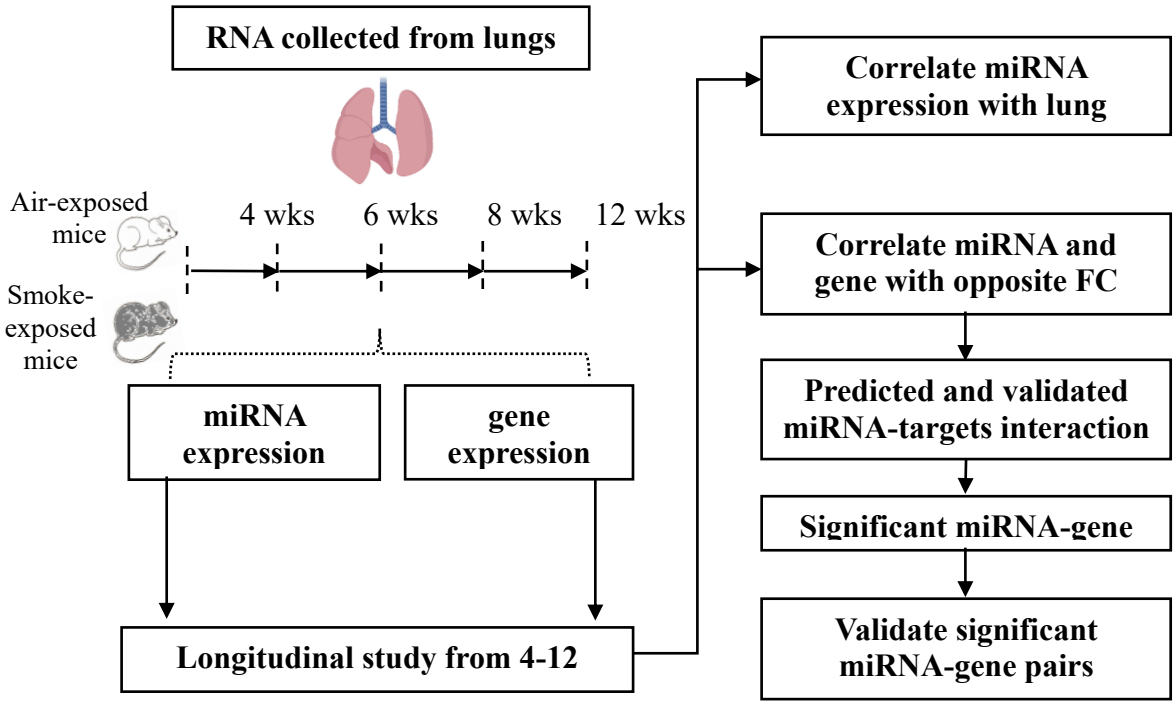

Figure 3.1: Data integration approach: *RNA from lung tissues of BALB/c mice from longitudinal CS exposure was profiled for miRNA and gene expression. The calculated mean of log normalized differential miRNA and gene expression at each time point was used for correlation. miRNAs and genes expressed differentially in opposite directions were correlated. Subsequently, validated and predicted direct targets of differentially expressed miRNAs were investigated, followed by the validation in an independent experimental COPD model. FC: fold change; miRNA: microRNA.*

3.3.1 Short-term CS-induced experimental COPD

BALB/c (n=4) and C57BL/6 (n=8-10) housed in specific pathogen-free and controlled environmental conditions were used. The mouse COPD model was developed according to an established protocol (102, 107, 241, 246, 311-313). Briefly, mice were exposed to nose-only directed flow inhalation to 12 cigarette smoke (3R4F reference cigarettes [the University of Kentucky, Lexington, Ky] twice per day and 5 times per week) for 12 weeks (102, 105, 107, 241, 311-313). Each exposure lasted for 75 minutes. Normal air-exposed mice were used as controls. Experimental procedures were approved by the institutional animal Ethics committee.

3.3.2 miRNA and gene expression profiling

In our investigation, miRNA expression profiling data sourced from an additional COPD model, accessible through the NCBI under the identifier GSE186955 (314), was integrated. This COPD model was established utilizing the identical procedural framework delineated in both Chapter 1 and Chapter 2 of our study. The raw data acquisition was facilitated through collaboration with Richard Kim, Emma Beckett, Andrew Jarnicki, and Andrew Deane affiliated with the Hunter Medical Research Institute and The University of Newcastle, New South Wales, Australia. To establish correlations with gene expression profiles, the data employed in Chapter 2 was referenced. It is noteworthy to mention that these experimental COPD models utilized for the canonical analysis (for gene expression profile and miRNA expression profile) were executed separately.

3.3.3 Data processing and statistical analysis

Raw miRNA expression profile data were processed using R statistical software version 4.0.2 (315) with RStudio version 1.3.1 (249) and Bioconductor (250). A preliminary exploratory analysis was performed to identify the outliers and other problems through principal

component analysis (PCA) using the "PCAtools" package (doi:10.18129/B9.bioc.PCAtools) (316). To obtain an estimate of the normalized signal for each miRNA, signals were background corrected and normalized with the quantile method by the AgiMicroRna package (doi:10.18129/B9.bioc.AgiMicroRna) (317). The measured intensities were log2 transformed. miRNAs that do not reach a minimum quality were filtered out. A linear model is fitted to each miRNA so that the FC between different experimental conditions and their standard errors can be estimated. The empirical Bayes method was applied to obtain moderated statistics. Differential expressions of miRNAs were analyzed with the "limma" package (doi:10.1093/nar/gkv007) (253, 254) of R and corrected for multiple testing using the Benjamini–Hochberg false discovery rate (FDR). Longitudinal changes in miRNA expression were measured using the following equation.

$$Longitudinal\ effect\ of\ smoke =$$

$$(Smoke12weeks - Smoke04weeks) - (Air12weeks - Air04weeks)$$

(eq.3)

3.3.4 Association of miRNAs with clinical features in a human cohort

To further investigate the association of dysregulated miRNAs identified during the longitudinal study with lung function parameters. We used a previously published GLUCOLD cohort study (available at NCBI: GSE36221 and GSE76774) (153, 308, 318). This data comprised COPD patients who participated in the GLUCOLD study. In our study, we incorporated a cohort of 38 humans. Demographic information including age, sex, smoking status, etc of the participating individuals is accessible via the NCBI accession numbers GSE36221 and GSE76774. COPD patients had lung function with postbronchodilator forced expiration volume in one second (FEV_1) < 80%. We correlated the miRNA expression in

human subjects with the FEV_1 using the Pearson correlation coefficient. Significant correlations (p-value < 0.05) with an absolute R-value (r >0.25) were considered biologically relevant.

3.3.5 Canonical correlation of miRNA expression level with gene expression levels

To identify the effect of the longitudinal shift of miRNA expression along with gene expression due to continuous CS exposure, we performed the canonical correlation analysis. Firstly, the expression of dysregulated miRNAs identified during the longitudinal effect of smoke study in our experimental COPD was correlated with the expression of dysregulated genes identified in Chapter 2 (during the longitudinal effect of smoke study). Briefly, for this, we first identified the genes and miRNAs using the same statistical model, defined as COPD genes and miRNAs in our model. We then calculated the mean of transformed normalized expression levels of all samples in each group. Subsequently, the expression of miRNAs and genes whose expressions were changed in opposite directions were correlated using the psych R package (319). Pearson correlation test was used to examine pair-wise correlations between the expression profile of miRNAs and genes expressed at opposite directions, with the p-value set to 0.05. Two-tailed probability for each correlation was calculated. Secondly, miRNA and genes expressed in opposite directions were interrogated for direct interactions using "multiMiR" R package (doi:10.1093/nar/gku631) (320), which has a comprehensive collection of predicted and validated miRNA-target interactions. The multiMiR database utilized in our study integrates human and mouse data sourced from 14 external databases, encompassing validated miRNA-target databases such as miRecords (321), miRTarBase (322), and TarBase (323). Additionally, predicted miRNA-target databases including DIANA-microT (324), ElMMo (325), MicroCosm (326), miRanda (327), miRDB (328), PicTar (329), PITA (330), and TargetScan (331) are incorporated, alongside disease-/drug-related miRNA databases including

miR2Disease (332), Pharmaco-miR (333), and PhenomiR (334). Thirdly, both factors including significant associations and direction interactions were incorporated together to investigate the significantly negatively correlated direct targets of miRNA in our time course COPD model.

3.3.6 Characterization of functionally enriched miRNAs

Functional enrichment analyses of significantly anti-correlated direct targeted genes of miRNA were performed by the g:GOSt module of the g:Profiler (https://biit.cs.ut.ee/gprofiler/gost) (335).

3.3.7 Investigation of significant correlated miRNA-gene expression in an independent experimental COPD model

As previously mentioned, we conducted correlation analyses between gene expression and miRNA expression using data generated from two separate COPD models. We now aim to elucidate the significance of identified miRNA alterations on gene expression within the same samples. To achieve this, we established a third independent COPD model in the C57BL/6 strain, adhering to an identical procedural framework at the Centenary Institute (Camperdown, NSW). However, it's important to highlight that the preceding dataset utilized for in-silico analysis was generated by our collaborative partners at the Hunter Medical Research Institute, using BALB/c mice. In our subsequent investigation, we employed the Pearson correlation method to systematically explore all associations, as delineated in the preceding section. Furthermore, the impact of identified miRNAs and gene expression changes on lung function during the time course of CS induction that leads to COPD was also evaluated.

3.3.8 Quantification of miRNA and mRNA expression by real-time qPCR

Lung tissue collected from biological replicates of each group for a time-course experiment (at 2-, 4-, 8-, and 12 weeks) was snap frozen and used for total RNA extraction with TRIzol® (Invitrogen) after blunt-dissected airway and parenchyma tissues. as previously described (242, 298, 336). RNA was quantified using a NanoDropTM, ND1000 spectrophotometer (NanoDrop TM USA). miRNA-specific cDNA and total cDNA were synthesized from RNA after DNAse-I treatment. For miRNAs, multiplex reverse transcription was performed using a combination of miRNA-specific reverse primers and the endogenous control small nuclear RNA (snRNA) U6 and small nucleolar RNA (snoRNA) U49. Expression of miRNA and genes were assessed by real-time qPCR (242, 298, 336) with SYBR Green Supermix (KAPA Biosystems). For miRNA, custom-designed primers for mir-181a-5p, miR-181b-5p, miR-181c-5p, let-7b-5p, miR-151-5p, miR-150-5p, and miR-30b-5p were synthesized from Integrated DNA Technologies (IDT) and QIAGEN (EXIQON) and used at an annealing temperature of 50^0C. For gene expression, pre-designed KiCqStart® SYBR® Green Primers (Sigma Aldrich) for *Arnt2, Bbox1, Kcnmb2, Mfsd4, Myb, Slco2b1,* and *Ttc25* were used at annealing temperature of 56^0C. For quantification of miRNAs in mouse lung tissues, the expression was normalized to the geometric mean of two endogenous controls U6 and U49. For gene quantification, the expression of genes was normalized to the hypoxanthine-guanine phosphoribosyl transferase (*Hprt*). Primer sequences of miRNAs and genes are provided in Table S3.1 and Table S3.2).

3.3.9 Statistical analysis

The data was assumed to be normally distributed. Statistical analyses were conducted utilizing GraphPad Prism (v8) (265). Group comparisons were executed through ordinary two-way ANOVA. Correlation analyses, employing Pearson's correlation coefficient (r) and associated probabilistic p-values, were computed using R statistical software version 4.0.2 (315).

3.4 Results

A PCA plot based on total miRNA expression revealed differences in the experimental groups (Figure 3.2A). The clear separation in miRNA expression between air-exposed (n=16; 4 mice at each timepoint 4, 6, 8, and 12 weeks of exposure) and smoke-exposed mice (n=16; 4 mice at each timepoint 4, 6, 8, and 12 weeks of exposure) mirrors its different cellular content (Figure 3.2A).

3.4.1 Longitudinal effects of smoke exposure on miRNA expression

As per the dataset GSE186955 used, miRNA profiling was conducted utilizing the Agilent unrestricted Mouse miRNA platform (314). This platform quantified a total of 672 mouse miRNAs. Following control-type filtration, 650 miRNAs were retained for further analysis. The influence of longitudinal smoke effect on miRNA expression was obtained by comparing the expression profile of 4 and 12 weeks in the CS-exposed group relative to the air-exposed mice group, using the equation (eq.3) mentioned in section 3.3.3. During an investigation of the longitudinal effect of smoke on miRNA expression from weeks 4 to 12, we identified only 48 miRNAs as differentially expressed with FDR < 0.1 (Figure 3.2B and Table S3.3). Of the 48 miRNAs, 09 were exhibited as highly expressed in CS-exposed mice to baseline expression. Where, 39 miRNAs showed a continuous reduction in their expression with increased exposure to CS till 12 weeks, compared to a baseline of 4 weeks of exposure (Figure 3.2C). Among them, miR-449a-5p was the most significant highly expressed miRNA with a $\log_2$FC of 1.64, whereas miR-195-5p was the most significant lower expressed miRNA (Figure 3.2B-C).

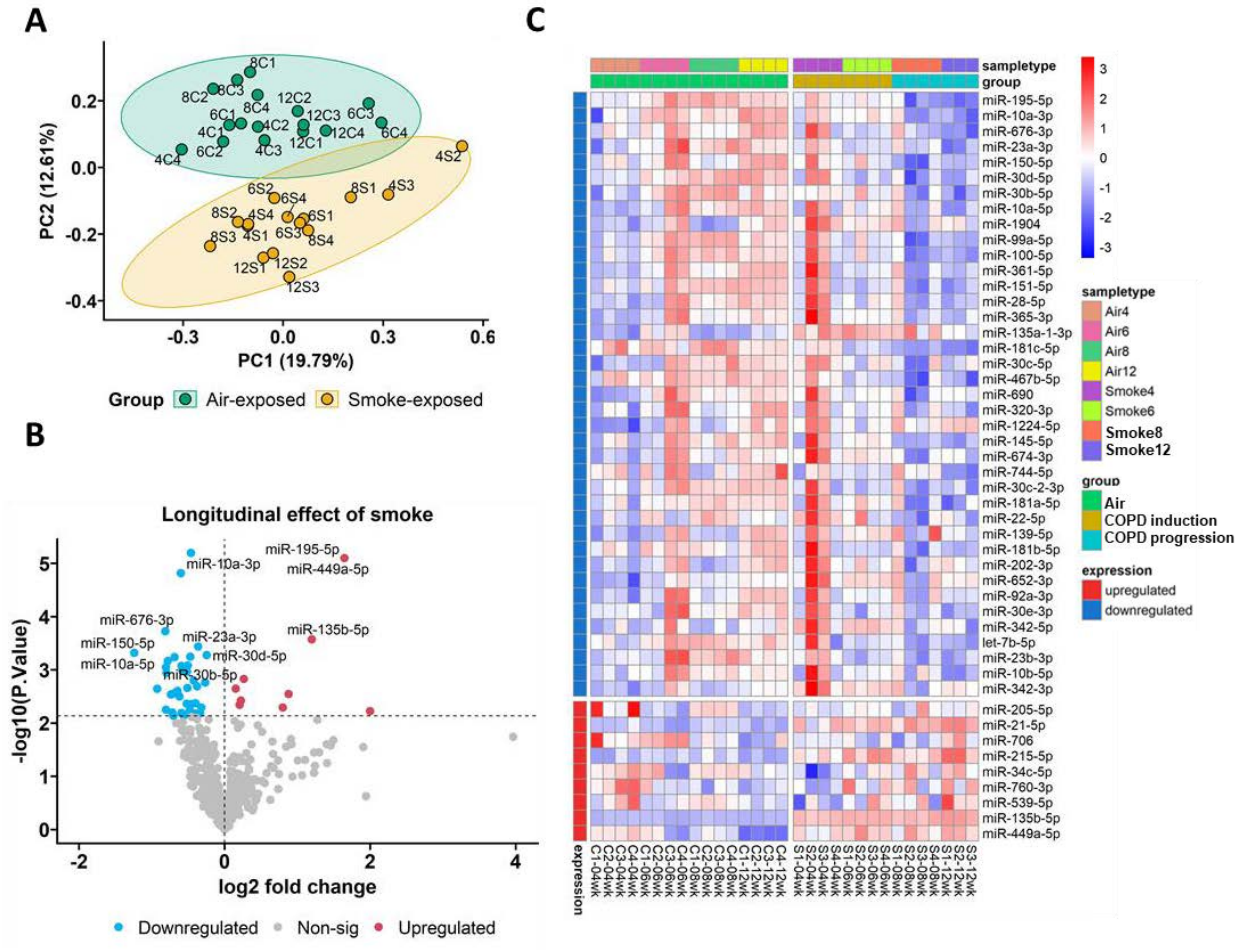

Figure 3.2*:* Expression profiling of miRNAs in lung tissues in CS-induced experimental COPD: ***A)*** *Principal component analysis (PCA) of miRNA expression levels in CS-induced experimental COPD showed air-exposed control group mice (n = 16; 4 mice at each timepoint 4, 6, 8 and 12 weeks of exposure) and smoke-exposed group mice (n = 16; 4 mice at each timepoint 4, 6, 8 and 12 weeks of exposure) are clustered separately.* ***B)*** *Volcano plot for differentially expressed miRNAs identified due to longitudinal effect of smoke from 4 to 12 weeks that leads to COPD development. The plot shows the differential miRNAs (FDR <0.1) in terms of significance (-log$_{10}$(p)) versus log$_2$(FC) on the y-axis and x-axis, respectively. Each blue, red, and grey dot indicates downregulated, upregulated, and non-significantly altered gene expressions.* ***C)*** *Heatmap representing the differential expressed miRNAs (longitudinal study) in each mouse sample. The level of gene expression was scaled from high (red) to low (blue).*

3.4.2 The association of differentially expressed miRNAs shows a significant correlation with lung function parameters in human COPD

We further investigated the role of miRNA expression with disease severity in the human cohort. Correlation between miRNA expression and post-bronchodilator Forced Expiratory Volume (FEV post) prediction, a key indicator of lung function identified significant associations with p-value < 0.05 and r > 0.25 (Figure 3.3). Among the 48 miRNAs examined, five miRNAs namely let-7b, miR-23b, miR-744, miR-100, and miR-320a exhibited positive correlations with FEV post-prediction, while only one, miR-181c, displayed a negative correlation (p-values < 0.05) (Figure 3.3). Notably, the remaining miRNAs did not demonstrate significant associations. It is crucial to understand that a higher predicted FEV post value suggests better lung function, whereas a lower value indicates impaired lung function. Interestingly, we observed that all downregulated miRNAs, except miR-181c, demonstrated a positive correlation. This suggests that modulating the expression of these miRNAs might explored as a strategy to modify lung function.

3.4.3 Correlation analysis showed significant correlations between differentially expressed miRNAs and their target genes

A total of 115 genes were identified during the longitudinal investigation of smoke exposure at transcriptional changes in chapter 2 (Figure 2.2D-E) and 48 miRNAs identified during the longitudinal effect of the smoke study (Figure 3.2B-C) were correlated with such as downregulated miRNAs with upregulated genes and vice versa. In total, there were 4,575 possible pairs (Figure 3.4A), among them, 487 significant correlations (p-value < 0.05) were observed between downregulated miRNAs and upregulated genes identified during longitudinal studies (Figure 3.4B). We observed that the downregulated miRNAs in our study i.e., miR-10a-3p, miR-145-5p, miR-676, miR467b-5p, and miR-181a-5p showed a higher

number of significant correlations with the expression of genes (Figure 3.4B). On the other hand, a total of 24 significant correlations with p-value < 0.05 were observed between upregulated miRNAs and downregulated genes that were identified during longitudinal analysis (Figure 3.4C). These 24 significant correlations are inclusive of 6 upregulated miRNAs and 17 downregulated genes. We observed that among all the upregulated miRNAs, the miR-135b-5p had the highest number of significant correlations with the expression of down-regulated genes (Figure 3.4C).

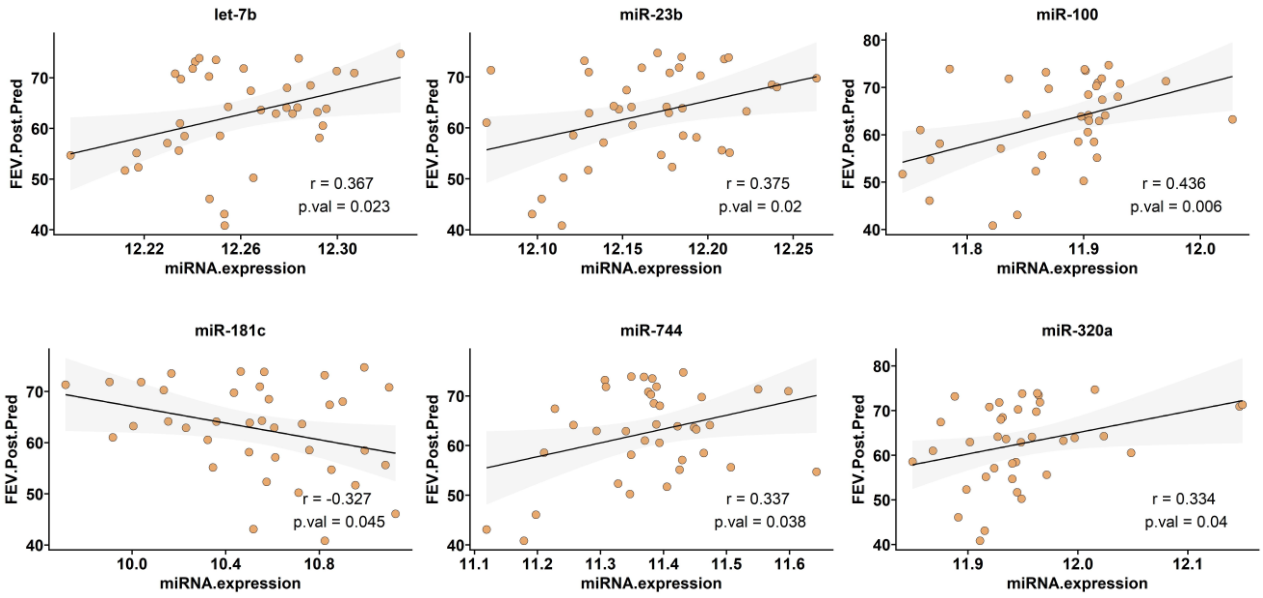

Figure 3.3: Correlation of longitudinal differentially expressed miRNAs with disease severity in human COPD (n=38): *Expression of miRNA was correlated with FEV post prediction in a human cohort using Pearson correlation showing a significant association with p< 0.05.*

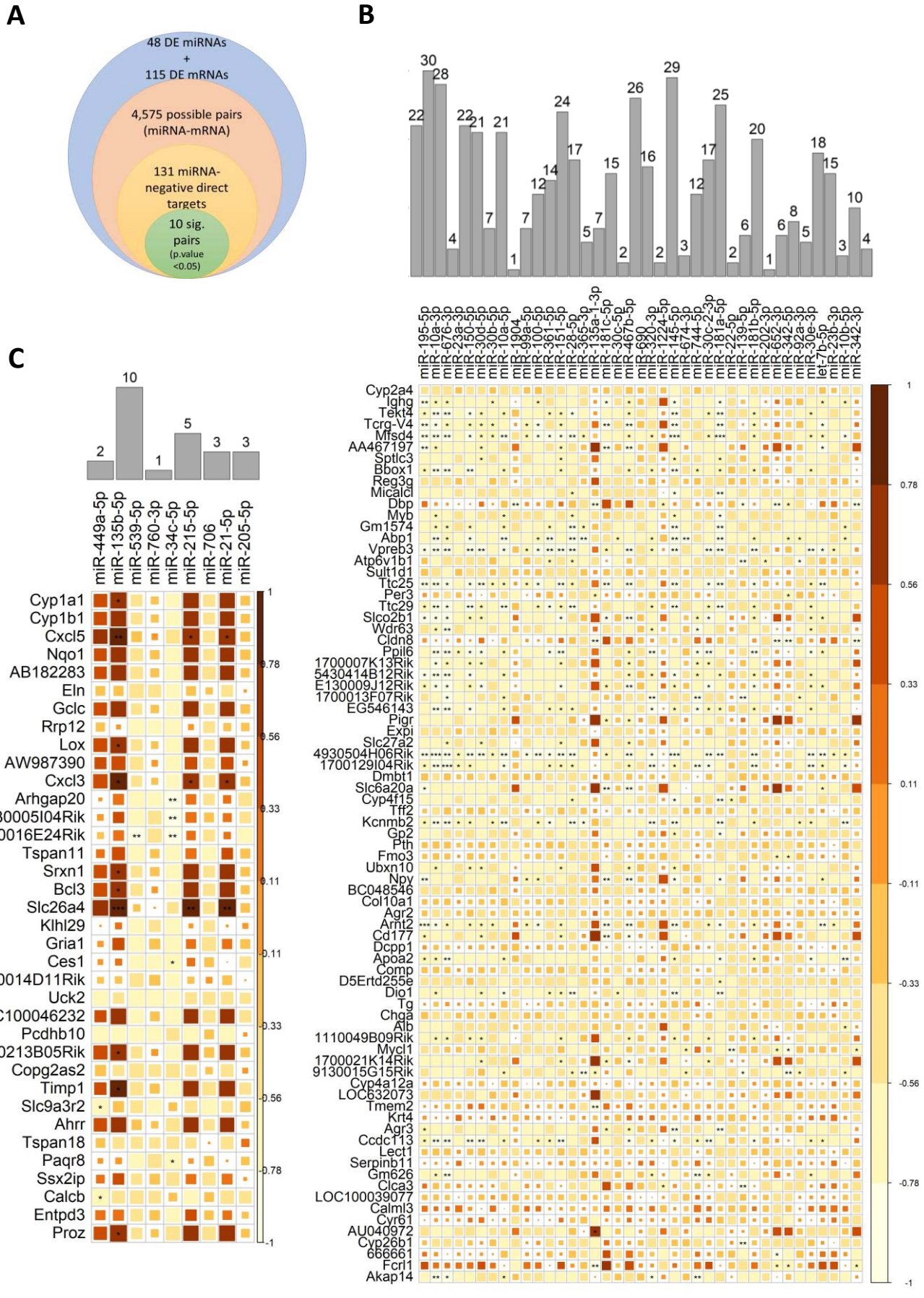
A
48 DE miRNAs
+
115 DE mRNAs
4,575 possible pairs
(miRNA-mRNA)
131 miRNA-negative direct targets
10 sig. pairs
(p.value <0.05)
B
C
miR-449a-5p
miR-135b-5p
miR-539-5p
miR-760-3p
miR-34c-5p
miR-215-5p
miR-706
miR-21-5p
miR-205-5p
Cyp1a1
Cyp1b1
Cxcl5
Nqo1
AB182283
Eln
Gclc
Rrp12
Lox
AW987390
Cxcl3
Arhgap20
A630005I04Rik
1200016E24Rik
Tspan11
Srxn1
Bcl3
Slc26a4
Klhl29
Gria1
Ces1
2310014D11Rik
Uck2
LOC100046232
Pcdhb10
9130213B05Rik
Copg2as2
Timp1
Slc9a3r2
Ahrr
Tspan18
Paqr8
Ssx2ip
Calcb
Entpd3
Proz
1
0.78
0.56
0.33
0.11
-0.11
-0.33
-0.56
-0.78
-1

Figure 3.4: Integration of miRNAs and gene expression profiles in CS-induced experimental COPD that were differentially expressed during longitudinal smoke exposure: ***A)*** *Overview of correlation study of miRNA expression with gene expression identified during longitudinal smoke exposure study.* ***B)*** *Heatmap shows the correlation between downregulated miRNAs with upregulated genes.* ***C)*** *Heatmap shows the correlation of upregulated miRNAs with downregulated genes, respectively. Bar plots on the top of each heat map indicate the total number of significant interactions of miRNAs with genes. The number of asterisks in each cell indicates the level of significance of correlation. The correlation was scaled from positive (red) to negative (yellow).*

3.4.4 Identification of direct miRNA targets highlights validated and predicted targets of dysregulated miRNAs in COPD

We further explored if the identified list of 115 genes are direct targets of any miRNAs out of 48 differentially expressed miRNAs. We identified that a total of 39 genes are direct targets of 33 miRNAs (Figure 3.5A). This makes up a total of 131 direct interactions between miRNAs and genes including both validated and predicted interactions (Figure 3.5A). Of all 131 direct interactions, 25 interactions are validated targets of miRNAs.

3.4.5 Correlation analyses showed strong correlations between the expression of miRNAs and their directs targets in COPD

Thereafter, we classified the direct targets in terms of significant correlations. However, not all the pairs were significantly correlated (p-value < 0.05). We found only 10 pairs were significantly anti-correlated with p-value < 0.05 (Figure 3.5B). In addition, among the significant negative correlation, we identified two validated target genes (*Arnt2* and *Slco2b1*) for miR-let-7b, one validated target gene (*Myb*) for miR-181a-5p, and one validated target gene for miR-151-5p (Bbox1), according to multiMiR outcome. The analysis conducted in our study using the multiMiR package relied upon three validated miRNA-target databases, namely

miRecords (321), miRTarBase (322), and TarBase (323). These databases were utilized to retrieve experimentally validated miRNA targets, employing a range of methodologies including Degradome sequencing, Luciferase reporter assays, Luciferase activity assays, qRT-PCR, and HITS-CLIP. Additionally, we also listed the predicted target genes for the miRNAs. We anticipated *Arnt2* as the predicted target for 3 miRNAs (miR-181a-5p, miR-181b-5p and miR-181c-5p), *Ttc25*, *Kcnmb2* and *Mfsd4* as the predicted targets for miR-150-5p, miR-30b-5p and let-7b-5p, respectively (Figure 3.5B). The differential expressed miRNAs from each group having a significant negative association with the differentially expressed targeted genes are depicted in Figure 3.6.

3.4.6 Functional enrichment analysis of negatively correlated direct targets of miRNAs in experimental COPD

To gain mechanistic insight into the significant anti-correlated miRNA targets, we conducted the functional enrichment analysis linking these targets with their respective molecular functions and biological pathways using g:Profiler. By interrogating target genes such as *Arnt2*, *Myb, Slco2b1, Bbox1, Ttc25, Mfsd4*, and *Kcnmb2* against two well-established databases, REACTOME and WikiPathways, we uncovered a spectrum of 12 pathways (Figure 3.7A). Pathways identified using the REACTOME database were mainly involved in cellular protection, metabolism, and transportation. Whereas, pathways identified via the WikiPathways database included the Wnt signaling pathway and Neural crest differentiation. Molecular functions associated with these genes predominantly revolved around transport and transcription activation activities (Figure 3.7A). Remarkably, during enrichment analysis, five genes (*Arnt2, Myb, Slco2b1, Bbox1,* and *Kcnmb2*) exhibited significant enrichment patterns (Figure 3.7B and Table S3.4), underscoring their potential importance in the observed regulatory dynamics.

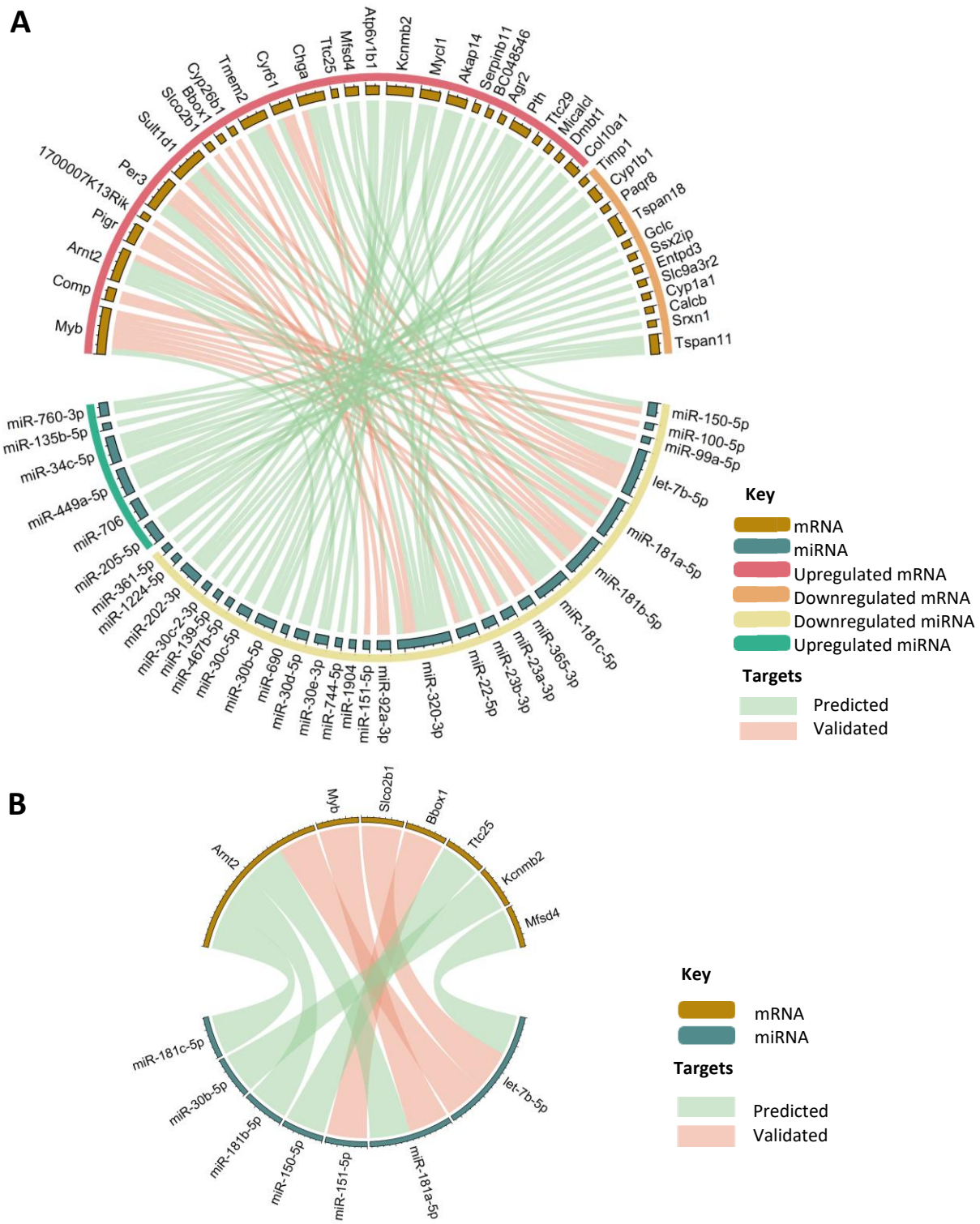

Figure 3.5: Integration of differentially expressed miRNAs with validated and predicted direct targets identified using multiMiR in COPD: ***A)*** *Circular plot showing all the experimentally validated and predicted direct targets of miRNAs from the list of identified differentially expressed genes.* ***B)*** *Circular plot showing only significant correlated direct targets of miRNAs having negative correlation coefficient with p-value < 0.05.*

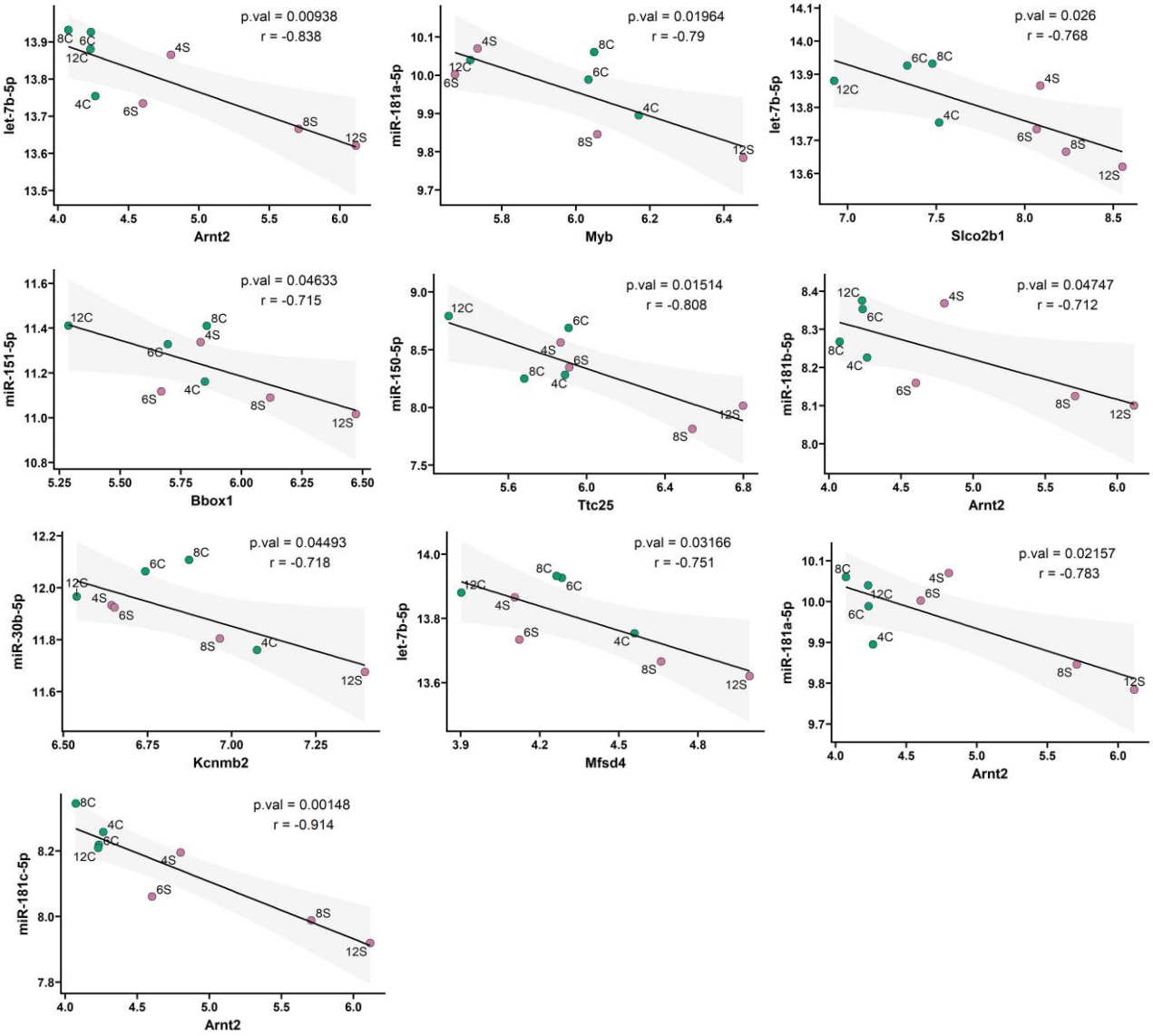

Figure 3.6: Anti-correlated miRNAs and their direct targeted genes identified through longitudinal analysis in CS-induced experimental COPD: *Pearson correlation scatter plots with linear regression are shown in black line with its confidence interval represented by grey area for miRNA expression vs. identified targeted gene expression during the time course CS-induced experimental COPD (n=8). Within each plot, pink and green circles represent the smoke-exposed (n=4) and air-exposed control group (n=4), respectively, labeled with their corresponding timepoints. Data are shown in Pearson correlation coefficient r and significance in terms of p-value.*

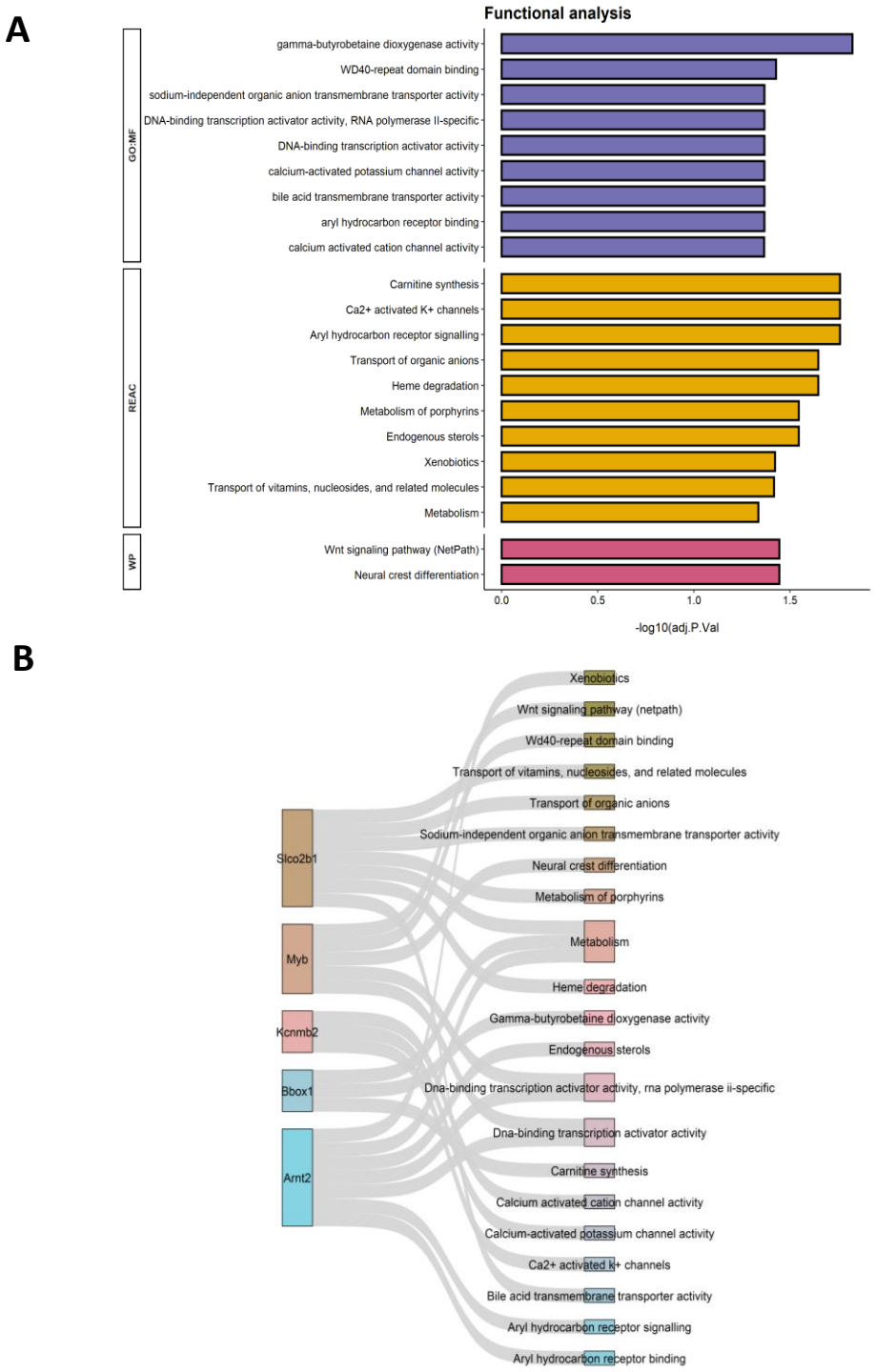

Figure 3.7: Functional enrichment analysis of negatively correlated direct targets of selected miRNAs in experimental COPD (BALB/c mice): ***A)*** *Significantly enriched pathways and molecular functions of miRNAs direct targets identified using g:Profiler.* ***B)*** *The Sanky diagram highlights the direct targets of miRNAs involved in functionally enriched pathways and molecular functions with adj. $p < 0.05$.*

3.4.7 Correlation shows *Arnt2*, *Bbox1,* and *Ttc25* may be direct targets of dysregulated miRNA in experimental COPD

The identified miRNAs and their respective validated and predicted target genes were quantified in terms of 1/delta Ct in a distinct COPD model utilizing the C57BL/6 strain (Figure S3.1 and Figure S3.2). The COPD model was generated following the same protocol as outlined in Chapter 1 and Chapter 2. Pearson correlation analysis between miRNA and gene expression revealed five miRNA-gene pairs exhibiting significant associations with p-value < 0.05 (Figure 3.8). Specifically, let-7b, miR-181a-5p, and miR-181c-5p displayed significant negative associations with their validated and predicted target gene *Arnt2* when examining expression using 1/delta Ct values. Additionally, miR-151-5p expression exhibited a significant negative association with its potential target *Bbox1* expression, while miR-150-5p expression was significantly associated ($p < 0.05$) with its target gene *Ttc25* expression, both showing negative coefficient correlations (Figure 3.8). These findings suggest that these miRNAs might play an important regulatory role in modulating their target genes during disease development.

Moreover, the longitudinal analysis of differentially expressed miRNAs in association with lung function parameters revealed a negative correlation of *Arnt2* with the lung function parameter FEV0.1/FVC (Figure 3.9). Given that a low FEV0.1/FVC ratio signifies airflow obstruction, the heightened expression of *Arnt2* may contribute to airflow obstruction, highlighting its detrimental role and suggesting its potential as a therapeutic target to mitigate disease progression. Notably, the association study with total lung capacity (TLC) unveiled a positive correlation with *Arnt2, Mfsd4, Myb, Slco2b1,* and *Ttc25* (Figure 3.9). A high TLC typically indicates increased lung volume or lung hyperinflation. Conditions such as COPD can lead to air trapping in the lungs, resulting in hyperinflation and an elevation in TLC. This

underscores the impact of *Arnt2* and its regulatory miRNAs during disease development, underscoring their potential as therapeutic targets.

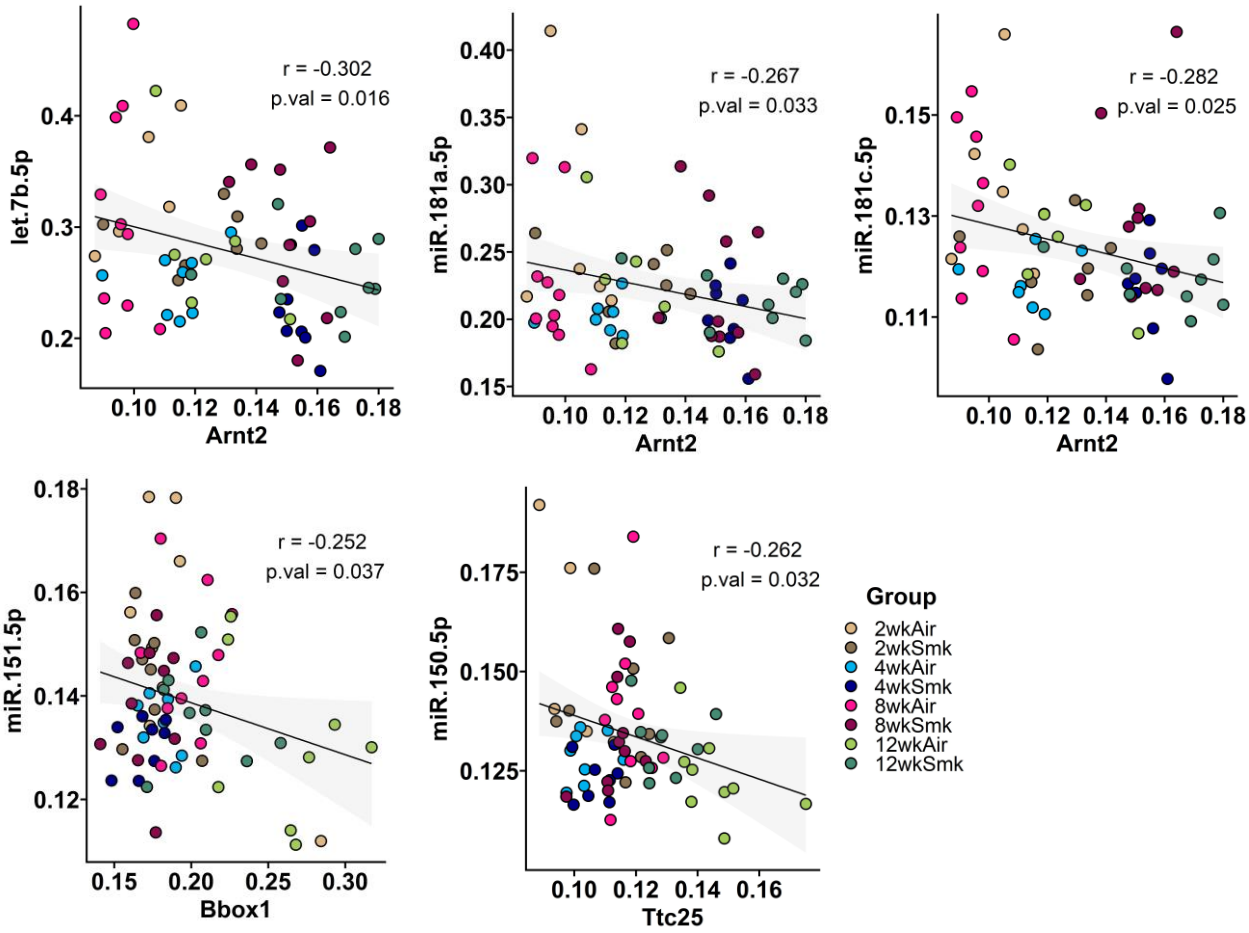

Figure 3.8: *Correlation* of negatively associated miRNAs and their target genes in CS-induced experimental COPD in C57BL/6 mice: *Pearson correlation scatter plots with linear regression are shown in black line with its confidence interval represented by grey area for miRNA expression vs. identified targeted gene expression during the time course CS-induced experimental COPD(n=8-10 mice were included in each group). Data are shown in Pearson correlation coefficient r and significance in terms of p-value.*

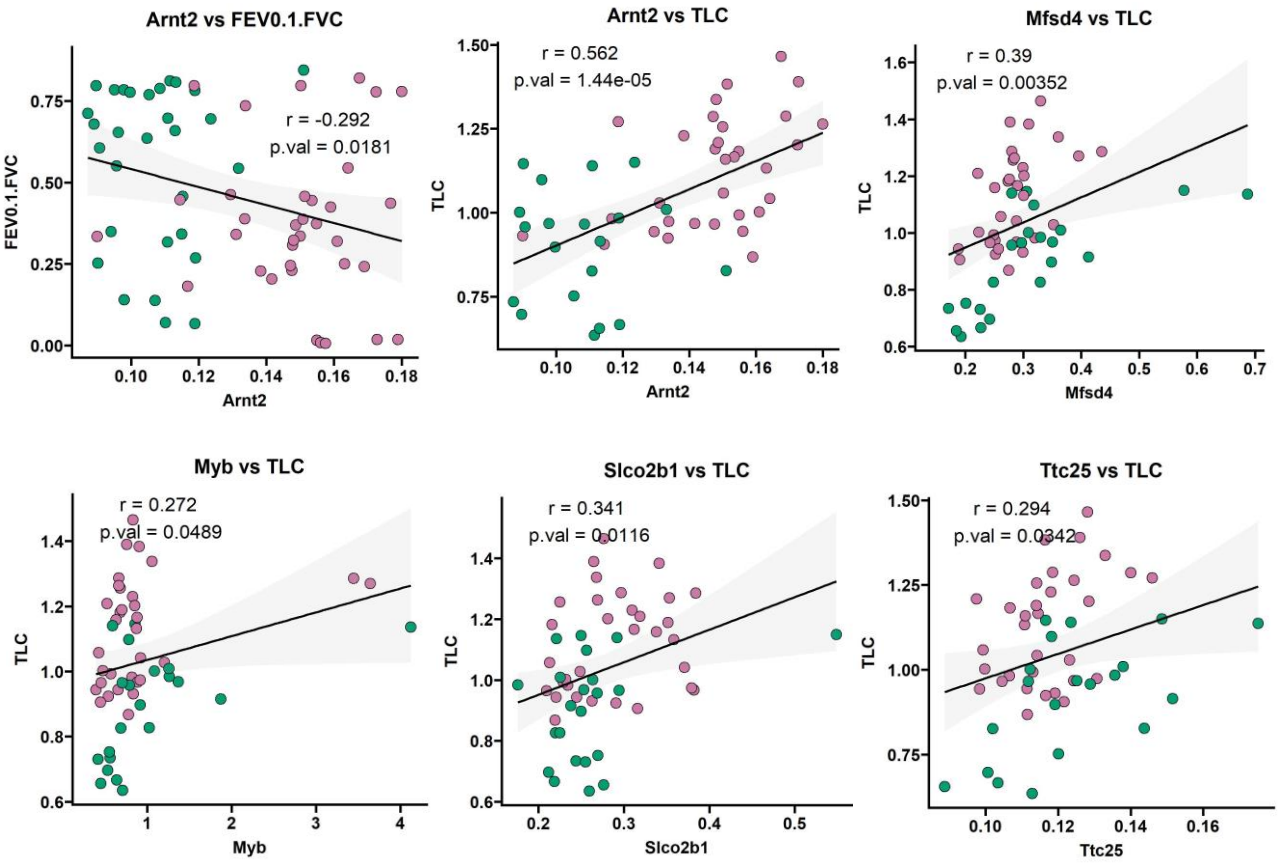

Figure 3.9: Correlation of longitudinal differentially expressed miRNAs with lung function parameters in experimental COPD in C57BL/6 mice: *A significant negative association was observed between Arnt2 expression and FEV0.1/FVC. A significant positive association was observed between Arnt2, Mfsd4, Myb, Slco2b1, and Ttc25 expression and total lung capacity (TLC) in the C57BL/6 COPD model during the time course of 12 weeks of exposure. Pearson correlation scatter plots with linear regression are shown in black line with its confidence interval represented by a grey area for expression of miRNA target genes and lung function parameters during the time course of CS-induced experimental COPD. Within each plot, pink and green circles represent the smoke-exposed and air-exposed control group, respectively, labeled with their corresponding timepoints. Data are shown in Pearson correlation coefficient r and significance in terms of p-value.*

3.5 Discussion

In recent years, studies have shown that modulating the expression of dysregulated miRNAs might improve COPD pathogenesis. Therefore, the primary goal of this study was to explore the longitudinal effect of smoke on the expression of miRNAs in lung tissues of mice using a CS-induced COPD mouse model (102, 241, 242, 293, 294, 296-298) and to propose new potential therapeutic targets through identifying new miRNA-gene pairs.

CS exposure has also been documented to induce alterations in lung function parameters (337-340). Some previous studies have elucidated the impact of smoking on the pulmonary function of individuals (337-341). To evaluate the lung function, the participants underwent an evaluation of their respiratory function, which included measurement of chest expansion, lung function assessment using a spirometer, and evaluation of respiratory muscle strength (337-341). These studies have consistently revealed a decrement in pulmonary function parameters, encompassing reductions in forced vital capacity (FVC), forced expiratory volume in one second (FEV_1), the ratio of FEV_1 to FVC (FEV1/FVC), and forced expiratory flow at 25–75% (FEF 25–75%) (341). Concurrently, the impact of CS exposure extends to changes in miRNA expression levels, as reported in relevant literature (295, 302, 303). Several studies in both humans and rats demonstrated that exposure to CS leads to the dysregulation of miRNAs (342-345). Despite these findings, the precise mechanisms through which cigarette smoke induces miRNA dysregulation remain incompletely understood.

This dual influence on both lung function parameters and miRNA expression underscores the intricate relationship between environmental exposures and molecular responses in the context of respiratory health. Therefore, we compared the miRNA expression with their lung function parameters in an independent human COPD cohort to identify the relationship between longitudinally dysregulated miRNA expression and disease severity. This endeavor aimed to

discern the associations between longitudinally dysregulated miRNA expression and disease severity. Notably, few miRNAs exhibited correlations, suggesting a potential influence on lung function.

We initially investigated the longitudinal impact of smoke exposure on miRNA expression in a Chronic Smoke (CS)-induced COPD model. Our findings highlighted miR-449a-5p and miR-135b-5p as the most significantly upregulated miRNAs compared to air-exposed mice, consistent with previous studies indicating elevated expression levels of these miRNAs in mouse lungs exposed to 24 weeks of chronic CS exposure (303). Notably, an increase in miR-135b-5p expression has also been documented in lung tissues of COPD smokers compared to non-smokers, suggesting a potential role in COPD pathogenesis (295).

Then we correlated the dysregulated miRNAs expression with genes that were altered due to the longitudinal effect of smoke exposure. A total of seven miRNAs (let-7b-5p, miR-30b-5p, miR-181a-5p, miR-181b-5p, miR-181c-5p, miR-150-5p, and miR-151-5p) with their validated and predicted targeted genes were significantly correlated during the longitudinal study. Notably, let-7b-5p and miR-181a-5p had more interactions with COPD-associated genes among their potential targets (Figure 3.4 and Figure 3.5). Our data suggest that let-7b-5p, miR-30b-5p, miR-151-5p, and miR-181(a,b,c)-5p may regulate the COPD pathogenesis via their potential disease-associated targets (e.g. *Arnt2, Bbox1, Kcnmb2, Myb, Slco2b1*) in our model. Consistent with our observation, the expression level of miR-181a-5p (346) and miR-181c-5p (163) has previously been found to be lower in the lung tissues of COPD mice. Similarly, lower expression of miR-181c-5p (163) has been observed in the lung tissues of COPD patients compared to non-smokers. Studies have found enrichment of the targeted genes of these remarkably down-regulated miRNAs in the Wnt signaling pathway, aryl hydrocarbon receptor

signaling, calcium and potassium channel activity, metabolism, and xenobiotic response elements/stress response (301).

Studies have validated the expression changes of miRNA associated with smoking, upon analyzing COPD patients and non-COPD controls (347). Moreover, findings from one study revealed a significant and inverse correlation between miRNA expression in COPD patients and the severity of airflow limitation, indicated by forced expiratory volume in 1 second (FEV_1) % predicted and the post-bronchodilator FEV_1/forced vital capacity (FVC) ratio, as well as with diffusing capacity of the lung for carbon monoxide (DL_{CO}) (348). Experiments conducted on miR-155 deficient mice revealed a significant attenuation in emphysema and lung functional alterations (348). These findings underscore the involvement of miRNAs in Chronic Smoke (CS)-induced inflammation and the pathogenesis of COPD, implicating miRNAs as a potential therapeutic target in COPD management.

In our investigation, the longitudinal analysis of differentially expressed miRNAs in conjunction with lung function parameters unveiled an association of *Arnt2* on specific lung function parameters, namely the Forced Expiratory Volume at 0.1 second to Forced Vital Capacity (FEV0.1/FVC) ratio and Total Lung Capacity (TLC), as depicted in Figure 3.9. A low FEV0.1/FVC ratio typically indicates airflow obstruction, while elevated TLC often signifies increased lung volume or hyperinflation. Our findings suggest an inverse correlation between *Arnt2* expression and FEV0.1/FVC, while a direct correlation was observed between *Arnt2* expression and TLC in our model. This observation aligns with the notion that *Arnt2* may potentially linked with impair lung function.

The major strength of this study is that we have explored the longitudinal effect in the COPD mice model and have highlighted changes in miRNA expression over time. Another strength

of our research is the co-analysis of miRNA and gene-expression changes over time, which helped identify the significance of experimentally validated and predicted miRNA targets.

A few limitations and some caveats are worth noting while inferring our results. Firstly, our study's miRNA and gene expression profiles were not from the same samples. Instead, the data generated was from independent but the same CS-induced COPD models. Therefore, for correlation, we calculated the mean of normalized expression at each time point to include the impact of every sample collected. Nevertheless, miRNA and gene expression profiling from the same samples would be more descriptive and correlational. Secondly, the longitudinal gene expression profiling through the RNA-Seq method would be more valuable and informative than microarrays for such studies due to its higher sensitivity and incomparable accuracy (349, 350). Our study's sample size is also limited, and applying multiple correction tests could markedly diminish statistical power, thereby complicating the detection of genuine effects. Furthermore, given the exploratory nature of our analysis, we have opted to exercise caution in result interpretation and prioritize the replication of findings in future investigations. Lastly, a closer investigation of the identified system in a serially designed human experiment can help understand the underlying biological mechanisms. Currently, to the best of our knowledge, there are only very few datasets like "NORM" and "GLUCOLD" (153, 308, 318) studies in which both miRNA and longitudinal gene profiles are available, reflecting an urgent need for more studies on this topic.

Overall, we have profiled miRNA expression in lung tissues of mice over a longitudinal time course of the development (8 weeks) and progression (12 weeks) of experimental COPD and correlated their levels with the expression of their putative target genes over time. We show that the predicted targets of these miRNAs are altered by longitudinal CS exposure. miRNAs, specifically let-7b-5p and miR-181a-5p, are associated with decreased lung function in human

COPD. We found seven miRNAs of interest that could regulate the expression of their target genes and promote COPD pathogenesis. Among them, four miRNAs (let-7b, miR-151-5p, miR-181a-5p, and miR-181c-5p) significantly correlated with their targeted genes including *Arnt2, Bbox1*, and *Ttc25* in a separate COPD model.

In summary, we identified miRNA-181a-5p, miR-181c-5p, and let-7b-5p as potential regulators of *Arnt2*, with likely roles in modulating its expression during disease development and progression. Notably, these regulators demonstrate a negative correlation with *Arnt2* expression levels. Consequently, targeting these regulators, potentially through overexpression strategies, may offer therapeutic benefits in mitigating disease progression. These findings present novel avenues for investigating therapeutic targets in the context of COPD, offering promising prospects for future research endeavors aimed at addressing the disease.

3.6 Proposed Mechanism

In the context of AhR signaling, the induction of *Cyp1a1* serves to suppress apoptosis. However, our observation hypothesizes that high levels of *Arnt2* during COPD development and progression may impede this pathway, contributing to increases apoptosis or cell destruction (Chapter 2). Additionally, we postulated that miR-181a-5p, miR-181c-5p, and let-7b-5p play a role in downregulating *Arnt2* expression. Our COPD model revealed diminished expression of these miRNAs (miR-181a-5p, miR-181c-5p, and let-7b-5p) which significantly correlated with elevated levels of *Arnt2* (Chapter 3). Therefore, we proposed that levels of let-7b, miR-181a-5p, and miR-181c-5p might regulate the expression of *Arnt2* that in turn possibly be involved in regulating the XREs, particularly *Cyp1a1*, which leads to the COPD development and progression (Figure 3.10).

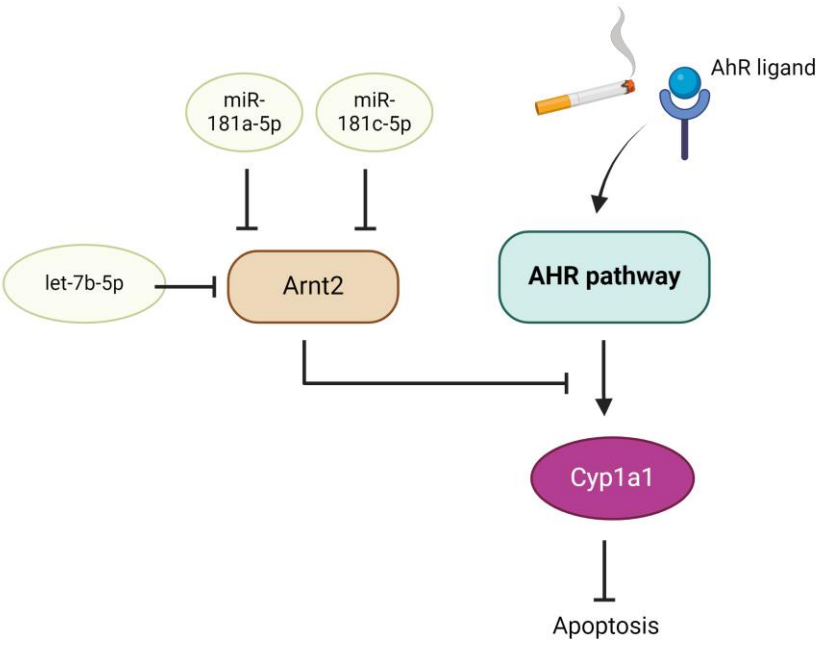

Figure 3.10: Proposed mechanism of miRNAs regulating the targeted gene *Arnt2*. *We proposed that let-7b-5p, miR-181a-5p, and miR-181c-5p are the negative regulators of Arnt2, which may be a potential repressor of the AhR pathway in CS-induced COPD. (Created with BioRender.com)*

Chapter 4: Genome-wide CRISPR-Cas9 Knock-out library screening to identify protective factors against cigarette smoke

4.1 Abstract

Introduction: Smoking is an established risk factor for respiratory diseases in particular chronic obstructive pulmonary disease (COPD). Cigarette smoke (CS) contains thousands of toxic substances that cause cell apoptosis. Elucidating the molecular biology of CS-induced death of lung structural cells has the potential to define new mechanisms that underpin the pathogenesis of COPD and define new therapeutic targets. Thus, we investigated the genes essential for survival against challenge with CS extract (CSE) in A549 and BCi-NS1 cells.

Methods: We performed CRISPR/Cas9 whole-genome pooled knockout (KO) screening with the Toronto Knockout (TKOv3) library, a CRISPR-Cas9 gene editing system. It has a 70,948-single guide RNAs (sgRNAs) library that targets 18,053 protein-coding genes (4 sgRNA/gene) in the genome. Cells were infected with TKOv3 3rd generation lentiviral library at a low multiplicity of infection (MOI = 0.3) to achieve ~300-400-fold coverage of the library. Transduced cells were challenged with an 80% lethal dose (LD_{80}) achieved with 3 consecutive days of CSE exposure (20% for A549 and 5% CSE for BCi NS1.1). Surviving cells were collected and sequenced to detect the alterations in sgRNA counts.

Results: Using the MAGeCK algorithm we found 31, 7, and 237 depleted genes in CSE-exposed A549 KO cells at three different timepoints namely T1, T2, and T3 with FDR < 0.1. Screening with CSE challenge in BCi NS1.1 KO cells identified 103 hits with FDR <0.1. Among them, the identified top 6 genes from A549 initial CSE exposure screen were *ANK2, BPIFB4, MMP15, MRPL11, NMI, and SNX31* with FDR < 0.1. Functional enrichment analysis shows the enriched pathways related to cellular response to oxidative stress, ribosomal and

protein biogenesis, and processing with adj. $p < 0.05$. Additionally, we identified *SEMA4F, ETV3L, YTHDF1*, and *FLYWCH2* as top hits in the BCi NS1.1 screen with FDR < 0.1.

Conclusion: Based on the biological relevance of identified hits, our findings have implications for understanding the potential mechanism of underlying CSE-induced cell death. However, there is a critical need to validate the identified hits individually.

Keywords: CRISPR-Cas9, Cigarette smoke, Library, Whole-genome, COPD.

4.2 Introduction

Cigarette smoke (CS) is an intricate mixture of over 4000 compounds including various toxins (351, 352), that could stimulate harmful effects on the respiratory tract. Toxic substances present in CS can lead to cellular malformations that can potentially lead to the pathogenesis of smoke-related lung diseases (353-355). COPD is one of the widely known lung diseases greatly affected by CS. Exposure to CS stimulates the production of reactive oxygen species (ROS) in the lung tissues which in turn can imbalance the cellular anti-oxidative system, airway inflammation, and lung emphysema, which are the key mechanisms of COPD onset (356).

Long-term CS can damage the structure of the air duct wall, alveolar septum (357-360), and bronchioles and increase mucous production that can lead to obstructive bronchiolitis (361, 362) which exacerbates lung tissue damage. These physiological changes within lung tissues due to CS can induce apoptosis (363-365). Apoptosis is one of the important upstream events that initiate and participate in the pathogenesis of COPD (366, 367). Both alveolar and endothelial apoptotic cells were found to be increased in the lungs of COPD patients (368). The rate of pulmonary structural cell apoptosis and necrosis cannot be redeemed by proliferation, thus causing lung tissue destruction that leads to the development of emphysema. The potential regulatory mechanisms of CS-induced lung structural cell death are complicated and need to be fully understood to maximize the resistance against CS -induced cell death.

The clustered regularly interspaced short palindrome repeat (CRISPR) associated Cas9 screening has revolutionized the research approach to identify new biological mechanisms and pathophysiology of diseases associated with specific cellular phenotypes. In CRISPR genomics screen role of individual genes is investigated against a phenotype by multiplexing thousands of gRNAs targeting the whole genome. CRISPR/Cas9 inhibition (loss-of-function) library has

been greatly used for the positive or negative selection of candidate genes (369), in a wide range of signaling pathways and biological processes (370).

We hypothesized that genome-wide CRISPR-Cas9 knock-out screening in human epithelial cell lines challenged with CS exposure will help determine the genes that are protective against CS. Thus, we performed an *in-vitro* screen (186, 191, 371, 372) in A549 and BCi-NS1.1 cell lines to identify the genes and pathways required for resistance against CS damage. We systematically cataloged the factors in human cells that were resistant to CS-induced death. For this, the pool of knock-out cells was exposed to high concentrations of CSE (20% for A549 and 5% CSE for BCi NS1.1) that induced 80% cell death, and the remaining cells were then subjected to next-generation sequencing. We identified numerous therapeutically targetable genes whose inactivation may significantly increase resistance to CSE-induced cell death. This approach has not been previously used to characterize the effects of CSE on cell death. This will define new therapeutic targets against CS-induced cell death that leads to emphysema and COPD.

4.3 Methodology

4.3.1 Cell lines

Adenocarcinoma human alveolar basal epithelial cells (A549), BCi-NS1.1 (human airway epithelium-derived basal cell line), and epithelial-like human embryonic kidney cells (HEK293T) cells were obtained from the American Type Culture Collection (ATCC). A549 cells were grown in RPMI-1640 supplemented with 5% fetal bovine serum (FBS), and 1% (100U/ml penicillin/streptomycin). The BCi-NS1.1 cell line was grown in BEGM™ Bronchial Epithelial Cell Growth Medium (CC-3171) supplemented with BEGM™ SingleQuots™ Supplements and Growth Factors (CC-4175). HEK293T cell line was grown in 10% FBS

medium supplemented with 1% (100U/ml penicillin/streptomycin). Cells kept growing under 5% CO_2 at 37ºC.

4.3.2 Human Toronto knockout CRISPR library version 3 (TKOv3)

Human TKOv3 CRISPR lentiviral library (Addgene Cat # 90294) containing 70,948 guides (4 guides/gene) targeting 18,053 genes was used in our screens (373). The library also contains 142 control guides targeting luciferase, LacZ, and EGFP. We used a one-component library (LCV2:TKOv3, Addgene pooled library #90294) that expresses Cas9 and sgRNAs on a single vector designed for genome-scale loss of function screens in mammalian cells. The library was amplified as per instructions given by MOFFAT LAB (University of Toronto).

4.3.3 Genome-scale lentiviral TKOv3 gRNA library construction

The lentiviral packaging cells HEK293T and plasmids psPAX2 (Addgene plasmid # 12260) and pCAG (Addgene plasmid # 35616) were used to generate large-scale lentiviral production of TKOv3 CRISPR library. The combination of plasmids used for transfection was a 3:3:1 DNA ratio of TKOv3 plasmid:psPAX2:pCAG in the form of DNA-lipid complex by using Lipofectamine™ 3000 Reagent. HEK293T cells at 80% confluency were transfected with DNA-lipid complex drop-wise over the surface. Media was changed 20 hours post-transfection to remove the transfection reagents. Forty-eight- and seventy-two-hours post-transfection media containing the virus was harvested. Viral media was spun down at 1000g for 5 mins at room temperature (RT) to remove the cell debris and filtered through 0.45um filter PES membrane syringe filter Millipore. For BCi-NS1.1 screening, after filtration of viral media, the virus was pelleted by ultracentrifugation at 3000g for 30-45 minutes at 4^0C using Piece protein concentrator PES, 100K MWCO (ref # 88533). The virus pellet was then resuspended in fresh media (2ml/flask). For A549 screening, the generated virus was concentrated using the

polyethylene glycol (PEG) precipitation method. For every 30ml filtered viral media 10ml of PEG was mixed and incubated overnight at 4^0C. The following day, the virus was pelleted by ultracentrifugation for 30 minutes at 3200g at 4^0C. Subsequently, the virus pellet was resuspended in fresh media (2ml/flask). Lentivirus aliquots were stored at -80^0C. For the human TKOv3 library, 5 x 175 flasks were seeded with $20x10^6$ HEK293T cells per flask.

4.3.4 Determination of lentivirus multiplicity of infection (MOI) for transducing the cell lines

Before the screen, the multiplicity of infection (MOI) of 30% (374) to integrate only one plasmid sequence into the genome of each cell during transduction was calculated. The MOI of lentivirus was determined with the aim of infecting cells with no more than one virion per cell. After infection, the integrated lentivirus sequence along with the cloned CRISPR Cas9 and gRNA expressed, eventually results in the production of a knockout cell for the integrated gRNA sequence (375).To determine the MOI for our pooled screens, $60x10^3$ cells were seeded in 24 well plates. Aliquots of the virus were thawed on ice. For BCi NS1.1 transduction, 3x dilution of the stock virus was used to define a range of viral titers from 0 to 32ul. For A549 transduction, 10x dilution of the stock virus was used to define a range of viral titers from 0 to 10ul. To enhance the lentivirus transduction efficiency 8ug/ml polybrene (Merk, Cat # TR-1003-G) with 1:1250 dilution into media (final volume of 500ul) was added per well. Each MOI was tested in 3 replicates with uninfected cells as a control. Cells were incubated overnight in 5% CO_2 at 37^0C. 24 hours post-transduction cells were selected for puromycin resistance for 72 hours to remove the transduced cells. Cell viability was determined by resazurin assay. We determined the MOI of 0.3 to maximize the number of cells with a single integration. For the A549 screen, a 2ul viral titer of 10x dilution of stock for 60,000 cells was selected, where 30% of the cell population was transduced with the lentivirus. Where, for BCi

NS1.1, 8ul viral titer of 3x dilution of stock for 60,000 cells was finalized to achieve the MOI of 0.3 (30% transduction efficiency) (Figure 4.1A).

4.3.5 Working concentration of puromycin for mammalian cell lines

To ensure the positive selection of transduced cells, mammalian cells were grown in puromycin antibiotic-containing media. The TKOv3 vector cassette also codes for the puromycin-N-acetyltransferase (PAC) gene. Sensitivity and response time to puromycin varies according to the cell type and cell culture conditions. Thus, we determined the minimum concertation of puromycin that is required to select all the positive A549 and BCi-NS1.1 cells within 72 hours. For this, we perform a kill curve using different concentrations of puromycin ranging from 0 to 32ug/ml. we seeded $10x10^3$ cells/well in opaque 96-well plates in 100ul media in triplicates. The following day, the media was replaced with puromycin-containing media, and cells were allowed to grow for 72 hours. After 72 hours being growing in the puromycin media, cell viability was determined by resazurin cell viability assay. Resazurin solution was prepared by dissolving 30ug resazurin powder in 1ml 1 x PBS. Cells were incubated for 3-4 hours in 10ul/ml resazurin solution in growth media. Measurements were taken using the Omega plate reader (BMG LABTECH) with excitation filter: EX544 and emission filter: 590-10. A minimum concentration of puromycin that killed 100% of uninfected cells was used for the whole genome knockout screen. We found that 1.5ug/ml and 16ug/ml of puromycin concentration are required to select 100% transduced A549 and BCi NS1.1 cells respectively when treated with puromycin for 3 days (Figure 4.1B). To ensure the complete removal of non-transduced cells, during the actual screen cells were allowed to keep growing in the determined concentration of puromycin media for 7 and 10 days for A549 and BCi NS1.1 cells, respectively.

4.3.6 Determination of cell death curve in cell lines against CSE

Next, we determined the CSE concentration that is required to kill 80% of CSE-exposed cells. We seeded $12.5x10^3$ A549 and $10x10^3$ BCi NS1.1 cells in 96 well plates followed by overnight incubation. For CSE preparation, smoke from one filter-less 3R4F reference cigarette (12mg tar and 1.0mg nicotine/cigarette) was continuously drawn through 10 ml of pre-warmed serum-free culture media. The resulting suspension was then filtered through sterile syringe filters 0.22um pore size. The obtained solution was considered 100% CSE and used within 20 minutes of preparation. The resulting 100% CSE was diluted to the desired concentration with a culture medium. Cells were then exposed to different concentrations of CSE for 3 consecutive days. CSE media was refreshed after every 24 hours. After 3 days of the CSE exposure challenge the cells were measured by MTT assay or by cell counting after trypan blue staining using a hemocytometer. Ten microliters of yellow tetrazolium salt (MTT) solution (5mg/ml) were added to each well. The cells were incubated for about 3-4 hours at 37^0C. MTT was removed, and the precipitants were solubilized in 100ul DMSO (Dimethyl sulfoxide). Measurements were taken at 570nm using a microplate reader. We observed 20% and 5% CSE is required to achieve 80% A549 and BCi NS1.1 cell death, respectively, upon 3 days of CSE exposure (Figure 4.1C).

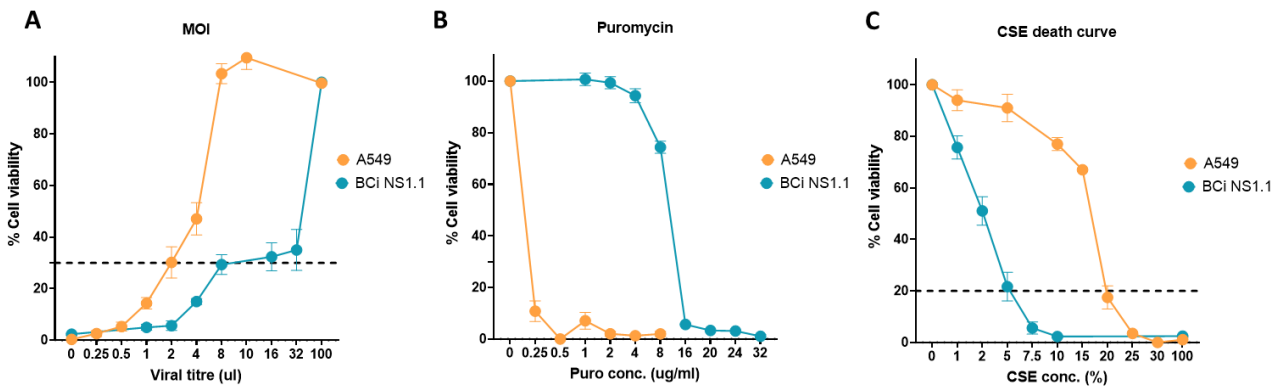

Figure 4.1: Estimation of percentage of viable A549 and BCi NS1.1 cell line: ***A)*** *Multiplicity of infection (MOI) with TKOv3 lentivirus library. The horizontal dotted line represents 30% cell viability, which corresponds to a 30% transfection efficiency of lentivirus.* ***B)*** *Puromycin concentration used for the selection of transduced cells.* ***C)*** *Determination of kill curve against CSE exposure for 3 days. The horizontal dotted line signifies 20% cell viability, indicative of achieving 80% cell death with the corresponding CSE concentration. Data (n=3 independent experiments with at least 3 technical replicates) is represented as means ± SEM. SEM: standard error of the mean.*

4.3.7 Large-scale pooled CRISPR depletion screening against CSE

About ~150–200 million cells at a low MOI of ~0.3 were infected with the CRISPR library to maintain the coverage of sgRNA. We transduced the A549 and BCi NS1.1 cells with a human TKOv3 lentiviral packed library at a low MOI of ~0.3. The cell suspension in the growth medium was supplemented with final polybrene concentrations to 8ug/ml for effective transduction. The media was changed after 24 hours to remove the polybrene. The transduced cells were then selected with 1.5ug/ml of puromycin (Sigma-Aldrich) media for 1 week to generate a mutant cell pool. Transduced cells were then allowed to grow for an additional 2-3 days for complete knockout of sgRNA-targeted genes. Approximately, $30x10^6$ cells (400x fold coverage) were collected for the T0 timepoint, and the remaining cells were seeded for screening against CSE. Cells were exposed to 5% and 20% CSE for BCi NS1.1 and A549 cells,

respectively, for 3 consecutive days. CSE media was refreshed after every 24 hours of exposure. After 3 days of exposure cells were allowed to grow for additional 2-3 days to recover from the stress and grow. Cells were then collected for T1 time points. Cells growing in normal media were considered control cells. For A549 three biological replicates (RepA, RepB, and RepC) were included (Figure 4.3A). After each round of exposure, $30x10^6$ cells were collected for genomic DNA extraction to ensure 400x coverage of the TKOv3 library. An overview of the pooled genome-wide CRISPR/Cas9 knock-out library screening is illustrated in Figure 4.2.

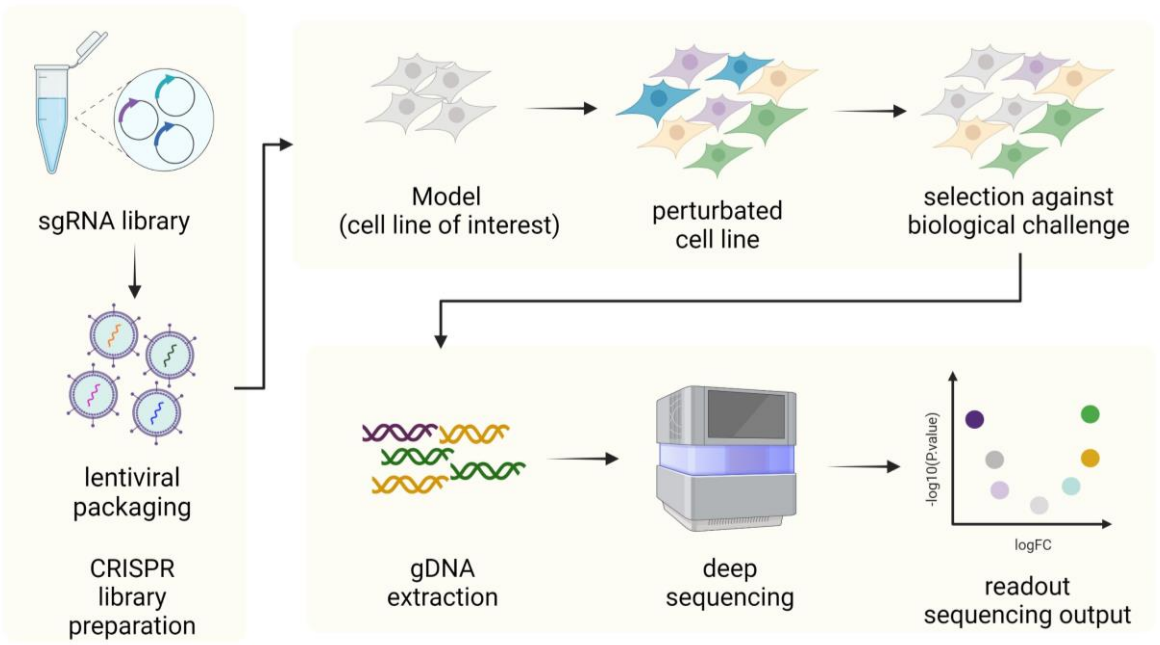

Figure 4.2: **Pooled genome-wide CRISPR library screen workflow:** *Schematic illustration of the workflow of genome-wide CRISPR-Cas9 knockout library screenings started with sgRNA library preparation to lentiviral packaging of sgRNA containing plasmids. The experimental cell line was subsequently infected with the lentivirus-packed CRISPR library to generate the pool of knockout cell population that underwent screening against a biological challenge. Genomic DNA was extracted from the surviving cells after screening and sequenced to determine the copy number of sgRNAs by deep sequencing. The resulting alterations in the number of gRNAs in cells that experienced a biological challenge as compared to control help determine the responsive genes. (Created with BioRender.com)*

4.3.8 Genomic DNA extraction and precipitation

Genomic DNA was isolated from control and CSE exposed surviving cells. Genomic DNA extraction was performed by BIO-52066 ISOLATE II Genomic Kit as described in the kit manual. The purity of genomic DNA was measured in terms of an A260/280 ratio by using NanoDrop for total nucleic acid content.

4.3.9 Library preparation for next-generation sequencing (NGS)

High-throughput sequencing was performed to determine changes in sgRNA representation compared to the control. CRISPR library for Illumina NGS was prepared from extracted genomic DNA through two-step PCR amplification. During 1st step of PCR the sgRNA cassette integrated into the genomic DNA was amplified to maintain the required coverage of the library and in the 2nd step of PCR Illumina TruSeq adapters with i5 and i7 indexes were attached with amplified gRNAs. For PCR2, forward primers (i5) were used as staggered primers in equal ratios with unique reverse primers (i7). During library preparation, these unique indexes were added to the PCR1 product to help in identification during multiplexing. DNA was amplified using NEBNext High-Fidelity 2X Master Mix (New England Biolabs) during the PCR1 and PCR2 amplification steps. For PCR1 and PCR2, primers were annealed at 66 °C. PCRs product was confirmed on SYBR-safe stained 2% agarose gel. The ~250bp amplicon was gel-extracted using Bioline Isolate II PCR and gel kit (BIO-52060). The quality of purified gel-extracted amplicon was assessed using nanodrop.

4.3.10 High-throughput DNA sequencing and computational analysis

Purified samples from PCR2 were subjected to massive parallel amplicon sequencing using a NovaSeq PE150 sequencing technology carried out by NovogeneAIT Genomics. Samples were demultiplexed based on i7 indices. MAGeCK/VISPR version 0.5.6 (376, 377) and

MAGeCKFlute were used for analyzing the genome-wide CRISPR KO screen. Sequenced sgRNAs were imported from raw FASTQ files and aligned to a reference file of all possible guide RNAs present in the TKOv3 library through the MAGeCK count module. Approximately 70% sequenced reads per sample successfully aligned to the reference library in both screens (Figure S4.1). For each library sequence, a table of raw counts was created based on uniquely aligned reads. The generated read counts were analyzed to test the differential enrichment of guides and ranked the gRNAs and genes using a robust ranking aggregation (α-RRA) algorithm available in the MAGeCK test module. MAGeCK α-RRA calculates the statistically significant positive or negative selected genes. Read counts were normalized with negative control sequences. MAGeCK test outputs the summary results at both the sgRNA and gene levels. Based on gene-level analysis, significant hits were identified with the FDR cut-off < 0.1.

4.3.11 Functional analysis of differentially enriched guides

Functional analysis was performed through Enrichr (378, 379) and the REACTOME database (380, 381) to explore the biological processes and signaling pathways in which significantly enriched guides targeting genes are involved. A list of significant positive selected genes targeted by enriched guides was used as search terms for the Enrichr and REACTOME databases. Genes represented by 1 guide were excluded from functional analysis as it might lead to false positive results. Hits were also searched against cancer dependency map project DepMap (382) to highlight the cancer vulnerabilities.

4.3.12 Statistical Analysis

Cell viability percentages in response to puromycin concentration, CSE, and MOI were evaluated utilizing GraphPad Prism (v8) (265). Data were derived from three independent experiments, each with at least three technical replicates. The variance among biological

replicates was assessed using the standard error of the mean (SEM). For the A549 screen, three independent experiments were conducted (n=3), whereas for BCi NS1.1, only one experiment was performed. Following the genome-wide CRISPR/Cas9 knockout screen, subsequent analysis aimed at identifying essential genes (FDR < 0.1) was performed using MAGeCK (376, 377).

4.4 Results

To identify the genes and pathways that are essential for resistance against CS-induced damage, A549 knock-out cells underwent three exposures to the CSE challenge (Figure 4.3A), whereas BCi NS1.1 knock-out cells were exposed to the CSE challenge only once.

4.4.1 Identification of sensitive genes to CSE in A549

Differential enrichment of gRNAs in CSE exposed KO cell population compared to unexposed KO population during the first round (termed as "time point T1") in A549 cells (n=3) showed enrichment of 31 guide targeting genes with FDR <0.1 (Figure 4.3B and Table S4.1). The most significantly enriched gRNAs in the A549 selected pool were found to target *CDK1, OR6K6, UPF2, CHEK1,* and *NOL6* (Figure 4.3C-D).

The A549 population from round 1 was exposed for an additional 3 days (round 2 termed as "time point T2") (n=2). Differential enrichment of gRNAs from round 2 exposure (n=2) showed enrichment of only 7 gRNA targeting genes with FDR < 0.1 (Figure 4.3E and Table S4.2). The significantly enriched multiple gRNAs were targeting *YEATS4, ZBTB17, HEATR1, RRM1,* and *MMS22L* (Figure 4.3F-G).

Lastly, cells from round 2 were exposed once again to CSE for 3 days referred to as time point T3. This round only comprised cells from one biological replicate (n=1) as most cells died by the time they reached the third round due to the strong cytotoxic effect of CSE. Following three

rounds of exposures to 20% CSE, the analysis of gRNA enrichment revealed 237 guides targeting genes with FDR of less than 0.1 (Figure 4.3H and Table S4.3). The topmost positive selected genes were *MTSO1, NF2, KRAS, PDCD2, ELL,* and *DDX49* (Figure 4.3I-J). Interestingly, there are no negative selected genes were observed with FDR < 0.1.

We did not observe consistent significant alterations of guides for any of the genes across all time points (Figure 4.6D). This inconsistency may be attributed to the severe and continuous exposure to CSE for 3 rounds, which were performed at different times, potentially resulting in divergent cellular responses and gene expression profiles. Additionally, the presence of knockout cells, which become cancerous following gene knockout, might overshadowed the alterations, as these cells proliferate at a faster rate compared to non-cancerous cells and protective knockout cells as the cells kept growing for multiple exposures.

4.4.2 The biological relevance of potential functional hits from the A549 screen

The biological importance of significantly enriched guides targeted hits from all time points of the A549 screen was explored by investigating the gene ontology for biological processes (Figure 4.4A). Several enriched biological processes found at time point T1 (FDR < 0.05) included cellular protein metabolic process, RNA processing, ribosome biogenesis, and rRNA processing. Likewise, enriched guides target genes at the time point T2 were involved in the regulation of rRNA processing. In regards to enriched guides target genes from timepoint T3 selected *vs* unselected group, enrichment of rRNA metabolic process, ribosome biogenesis, regulation of transcription from RNA polymerase II promoter in response to stress, nuclear-transcribed mRNA catabolic process, cellular response to hypoxia, and cellular response to oxidative stress was observed (Figure 4.4A). REACTOME-based pathway enrichment analysis further confirmed the results of GO enrichment analysis (Figure 4.4B). The complete list of significant pathways and enriched genes can be seen in Table S4.4, Table S4.5, and Table S4.6.

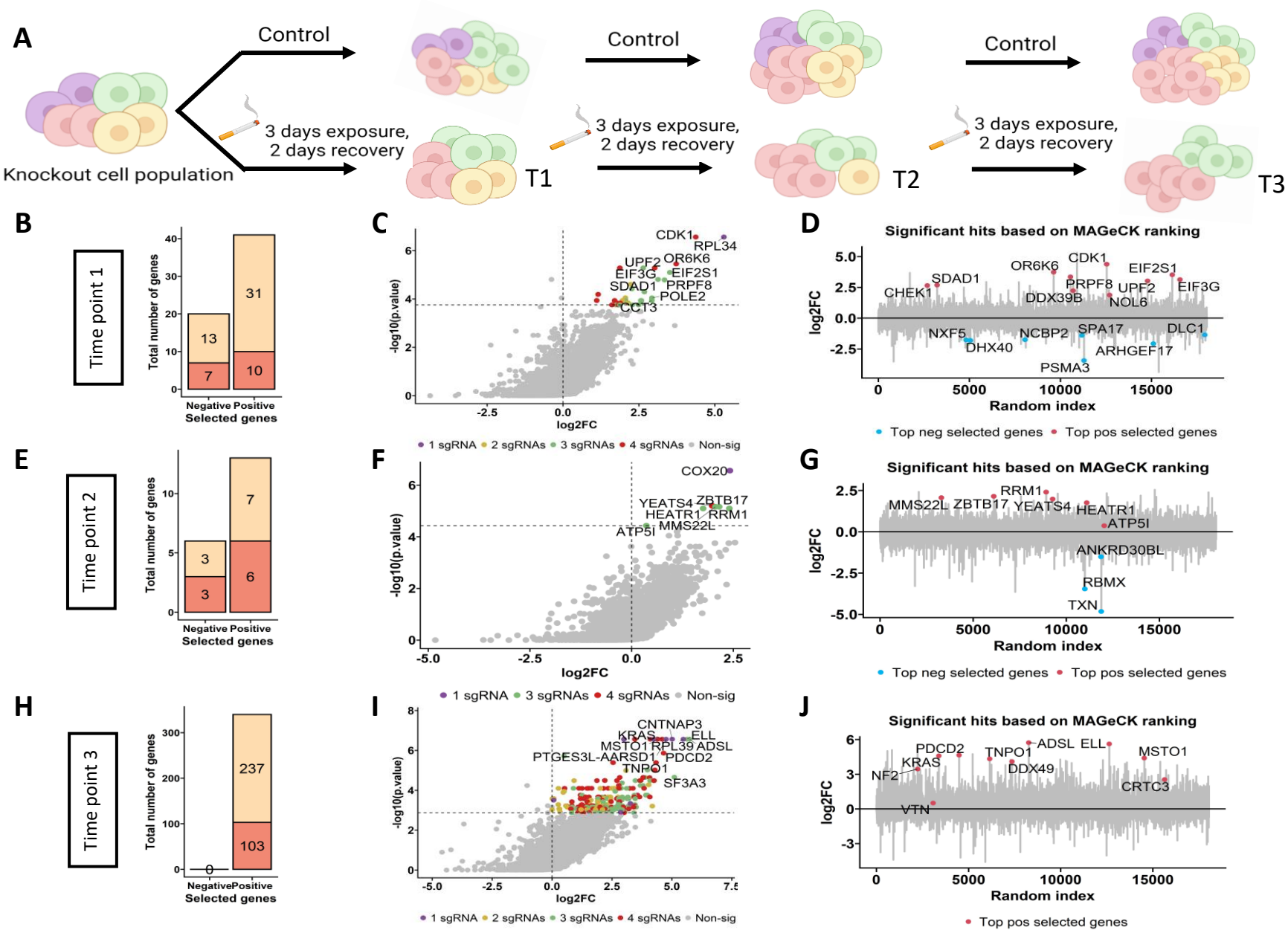
A
Control
Control
Control
3 days exposure, 2 days recovery
3 days exposure, 2 days recovery
3 days exposure, 2 days recovery
Knockout cell population
T1
T2
T3
B
Time point 1
Total number of genes
Negative Positive
Selected genes
13
7
31
10
C
CDK1
RPL34
UPF2 OR6K6
EIF3G EIF2S1
SDAD1 PRPF8
POLE2
CCT3
-log10(p.value)
log2FC
1 sgRNA 2 sgRNAs 3 sgRNAs 4 sgRNAs Non-sig
D
Significant hits based on MAGeCK ranking
log2FC
Random index
Top neg selected genes Top pos selected genes
E
Time point 2
3
3
7
6
F
COX20
YEATS4 ZBTB17
HEATR1 RRM1
ATP5I MMS22L
G
Significant hits based on MAGeCK ranking
RRM1
MMS22L ZBTB17 YEATS4 HEATR1
ATP5I
ANKRD30BL
RBMX
TXN
H
Time point 3
0
237
103
I
CNTNAP3
KRAS ELL
MSTO1 RPL39 ADSL
PTGES3L-AARSD1 PDCD2
TNPO1
SF3A3
J
Significant hits based on MAGeCK ranking
PDCD2
KRAS
TNPO1
ADSL ELL
MSTO1
DDX49
NF2
CRTC3
VTN

Figure 4.3: Distribution of significantly enriched guides targeting genes in CSE exposed knock-out A549 cells: ***A)*** *Workflow of CRISPR screen for collecting data at different time points (T1 (n=3), T2 (n=2) and T3 (n=1)).* ***B)*** *A stacked bar plot for T1 representing the total number of guides targeting genes significantly changed. The dark color portion represents the hits with FDR <0.05, light color portion represents the hits with FDR < 0.1.* ***C)*** *Volcano plot representing statistically significant guides targeting genes with FDR <0.1 at T1. The y-axis represents the -log_{10}(p-value) and the x-axis represents the log_2fold change. Each dot on the volcano plot with different colors represents the number of good sgRNA for selected genes.* ***D)*** *Scatterplot at T1 with randomly indexed guides targeting genes highlighted the top positive or negative selected genes with log_2FC on the y-axis and random indexed on the x-axis.* ***E)*** *Stacked bar plot for T2 representing the total number of guides targeting genes significantly changed. The dark color portion represents the hits with FDR <0.05, light color portion represents the hits with FDR < 0.1.* ***F)*** *Volcano plot representing statistically significant guides targeting genes with FDR <0.1 at T2. The y-axis represents the -log_{10}(p-value) and the x-axis represents the log_2fold change. Each dot on the volcano plot with different colors represents the number of good sgRNA for selected genes.* ***G)*** *Scatterplot at T2 with randomly indexed guides targeting genes highlighted the top positive or negative selected genes with log_2FC on the y-axis and random indexed on the x-axis.* ***H)*** *A stacked bar plot for T3 representing the total number of guides targeting genes significantly changed. The dark color portion represents the hits with FDR <0.05, light color portion represents the hits with FDR < 0.1.* ***I)*** *Volcano plot representing statistically significant guides targeting genes with FDR <0.1 at T3. The y-axis represents the -log_{10}(p-value) and the x-axis represents the log_2fold change. Each dot on the volcano plot with different colors represents the number of good sgRNA for selected genes.* ***J)*** *Scatterplot at T3 with randomly indexed guides targeting genes highlighted the top positive or negative selected genes with log2FC on the y-axis and random indexed on the x-axis. T1 has n = 3 biologically independent replicates, T2 has n =2 biologically independent replicates and T1 has n =1 biological replicate. (Panel A: Created with BioRender.com)*

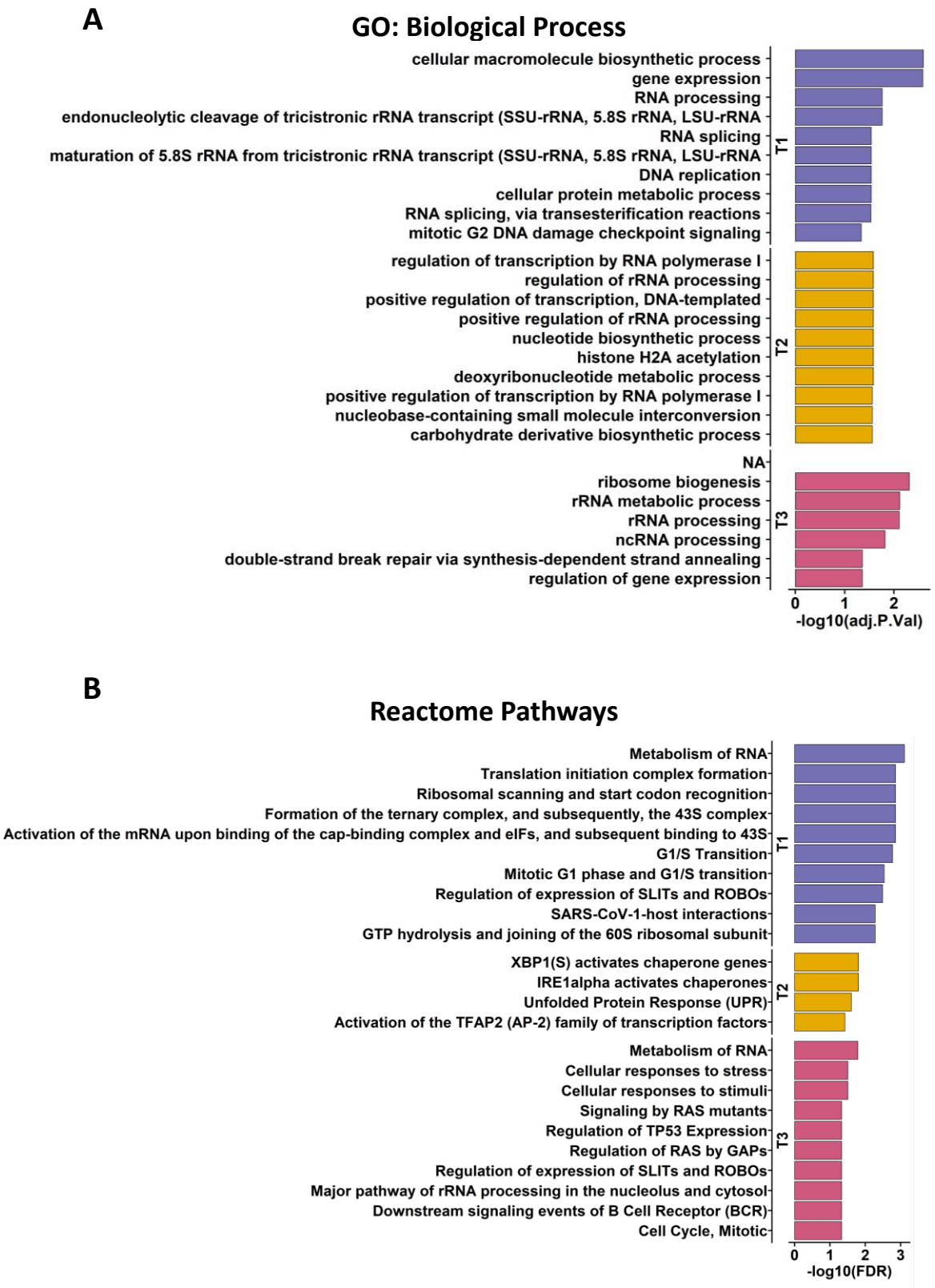

Figure 4.4: Functional enrichment analysis for identified hits from the A549 screen: *Each color represents a different time point. The x-axis shows -log10 false discovery rate (FDR).*

4.4.3 Selection of biologically relevant enriched sgRNA guides targeting genes from the A549 screen

Following the identification of positive selected genes, we proceeded to explore the biologically relevant genes. For this, the subsequent step involved screening for the most vulnerable knockouts, specifically targeting genes associated with tumor suppression. Knocking out these genes can result in increased proliferation and progression of cancer. Thus, the positive selected genes were searched against a cancer dependency map DepMap (382), to filter out the genes associated with cancer development. As expected, we identified several cancer causative genes including *NF2*, *TP53*, *VHL*, and *RAD51*. To remove the cancerous genes, the search against DepMap, a cancer dependency map highlighted 6, and 66 functionally relevant hits from time points 1 and 3, respectively, with gene effect chronos score ranges from -1 to 0.5. We could not find any significant hits for time point 2 during our search against DepMap with the above-mentioned score criteria. Fold change pattern of 6 potential hits from 1st round of CSE exposure (T1) including *ANK2, BPIFB4, NMI, MMP15, MRPL11,* and *SNX31* were explored at gene as well as sgRNA level at each time point (Figure 4.5A-B). On the other hand, 13 out of 66 potential hits (*ALPK1, C3orf18, CCDC96, CDHR1, CELA3B, CSRNP1, DNAJA4, EZH1, ITPR1, KIF1C, PAPPA2, SLC17A8,* and *SLURP1*) from 3rd round of CSE exposure (T3) is shown in (Figure 4.5C-D). Interestingly, most of the hits also show positive logFC throughout the screening against CSE exposure (Figure 4.5C-D).

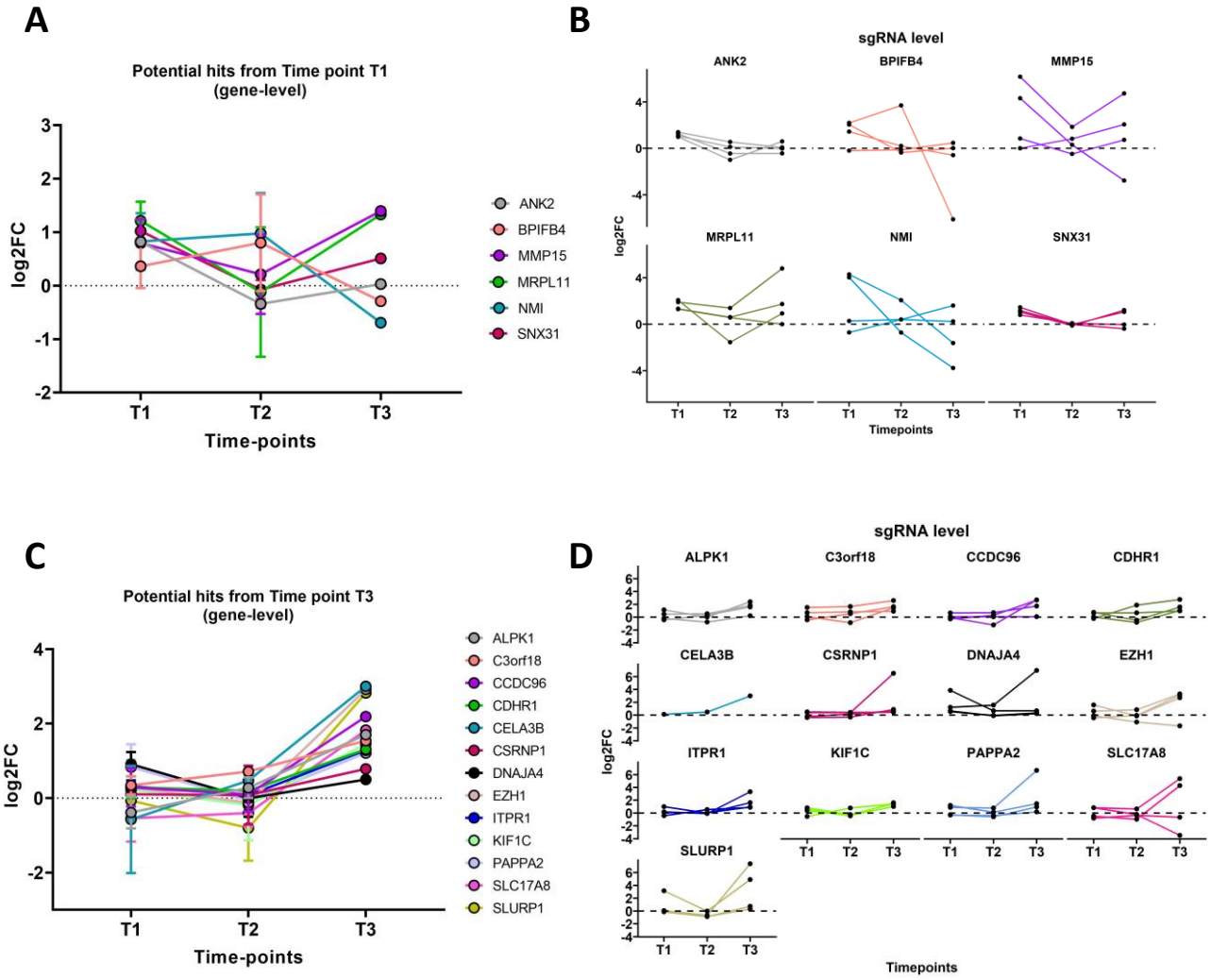

Figure 4.5: Gene level and sgRNA level analysis of top identified hits from T1 and T3 of A549 screen: ***A)*** *Top selected hits from time point 1 showing the distribution of guides targeting genes (at gene level) at each time point.* ***B)*** *Top selected hits from time point 1 showing the distribution of guides targeting genes (at sgRNA level) at each time point.* ***C)*** *Top selected hits from time point 3 showing the distribution of guides targeting genes (at gene level) at each time point.* ***D)*** *Few of selected hits from time point 3 showing the distribution of guides targeting genes (at sgRNA level) at each time point.*

4.4.4 Identification of genes sensitive to CSE in BCi NS1.1 cells

To replicate our findings in a non-cancerous cell line we exposed the minimally immortalized broncho-epithelial cell line BCi NS1.1 with 5% CSE that kills 80% of the KO cell population after 3 days of exposure. Differential enrichment of guides in CSE exposed BCi NS1.1 knockout pooled cell population as compared to unexposed cells showed enrichment of 103 guides targeting genes (FDR <0.1) (Figure 4.6A and Table S4.7). Interestingly, there were no negative selected genes observed with FDR < 0.1. Based on MAGeCK ranking, *SEMA4F, DEFB129, ETV3L, YTHDF1*, and *FLYWCH2* were identified as top positive selected hits (Figure 4.6B) having more than one guide, representing positive fold change (Figure 4.6C).

4.4.5 Similarities between the two screens

Next, we investigated the overlap of screens in two different cell lines (A549 and BCi NS1.1). For this purpose, the significantly enriched positive selected genes (FDR <0.1) from both screens were overlapped (Figure 4.6D). Surprisingly, only one gene i.e., Deoxyhypusine synthase (*DHPS*) found common between the two screens (Figure 4.6E-F). This divergence is possibly due to the variability in the strength of the CSE effect for long-term exposure and the nature of cell lines used, which might led to the statistical computational variance for the most significant selected genes.

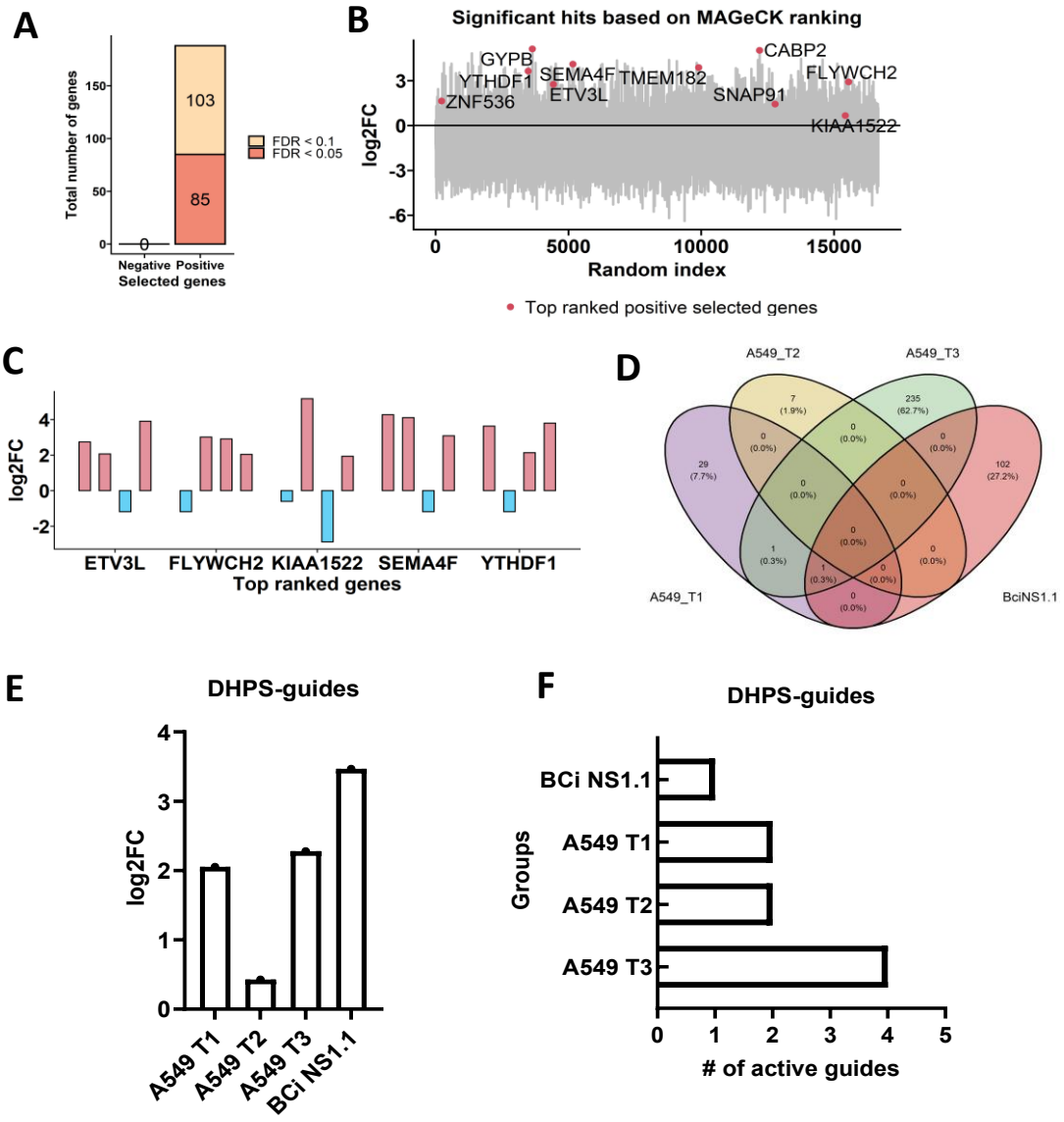

Figure 4.6: Distribution of significantly enriched and depleted sgRNA guides targeting genes in CSE exposed knock-out BCi NS1 cells: ***A)*** *Stacked bar plot representing the total number of guides targeting genes significantly changed. The dark color portion represents the hits with FDR <0.05, light color portion represents the hits with FDR < 0.1.* ***B)*** *Scatterplot showing the top positive or negative selected genes with log2FC on the y-axis and random indexed on the x-axis.* ***C)*** *A bar plot of each representing guides for the top 5 positive selected genes.* ***D)*** *Venn diagram showing the overlap of positive selected genes from T1, T2, and T3 of A549 and BCi NS1.1,* ***E)*** *Bar plot representing log_2FC of the commonly found gene (DHPS) at T1, T2, T3 of A549 and BCi NS1,* ***F)*** *Bar plot represents the number of DHPS guides at T1, T2, T3 of A549 and in BCi NS1.*

4.5 Discussion

Loss-of-functional (LOF) screening has great potential to identify context-dependent factors. Therefore, we performed a genome-wide loss-of-functional CRISPR/Cas9 library screening against cigarette smoke exposure (CSE). Important genes and pathways responsible for CSE resistance were investigated in pooled genome-wide loss-of-function in A549 and BCi NS1.1 cell lines. Our screen identified multiple genetic targets whose depletion can potentially create a significant resistance against CSE exposure and thus can reduce CSE-induced cell death. We hypothesized that knocking out the factors responsible for suppressing the ribosome and protein biogenesis during CSE stress and factors that activate the apoptotic pathway could reduce the sensitization of cells to CSE exposure.

Many studies demonstrated CSE-induced apoptosis and inflammation (383, 384). Cell death in response to CSE exposure is mainly mediated through intrinsic apoptotic pathways, but the mechanism is unclear (383, 384). In line with this, we found guides targeting the genes reported to be essential for apoptosis. This includes apoptosis-linked gene 2 (*ALG-2*), also known as *PDCD6*, a member of Ca^{2+}-binding proteins. A pro-apoptotic protein CDIP1 interreacts with ALG2 in a Ca^{2+}-binding manner whose overexpression induces the caspase-3/7 mediated cell death (385, 386). It has also been reported that mitochondrial ribosomal proteins (MRPs) participate as apoptosis-inducing factors (387). We also found the mitochondrial ribosomal protein L11 (MRPL11) guide significantly enriched in our CSE-exposed group. Limited information is available regarding the function of MRPL11 in response to CSE, rendering it an appealing target for further investigation into its functional role in CSE responses. Proteolytic enzymes such as matrix metalloproteinases have been linked to COPD inflammation and degradation of extra cellular matrix components. Previously MMP-9 and MMP-2 has been

found dysregulated in COPD (388, 389). In our study we identified *MMP-15* as one of the potential hits, suggesting that this may also be important in response to CS-induced damage.

The Sorting nexin 31 (*SNX31*) gene's function is poorly understood, but it is predicted to facilitate phosphatidylinositol binding activity. This likely involves regulating intracellular protein transport, including endocytosis, as well as signaling and sorting within endosomes (390, 391). This gene is implicated in facilitating the endocytic degradation of proteins, providing protection against toxic substances in the urinary tract (391). Knockout studies have demonstrated lysosome accumulation in urothelial umbrella cells, potentially contributing to resistance against toxic substances (392). Lysosomes play a vital role in autophagic degradation and regulation (393), which may influence resistance mechanisms. Increased autophagy has been observed in COPD patients and mice exposed to cigarette smoke (CS), suggesting a role in disease pathology (392-394). Inhibition of autophagic proteins like LC3B has shown promise in preventing CS-induced cell death *in-vitro* (395), highlighting potential therapeutic targets. However, the precise functional role of CS-induced autophagy in different airway cell types remains incompletely understood, suggesting further investigation.

Moreover, during our investigation of constant exposure of CSE on guides survival, we found that the guides targeting genes belong to the pathways involved in apoptosis, ribosomal biogenesis, and RNA metabolism. These guides represent the genes including *MSTO1, PDCD2, SF3A3, ELL,* and *KAT8*. Among them, *PDCD2* accelerates apoptosis (396). Cell death induced through an elongation factor for RNA polymerase II (ELL) has also been reported (397). Besides, another factor, cellular stress response 1 (*CSR1*), mediates cell death essentially through interacting with splicing factor 3A3 (*SF3A3*) (398). In addition, *KAT8* was reported as a cell apoptosis mediator through p53 in response to DNA damage (399-402). One study reported the role of DAPK1 in the association of caspase-3 regulation and apoptosis in the

cigarette-inflamed bronchial epithelium (403). All this supports our findings that our identified enriched guides against CSE exposure belong to the apoptosis pathway and thus can play a critical role in protecting against CSE-induced cell death.

Regarding ribosome and protein biogenesis, upon e-cig smoke solution exposure, these biological pathways were inhibited in normal HBE cells, and the expression of corresponding proteins was reduced (404). This indicates that the genes responsible for inhibiting the ribosome and protein biogenesis can be important therapeutic targets during CSE exposure. Interestingly, during our screening, we observed significantly enriched guides that target the molecules responsible for suppressing the ribosome and protein biogenesis, including RPS25, RPL11, RPS3, and transportin-1 (TNPO1). The processes through which ribosomal proteins (RPs) can control cell death include the MDM2/p53 pathway (405) and via interruption of the NF-κB activity (406). One study has demonstrated the protective role of silencing 40S ribosomal protein, RPS3, in CS-induced acute lung injury in a mouse model (406). Regarding the role of ribosomal proteins, RPL11 and RPS25 are essential for p53 response under ribosomal stress, where p53 is critical for regulating cell apoptosis and proliferation (407, 408). RPL11 has been reported as a positive regulator of p53 (409), whose silencing eliminates the p53 activation that ultimately results in the attenuation of apoptosis (410). Also, RPS25 silencing is reported to attenuate the p53 response under ribosomal stress (411).

To increase our findings' translational value, we exposed the cells multiple times with CSE to enable robust identification of hits. Overall, our screen highlighted some novel potential candidates whose suppression can increase CSE resistance by targeting the stress-induced apoptotic pathways and targets that inhibit the ribosomal and protein biogenesis pathways upon CSE exposure. Furthermore, selecting candidates based on their functional importance and biological role can further increase the possibility of being therapeutic targets against CS-

induced cell death. Our screen also supports previously identified candidates like *PRMT5* and *RPS3* whose knockdown significantly increased the CSE resistance *in-vitro* and attenuated the CSE injury in mice (406, 412). Importantly, we also excluded the cancer-causing genes from our inventory of potential hits as, some of the candidates are tumor suppressor genes including *NF2, TP53, VHL,* and *RAD51.*

The main strength of our study is that we performed the screen longitudinally and collected the guides at different intervals after multiple exposures of rounds. However, there are some limitations. The first and foremost limitation of the CRISPR-Cas9 system is that having off-target effects of Cas9 could lead to false positives and negatives in the final readouts of the functional screens. Also, the lentivirus that transfers the sgRNAs and Cas9 into the cells integrates the sgRNA and Cas9 sequence into the genome, resulting in continuous expression and increasing the probability of false positives. To minimise the effect, we filtered and ranked top hits represented by either 3 or 4 sgRNAs thus, multiple guides showing the same phenotype can significantly reduce the off-target impact. Furthermore, WT Cas9 may not always result in knockout due to the potential for in-frame mutations generated by cellular repair mechanisms to maintain gene function. This can create a mixed population, complicating the screen and data interpretation. Additionally, the identified targets have not been validated individually to confirm the resistance against CSE-induced cell death. Further, the guide-targeted genes identified in *in-vitro* functional screens could be limited to the model cell lines in the study, as not much overlap was seen between two different cell lines either. The lack of overlapping guide-targeted genes between different cell lines used may be due to the differences in cellular context and microenvironment, impacting the interpretation and relevance of screening results. Also, BCi NS1.1 cells lost ~45% of the cell population, perhaps due to long culture time. Despite these limitations, we proposed some interesting potential candidates whose suppression could provide resistance against CSE-induced cell death.

To further investigate the relevance and biological significance of identified candidates, future experiments are essential to determine their potential involvement in CS-induced cell death and COPD. Initial validation should encompass independent assays confirming their functional relevance in response to CSE exposure, including cell viability and apoptosis assessments. Subsequent analysis of gene expression profiles in knockout versus wild-type under CSE exposure will elucidate downstream targets and pathways affected by these candidates. Evaluation of inflammatory response modulation by the candidates should also be conducted through cytokine secretion assays. Crucially, *in-vivo* validation using CS-induced COPD models with transgenic or knockout animal models is imperative, alongside assessments of lung function, histology, and inflammatory markers. Validation in smokers and COPD patient-derived samples will further provide insights into their relevance in physiological contexts and disease models. Efficacy testing of potential therapeutic agents targeting these candidates can be performed using cell-based assays or animal models, facilitating an understanding of the molecular mechanisms underlying CS-induced cell death and COPD pathogenesis, and aiding in the identification of therapeutic targets.

In conclusion, our unbiased genome-wide CRISPR-Cas9 KO functional screens have identified potential genes and pathways whose silencing could induce resistance and enhance survival against CSE stress. The identified genes are context-specific and thus establish the potential relevance of identified pathways in response to CS exposure, providing new research directions to better understand the mechanisms of pathogenesis of CS-induced cell death and COPD.

4.6 Summary of CRISPR library screen outcome

Our unbiased genome-wide knock-out screen revealed several potential candidates, implying that their loss of function may increase protection against cell damage during CSE exposure. We also generated an inventory of potential targets mainly involved in ribosomal/protein biogenesis, unfolded protein responses, cellular response to stress, and apoptosis, suggesting their possible role in conferring resistance against CS-induced damage (Figure 4.7)

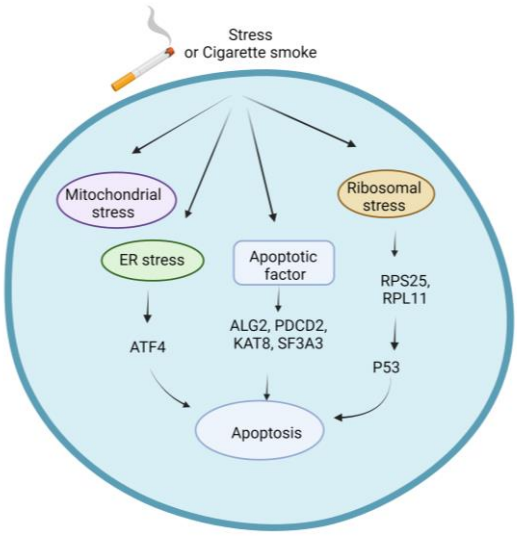

Figure 4.7: Hits from the A549 CRISPR library screen that might protect against CS-induced cell death and COPD. *(Created with BioRender.com)*

Chapter 5: General Discussion

Chronic obstructive pulmonary disease (COPD) is a progressive respiratory disease and a massive cause of death worldwide. It imparts a high socioeconomic burden that continues to climb. Despite many studies highlighting the harmful effects of smoking, it endures being the leading cause of COPD. Other noxious exposures are becoming major causes such as air pollution, e-cigarettes, and vaping and by 2050 >50% of patients with COPD will never have smoked. CS is a good experimental stimulus as research-grade cigarettes are highly quality-controlled. Researchers have made significant efforts to decipher the pathogenic mechanisms that underpin COPD at different molecular levels including genetics, transcriptomics, and proteomics. Despite these advancements and extensive scientific endeavors, the identity and dynamics of the molecular players responsible for the onset and progression of COPD are still incompletely known. Current treatment approaches only provide symptomatic relief for patients but do not halt or reverse the disease. There is an urgent for therapeutic targets for molecular drivers of entire disease pathways in COPD. Indeed, understanding of the COPD underlying mechanisms needs to be more deeply elucidated. Recent studies have also shown the importance of longitudinal over cross-sectional studies because of their ability to characterize changing patterns over time. Additionally, it is frequently suggested that multi-omics research yields deeper more reliable, and more convincing results than single-omics studies. For this reason, we hypothesized that investigating the longitudinal effects of CS that contribute to the development of COPD at the transcriptional level in combination with post-transcriptional changes in CS-induced experimental COPD will identify novel potential therapeutic targets that more globally induce COPD. Furthermore, CRISPR is a state-of-the-art technique based on genome-wide loss-of-function screening against CS that has a great potential to identify in an untargeted fashion the factors that protect against CS-induced lung cell damage. Overall, the entire purpose of conducting our studies was to identify critical

therapeutic targets that have the potential to help abate CS-induced damage in COPD utilizing current molecular techniques and knowledge. Herein, I discuss the critical findings of this body of work along with the limitations of the studies as well as suggest the most important future directions.

5.1 Key findings from this research

The findings highlighted in this thesis are based on *in-vitro* and *in-vivo* studies. For the *in-vivo* studies, our chronic (12 weeks) mouse model of CS-induced experimental COPD mouse model was used, which effectively replicates the primary clinical hallmark features of human COPD. A distinct feature of this model is its clearly defined time points, which enable research into the disease's induction (4 and 6 weeks) and progression (8 and 12 weeks) stages. Using this well-established CS-induced COPD model, we have thoroughly characterized disease progression by exploring the transcriptional profile (Chapter 2) and integrating it with post-transcriptional data (Chapter 3).

In Chapter 2, the investigation of longitudinal transcriptional changes due to constant CS exposure that leads to COPD in CS-induced experimental COPD showed the reduction in expression of novel smoke-protective genes led to COPD progression. These genes involved are oxidative stress-related and belong to xenobiotic response elements (XREs). We discovered that the transcriptional repressor *Arnt2* may be involved in disease progression, as its expression increases when the disease first manifests. We also found that the downregulation of *ARNT2* in human bronchial epithelial cells *in vitro* was protective against the detrimental effect of CS-induced cell death.

In Chapter 3, we investigated the longitudinal miRNA changes and integrated them with the transcriptional changes identified in Chapter 2. The investigation of direct miRNA targets in

CS-induced experimental COPD led to a more comprehensive and deeper understanding of transcriptional and post-transcriptional events that may promote COPD progression. This identified the potential involvement of critical miRNAs (let-7b-5p, miR-151-5p, miR-181a-5p, and miR-181c-5p) along with their target genes (*Arnt2, Bbox1,* and *Ttc25*) in regulating COPD progression. We also observed a significant correlation of the expression of let-7b-5p, miR-181a-5p, and miR-181c-5p with *Arnt2*, miR-151-5p with *Bbox1,* and miR-150-5p with *Ttc25* in a separate COPD model. However, additional cause-and-effect studies can help confirm these findings.

In Chapter 4, we identified genes that are important in providing resistance against CS-induced cell death. For this, in a genome-wide loss-of-function CRISPR screen in human lung cell lines, we knocked out 18,053 protein-coding genes and exposed the knock-out pool of cells to potent CSE-induced cell death. The remaining viable cells were subjected to high-throughput sequencing, to identify the knock-out genes whose targeting made them resistant to CS-induced cell death. Through this CRISPR-based functional genomics study, we generated an inventory of potential targets that might have the potential to confer resistance against CS-induced damage. In this inventory, the genes involved are mainly potential contributors to ribosomal and protein biogenesis and unfolded protein responses and are important regulators of apoptosis. We propose that knockout of these potential targets such as *MMP-15* and *MRPL11* might increase protection against CS damage.

When comparing the outcomes of the mouse model with those of the unbiased CRISPR survival screen, no overlap was noted, with *Bbox1* being the exception. One possible explanation for this inconsistency is the difference in experimental models. The mouse model of COPD was developed over time with continuous CS exposure, while candidates obtained from the screen experienced acute CSE exposure with harsh selection criteria after 80% of cell

death. Additionally, *in-vitro* screens assessed short-term effects of gene perturbations, whereas mouse models involved longitudinal studies during disease development and progression. This temporal disparity in gene function and disease development may contributed to outcome disparities. Furthermore, *in-vitro* screening assays often lack the complexity and physiological relevance of *in-vivo systems*, limiting the understanding of gene function within the whole organism.

Taken together, our studies define potential therapeutic targets that might hinder CS-induced damage that leads to emphysema and COPD.

5.2 Limitations of the current studies

This study has produced substantial new information and potential new therapeutic targets that are molecular drivers of CS-induced damage. Our studies had some limitations that could be improved in our future work. Firstly, the observations from our longitudinal transcriptional analysis coupled with the longitudinal post-transcriptional study are very promising. However, the data was generated using microarrays instead of the latest RNA-Seq-based technology. RNA-Seq sequences the whole transcriptome instead of pre-defined transcripts through hybridization that occurs in microarrays. Secondly, despite the valuable application of the integrative approach that uncovers the potential gene determinants and potential therapeutic targets for COPD, gene and miRNA expression data was generated from separate mouse COPD models potentially masking the findings of significance. Transcriptional and post-transcriptional data generated from the same mouse sample would be more informative and have more potential to truly delineate the flow of information. However, this could all also be a positive in that the identification of targets in different models shows that the findings are consistent. Lastly, regarding the use of an unbiased functional genomic CRISPR screen. Although the functional screen is a mature scientific approach employed across many fields

and has identified potential therapeutic candidate targets against many phenotypes, it is not 100% efficient and accurate due to "off-target" edits. Also, the Cas-9 continues to express due to the sequence integration within the genome. Thus, is essential to validate potential candidates individually.

5.3 Future directions

Future experiments are required to further explore the relevance and biological importance of the identified potential candidates we have discovered. Candidates identified in Chapters 2 and 3 hold promising therapeutic importance and their functional targeting needs to be investigated in the pathogenesis of COPD. Firstly, the identified targets should be studied individually *in vitro* to delineate the relationship of miRNAs with their predicted and validated direct targets with smoke exposure. Following, the identified targets including let-7b-5p, miR-181a-5p, and miR-151c-5p and *Arnt2* should be targeted individually during different phases of smoke exposure in mice, and the downstream impact on the developing features of COPD determined. Such experiments will shed further light on the significance of identified targets to different phases of COPD development and progression.

We plan to further extend this work in more in-depth investigations. There is also a great need to test highly enriched sgRNAs from our screens individually for their ability to confer resistance to the CSE challenge of WT cells. Future gene manipulation studies should focus on targeted deletion using multiple guides per target gene identified in Chapter 4 followed by CSE exposure. This will help to explore new therapeutic targets to prevent CS-induced cell death that leads to emphysema and COPD.

5.4 Conclusion

In conclusion, we identified *ARNT2* as a potential repressor of smoke-protective genes. We also discovered that *Arnt2* expression may act in conjugation with let-7b-5p, miR-181a-5p, and/or miR-181c-5p in the CS-induced COPD model. Additionally, during an unbiased genome-wide knockout screen we identified several potential genes including *MMP15*, whose loss of function may increase protection against CS-induced cell damage. Previous studies have also reported different MMP dysregulation to COPD. Altogether, we have identified new targets and pathways that can open new avenues in developing new treatment regimens against CS-induced cell death in COPD.

References

1. GOLD. Global Initiative for Chronic Obstructive Lung Disease (GOLD), Global strategy for the diagnosis, management and prevention of COPD, Global Initiative for Chronic Obstructive Lung Disease (GOLD). 2018.
2. Barnes PJ, Burney PG, Silverman EK, Celli BR, Vestbo J, Wedzicha JA, et al. Chronic obstructive pulmonary disease. Nat Rev Dis Primers. 2015;1:15076.
3. Yan F, Gao H, Zhao H, Bhatia M, Zeng Y. Roles of airway smooth muscle dysfunction in chronic obstructive pulmonary disease. J Transl Med. 2018;16(1):262.
4. Vestbo J, Hurd SS, Agusti AG, Jones PW, Vogelmeier C, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med. 2013;187(4):347-65.
5. Yang IA, Relan V, Wright CM, Davidson MR, Sriram KB, Savarimuthu Francis SM, et al. Common pathogenic mechanisms and pathways in the development of COPD and lung cancer. Expert opinion on therapeutic targets. 2011;15(4):439-56.
6. López-Campos JL, Tan W, Soriano JB. Global burden of COPD. Respirology. 2016;21(1):14-23.
7. Miravitlles M, Ribera A. Understanding the impact of symptoms on the burden of COPD. Respir Res. 2017;18(1):67.
8. Zinellu E, Zinellu A, Fois AG, Fois SS, Piras B, Carru C, et al. Reliability and Usefulness of Different Biomarkers of Oxidative Stress in Chronic Obstructive Pulmonary Disease. Oxid Med Cell Longev. 2020;2020:4982324.
9. Smith MC, Wrobel JP. Epidemiology and clinical impact of major comorbidities in patients with COPD. Int J Chron Obstruct Pulmon Dis. 2014;9:871-88.
10. Gershon AS, Warner L, Cascagnette P, Victor JC, To T. Lifetime risk of developing chronic obstructive pulmonary disease: a longitudinal population study. Lancet. 2011;378(9795):991-6.
11. Jindal SK. Chronic obstructive pulmonary disease in non-smokers - Is it a different phenotype? Indian J Med Res. 2018;147(4):337-9.
12. Aryal S, Diaz-Guzman E, Mannino DM. Influence of sex on chronic obstructive pulmonary disease risk and treatment outcomes. Int J Chron Obstruct Pulmon Dis. 2014;9:1145-54.
13. Aryal S, Diaz-Guzman E, Mannino DM. COPD and gender differences: an update. Transl Res. 2013;162(4):208-18.
14. Adeloye D, Chua S, Lee C, Basquill C, Papana A, Theodoratou E, et al. Global and regional estimates of COPD prevalence: Systematic review and meta-analysis. J Glob Health. 2015;5(2):020415.
15. Doiron D, de Hoogh K, Probst-Hensch N, Fortier I, Cai Y, De Matteis S, et al. Air pollution, lung function and COPD: results from the population-based UK Biobank study. Eur Respir J. 2019;54(1).
16. Pini L, Paoletti G, Heffler E, Tantucci C, Puggioni F, Asthma, et al. Alpha1-antitrypsin deficiency and asthma. Curr Opin Allergy Clin Immunol. 2021;21(1):46-51.
17. Lokke A, Lange P, Scharling H, Fabricius P, Vestbo J. Developing COPD: a 25 year follow up study of the general population. Thorax. 2006;61(11):935-9.
18. Barnes PJ. Oxidative stress-based therapeutics in COPD. Redox Biol. 2020;33:101544.
19. McGuinness AJ, Sapey E. Oxidative Stress in COPD: Sources, Markers, and Potential Mechanisms. J Clin Med. 2017;6(2).

20. Suissa S, Dell'Aniello S, Ernst P. Long-term natural history of chronic obstructive pulmonary disease: severe exacerbations and mortality. Thorax. 2012;67(11):957-63.
21. Chung K. Cytokines in chronic obstructive pulmonary disease. European Respiratory Journal. 2001;18(34 suppl):50s-9s.
22. Hartl D, Tirouvanziam R, Laval J, Greene CM, Habiel D, Sharma L, et al. Innate immunity of the lung: from basic mechanisms to translational medicine. Journal of innate immunity. 2018;10(5-6):487-501.
23. Barnes PJ. The cytokine network in asthma and chronic obstructive pulmonary disease. The Journal of clinical investigation. 2008;118(11):3546-56.
24. Chatfield SM, Thieblemont N, Witko-Sarsat V. Expanding neutrophil horizons: new concepts in inflammation. Journal of innate immunity. 2018;10(5-6):422-31.
25. Tanner L, Single AB. Animal models reflecting chronic obstructive pulmonary disease and related respiratory disorders: translating pre-clinical data into clinical relevance. Journal of innate immunity. 2020;12(3):203-25.
26. Qiu SL, Zhong XN. [Immunologic aspects of chronic obstructive pulmonary disease and emphysema]. Zhonghua Jie He He Hu Xi Za Zhi. 2010;33(4):298-300.
27. Alfahad AJ, Alzaydi MM, Aldossary AM, Alshehri AA, Almughem FA, Zaidan NM, et al. Current views in chronic obstructive pulmonary disease pathogenesis and management. Saudi Pharm J. 2021;29(12):1361-73.
28. Jha P, Peto R. Global effects of smoking, of quitting, and of taxing tobacco. N Engl J Med. 2014;370(1):60-8.
29. Jamison DT, Summers LH, Alleyne G, Arrow KJ, Berkley S, Binagwaho A, et al. Global health 2035: a world converging within a generation. Lancet. 2013;382(9908):1898-955.
30. Halpin DM, Decramer M, Celli B, Kesten S, Liu D, Tashkin DP. Exacerbation frequency and course of COPD. Int J Chron Obstruct Pulmon Dis. 2012;7:653-61.
31. Wu J, Sin DD. Improved patient outcome with smoking cessation: when is it too late? Int J Chron Obstruct Pulmon Dis. 2011;6:259-67.
32. Maddocks M, Kon SS, Canavan JL, Jones SE, Nolan CM, Labey A, et al. Physical frailty and pulmonary rehabilitation in COPD: a prospective cohort study. Thorax. 2016;71(11):988-95.
33. Kopsaftis Z, Wood-Baker R, Poole P. Influenza vaccine for chronic obstructive pulmonary disease (COPD). Cochrane Database Syst Rev. 2018;6(6):CD002733.
34. Walters JA, Tang JNQ, Poole P, Wood-Baker R. Pneumococcal vaccines for preventing pneumonia in chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews. 2017(1).
35. Oba Y, Keeney E, Ghatehorde N, Dias S. Dual combination therapy versus long-acting bronchodilators alone for chronic obstructive pulmonary disease (COPD): a systematic review and network meta-analysis. Cochrane Database Syst Rev. 2018;12(12):CD012620.
36. Yang M, Du Y, Chen H, Jiang D, Xu Z. Inhaled corticosteroids and risk of pneumonia in patients with chronic obstructive pulmonary disease: a meta-analysis of randomized controlled trials. International immunopharmacology. 2019;77:105950.
37. Harries TH, Rowland V, Corrigan CJ, Marshall IJ, McDonnell L, Prasad V, et al. Blood eosinophil count, a marker of inhaled corticosteroid effectiveness in preventing COPD exacerbations in post-hoc RCT and observational studies: systematic review and meta-analysis. Respir Res. 2020;21(1):3.
38. Houben-Wilke S, Augustin IM, Vercoulen JH, van Ranst D, Bij de Vaate E, Wempe JB, et al. COPD stands for complex obstructive pulmonary disease. Eur Respir Rev. 2018;27(148).

39. "Continuous fat-free mass decline in COPD: fact or fiction?" Erica P.A. Rutten, Martijn A. Spruit, Merry-Lynn N. McDonald, Stephen Rennard, Alvar Agusti, Bartolome Celli, Bruce E. Miller, Courtney Crim, Peter M.A. Calverley, Corrine Hanson, William MacNee, Frits M.E. Franssen, Lowie Vanfleteren and Emiel F.M. Wouters on behalf of the ECLIPSE (Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points) Investigators. Eur Respir J 2015; 46: 1496-1498. Eur Respir J. 2016;48(5):1533.
40. Duffy S, Weir M, Criner GJ. The complex challenge of chronic obstructive pulmonary disease. Lancet Respir Med. 2015;3(12):917-9.
41. Divo MJ, Casanova C, Marin JM, Pinto-Plata VM, de-Torres JP, Zulueta JJ, et al. COPD comorbidities network. Eur Respir J. 2015;46(3):640-50.
42. Chung KF. Cytokines in chronic obstructive pulmonary disease. Eur Respir J Suppl. 2001;34:50s-9s.
43. Hartl D, Tirouvanziam R, Laval J, Greene CM, Habiel D, Sharma L, et al. Innate Immunity of the Lung: From Basic Mechanisms to Translational Medicine. J Innate Immun. 2018;10(5-6):487-501.
44. Barnes PJ. The cytokine network in asthma and chronic obstructive pulmonary disease. J Clin Invest. 2008;118(11):3546-56.
45. Chatfield SM, Thieblemont N, Witko-Sarsat V. Expanding Neutrophil Horizons: New Concepts in Inflammation. J Innate Immun. 2018;10(5-6):422-31.
46. Zeiger RS, Tran TN, Butler RK, Schatz M, Li Q, Khatry DB, et al. Relationship of Blood Eosinophil Count to Exacerbations in Chronic Obstructive Pulmonary Disease. J Allergy Clin Immunol Pract. 2018;6(3):944-54 e5.
47. Verhamme FM, Bracke KR, Joos GF, Brusselle GG. Transforming growth factor-beta superfamily in obstructive lung diseases. more suspects than TGF-beta alone. Am J Respir Cell Mol Biol. 2015;52(6):653-62.
48. Jang JH, Chand HS, Bruse S, Doyle-Eisele M, Royer C, McDonald J, et al. Connective Tissue Growth Factor Promotes Pulmonary Epithelial Cell Senescence and Is Associated with COPD Severity. COPD. 2017;14(2):228-37.
49. Matera MG, Calzetta L, Rogliani P, Cesario A, Cazzola M. New treatments for COPD in the elderly. Curr Pharm Des. 2014;20(38):5968-82.
50. Barnes PJ, Shapiro SD, Pauwels RA. Chronic obstructive pulmonary disease: molecular and cellular mechanisms. Eur Respir J. 2003;22(4):672-88.
51. Russell RE, Thorley A, Culpitt SV, Dodd S, Donnelly LE, Demattos C, et al. Alveolar macrophage-mediated elastolysis: roles of matrix metalloproteinases, cysteine, and serine proteases. Am J Physiol Lung Cell Mol Physiol. 2002;283(4):L867-73.
52. Malhotra D, Thimmulappa R, Navas-Acien A, Sandford A, Elliott M, Singh A, et al. Decline in NRF2-regulated antioxidants in chronic obstructive pulmonary disease lungs due to loss of its positive regulator, DJ-1. American journal of respiratory and critical care medicine. 2008;178(6):592-604.
53. Hwang J-w, Rajendrasozhan S, Yao H, Chung S, Sundar IK, Huyck HL, et al. FOXO3 deficiency leads to increased susceptibility to cigarette smoke-induced inflammation, airspace enlargement, and chronic obstructive pulmonary disease. The Journal of Immunology. 2011;187(2):987-98.
54. Lee J, Jang J, Park S-M, Yang S-R. An update on the role of Nrf2 in respiratory disease: molecular mechanisms and therapeutic approaches. International Journal of Molecular Sciences. 2021;22(16):8406.
55. Barnes PJ. Oxidative stress-based therapeutics in COPD. Redox biology. 2020;33:101544.
56. Zhang X-Y, Li W, Zhang J-R, Li C-Y, Zhang J, Lv X-J. Roles of sirtuin family members in chronic obstructive pulmonary disease. Respiratory Research. 2022;23(1):1-10.

57. Turino G. Emphysema in COPD: consequences and causes. Thorax. 2006;61(12):1031.
58. Gharib SA, Manicone AM, Parks WC. Matrix metalloproteinases in emphysema. Matrix Biology. 2018;73:34-51.
59. Heinz A. Elastases and elastokines: elastin degradation and its significance in health and disease. Critical reviews in biochemistry and molecular biology. 2020;55(3):252-73.
60. Houghton AM, Quintero PA, Perkins DL, Kobayashi DK, Kelley DG, Marconcini LA, et al. Elastin fragments drive disease progression in a murine model of emphysema. The Journal of clinical investigation. 2006;116(3):753-9.
61. Jones RL, Noble PB, Elliot JG, James AL. Airway remodelling in COPD: It's not asthma! Respirology. 2016;21(8):1347-56.
62. Theocharis AD, Skandalis SS, Gialeli C, Karamanos NK. Extracellular matrix structure. Advanced drug delivery reviews. 2016;97:4-27.
63. Strange C. Alpha-1 antitrypsin deficiency associated COPD. Clinics in Chest Medicine. 2020;41(3):339-45.
64. Barre-Sinoussi F, Montagutelli X. Animal models are essential to biological research: issues and perspectives. Future Sci OA. 2015;1(4):FSO63.
65. Reczynska K, Tharkar P, Kim SY, Wang Y, Pamula E, Chan HK, et al. Animal models of smoke inhalation injury and related acute and chronic lung diseases. Adv Drug Deliv Rev. 2018;123:107-34.
66. Groneberg DA, Chung KF. Models of chronic obstructive pulmonary disease. Respiratory research. 2004;5(1):1-16.
67. Eltom S, Stevenson C, Birrell MA. Cigarette smoke exposure as a model of inflammation associated with COPD. Current Protocols in Pharmacology. 2013;60(1):5.64. 1-5.. 18.
68. Mortaz E, Adcock IA. Limitation of COPD studies in animal modeling. Tanaffos. 2012;11(3):7.
69. Ghorani V, Boskabady MH, Khazdair MR, Kianmeher M. Experimental animal models for COPD: a methodological review. Tobacco induced diseases. 2017;15:1-13.
70. Li Y, Li S-Y, Li J-S, Deng L, Tian Y-G, Jiang S-L, et al. A rat model for stable chronic obstructive pulmonary disease induced by cigarette smoke inhalation and repetitive bacterial infection. Biological and Pharmaceutical Bulletin. 2012;35(10):1752-60.
71. Brusselle G, Bracke K, Maes T, D'hulst A, Moerloose K, Joos G, et al. Murine models of COPD. Pulmonary pharmacology & therapeutics. 2006;19(3):155-65.
72. Bonfield TL. In vivo models of lung disease. Lung diseases-selected state of the art Reviews, InTech. 2012:407-28.
73. O'byrne P, Postma D. The many faces of airway inflammation: asthma and chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine. 1999;159(supplement_2):S1-S63.
74. Pera T. Inflammation and remodelling in experimental models of COPD-Mechanisms and therapeutic perspectives. 2011.
75. Churg A, Tai H, Coulthard T, Wang R, Wright JL. Cigarette smoke drives small airway remodeling by induction of growth factors in the airway wall. American journal of respiratory and critical care medicine. 2006;174(12):1327-34.
76. Wright JL, Cosio M, Churg A. Animal models of chronic obstructive pulmonary disease. american journal of physiology-lung cellular and molecular physiology. 2008;295(1):L1-L15.
77. Wright J, Postma D, Kerstjens H, Timens W, Whittaker P, Churg A. Airway remodeling in the smoke exposed guinea pig model. Inhalation toxicology. 2007;19(11):915-23.

78. Bracke K, D'hulst A, Maes T, Demedts I, Moerloose K, Kuziel W, et al. Cigarette smoke-induced pulmonary inflammation, but not airway remodelling, is attenuated in chemokine receptor 5-deficient mice. Clinical & Experimental Allergy. 2007;37(10):1467-79.
79. Churg A, Cosio M, Wright JL. Mechanisms of cigarette smoke-induced COPD: insights from animal models. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2008;294(4):L612-L31.
80. Wright JL, Churg A. Animal models of cigarette smoke-induced COPD. Chest. 2002;122(6):301S-6S.
81. Wright JL, Churg A. Animal models of cigarette smoke-induced chronic obstructive pulmonary disease. Expert review of respiratory medicine. 2010;4(6):723-34.
82. Canning BJ. Modeling asthma and COPD in animals: a pointless exercise? Current opinion in pharmacology. 2003;3(3):244-50.
83. Canning BJ, Chou Y. Using guinea pigs in studies relevant to asthma and COPD. Pulmonary pharmacology & therapeutics. 2008;21(5):702-20.
84. Bourbon JR, Boucherat O, Boczkowski J, Crestani B, Delacourt C. Bronchopulmonary dysplasia and emphysema: in search of common therapeutic targets. Trends in molecular medicine. 2009;15(4):169-79.
85. Ressmeyer A, Larsson A, Vollmer E, Dahlén S-E, Uhlig S, Martin C. Characterisation of guinea pig precision-cut lung slices: comparison with human tissues. European Respiratory Journal. 2006;28(3):603-11.
86. Plopper CG, Hyde DM. The non-human primate as a model for studying COPD and asthma. Pulmonary pharmacology & therapeutics. 2008;21(5):755-66.
87. Chapman RW. Canine models of asthma and COPD. Pulmonary pharmacology & therapeutics. 2008;21(5):731-42.
88. John G, Kohse K, Orasche J, Reda A, Schnelle-Kreis J, Zimmermann R, et al. The composition of cigarette smoke determines inflammatory cell recruitment to the lung in COPD mouse models. Clinical science. 2014;126(3):207-21.
89. Mallia P, Johnston SL. Mechanisms and experimental models of chronic obstructive pulmonary disease exacerbations. Proceedings of the American Thoracic Society. 2005;2(4):361-6.
90. Leberl M, Kratzer A, Taraseviciene-Stewart L. Tobacco smoke induced COPD/emphysema in the animal model—are we all on the same page? Frontiers in physiology. 2013;4:91.
91. Peters LL, Robledo RF, Bult CJ, Churchill GA, Paigen BJ, Svenson KL. The mouse as a model for human biology: a resource guide for complex trait analysis. Nat Rev Genet. 2007;8(1):58-69.
92. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, et al. A comparative encyclopedia of DNA elements in the mouse genome. Nature. 2014;515(7527):355-64.
93. Mahadeva R, Shapiro SD. Chronic obstructive pulmonary disease * 3: Experimental animal models of pulmonary emphysema. Thorax. 2002;57(10):908-14.
94. Hardaker L, Bahra P, de Billy BC, Freeman M, Kupfer N, Wyss D, et al. The ion channel transient receptor potential melastatin-2 does not play a role in inflammatory mouse models of chronic obstructive pulmonary diseases. Respiratory research. 2012;13(1):1-10.
95. Motz GT, Eppert BL, Wesselkamper SC, Flury JL, Borchers MT. Chronic cigarette smoke exposure generates pathogenic T cells capable of driving COPD-like disease in Rag2−/− mice. American journal of respiratory and critical care medicine. 2010;181(11):1223-33.
96. Motz GT, Eppert BL, Wortham BW, Amos-Kroohs RM, Flury JL, Wesselkamper SC, et al. Chronic cigarette smoke exposure primes NK cell activation in a mouse model of chronic obstructive pulmonary disease. The Journal of Immunology. 2010;184(8):4460-9.

97. Cremona TP, Tschanz SA, von Garnier C, Benarafa C. SerpinB1 deficiency is not associated with increased susceptibility to pulmonary emphysema in mice. American journal of physiology-lung cellular and molecular physiology. 2013;305(12):L981-L9.
98. Jobse BN, Rhem RG, Wang IQ, Counter WB, Stämpfli MR, Labiris NR. Detection of lung dysfunction using ventilation and perfusion SPECT in a mouse model of chronic cigarette smoke exposure. Journal of Nuclear Medicine. 2013;54(4):616-23.
99. Barrett EG, Wilder JA, March TH, Espindola T, Bice DE. Cigarette smoke-induced airway hyperresponsiveness is not dependent on elevated immunoglobulin and eosinophilic inflammation in a mouse model of allergic airway disease. American journal of respiratory and critical care medicine. 2002;165(10):1410-8.
100. Tanabe N, Hoshino Y, Marumo S, Kiyokawa H, Sato S, Kinose D, et al. Thioredoxin-1 protects against neutrophilic inflammation and emphysema progression in a mouse model of chronic obstructive pulmonary disease exacerbation. PloS one. 2013;8(11):e79016.
101. Zhou S, Wright JL, Liu J, Sin DD, Churg A. Aging does not enhance experimental cigarette smoke-induced COPD in the mouse. PloS one. 2013;8(8):e71410.
102. Beckett EL, Stevens RL, Jarnicki AG, Kim RY, Hanish I, Hansbro NG, et al. A new short-term mouse model of chronic obstructive pulmonary disease identifies a role for mast cell tryptase in pathogenesis. Journal of Allergy and Clinical Immunology. 2013;131(3):752-62. e7.
103. Hansbro P, Beckett E, Stevens R, Jarnicki A, Wark P, Foster P. A short-term model of COPD identifies a role for mast cell tryptase. Eur Respiratory Soc; 2013.
104. Stevenson CS, Docx C, Webster R, Battram C, Hynx D, Giddings J, et al. Comprehensive gene expression profiling of rat lung reveals distinct acute and chronic responses to cigarette smoke inhalation. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2007;293(5):L1183-L93.
105. Haw T, Starkey M, Nair P, Pavlidis S, Liu G, Nguyen D, et al. A pathogenic role for tumor necrosis factor-related apoptosis-inducing ligand in chronic obstructive pulmonary disease. Mucosal immunology. 2016;9(4):859-72.
106. Fricker M, Goggins BJ, Mateer S, Jones B, Kim RY, Gellatly SL, et al. Chronic cigarette smoke exposure induces systemic hypoxia that drives intestinal dysfunction. JCI insight. 2018;3(3).
107. Liu G, Cooley MA, Jarnicki AG, Hsu AC, Nair PM, Haw TJ, et al. Fibulin-1 regulates the pathogenesis of tissue remodeling in respiratory diseases. JCI insight. 2016;1(9).
108. Jones B, Donovan C, Liu G, Gomez HM, Chimankar V, Harrison CL, et al. Animal models of COPD: W hat do they tell us? Respirology. 2017;22(1):21-32.
109. Van Eeckhoutte HP, Donovan C, Kim R, Khan H, Jayaraman R, Dondelinger Y, et al. Inhibiting RIPK1 kinase activity is protective in experimental models of COPD. Eur Respiratory Soc; 2022.
110. Kapellos T, Bassler K, Fujii W, Pecht T, Bonaguro L, Galvao I, et al. Inflammatory blood neutrophils in COPD stem from activated bone marrow progenitors. Eur Respiratory Soc; 2022.
111. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57-63.
112. Hager J. Making and using spotted DNA microarrays in an academic core laboratory. Methods Enzymol. 2006;410:135-68.
113. Bumgarner R. Overview of DNA microarrays: types, applications, and their future. Curr Protoc Mol Biol. 2013;Chapter 22:Unit 22 1.
114. Afzal M, Manzoor I, Kuipers OP. A Fast and Reliable Pipeline for Bacterial Transcriptome Analysis Case study: Serine-dependent Gene Regulation in Streptococcus pneumoniae. J Vis Exp. 2015(98).

115. Campbell JD, Spira A, Lenburg ME. Applying gene expression microarrays to pulmonary disease. Respirology. 2011;16(3):407-18.
116. Koltai H, Weingarten-Baror C. Specificity of DNA microarray hybridization: characterization, effectors and approaches for data correction. Nucleic Acids Res. 2008;36(7):2395-405.
117. Kim WJ, Lim JH, Lee JS, Lee SD, Kim JH, Oh YM. Comprehensive Analysis of Transcriptome Sequencing Data in the Lung Tissues of COPD Subjects. Int J Genomics. 2015;2015:206937.
118. Pineau F, Shumyatsky G, Owuor N, Nalamala N, Kotnala S, Bolla S, et al. Microarray analysis identifies defects in regenerative and immune response pathways in COPD airway basal cells. ERJ Open Res. 2020;6(4).
119. Bhattacharya S, Srisuma S, Demeo DL, Shapiro SD, Bueno R, Silverman EK, et al. Molecular biomarkers for quantitative and discrete COPD phenotypes. Am J Respir Cell Mol Biol. 2009;40(3):359-67.
120. Spira A, Beane J, Pinto-Plata V, Kadar A, Liu G, Shah V, et al. Gene expression profiling of human lung tissue from smokers with severe emphysema. Am J Respir Cell Mol Biol. 2004;31(6):601-10.
121. Wang IM, Stepaniants S, Boie Y, Mortimer JR, Kennedy B, Elliott M, et al. Gene expression profiling in patients with chronic obstructive pulmonary disease and lung cancer. Am J Respir Crit Care Med. 2008;177(4):402-11.
122. Ning W, Li CJ, Kaminski N, Feghali-Bostwick CA, Alber SM, Di YP, et al. Comprehensive gene expression profiles reveal pathways related to the pathogenesis of chronic obstructive pulmonary disease. Proc Natl Acad Sci U S A. 2004;101(41):14895-900.
123. Zhang W, Yan SD, Zhu A, Zou YS, Williams M, Godman GC, et al. Expression of Egr-1 in late stage emphysema. Am J Pathol. 2000;157(4):1311-20.
124. Takizawa H, Tanaka M, Takami K, Ohtoshi T, Ito K, Satoh M, et al. Increased expression of transforming growth factor-beta1 in small airway epithelium from tobacco smokers and patients with chronic obstructive pulmonary disease (COPD). Am J Respir Crit Care Med. 2001;163(6):1476-83.
125. Zhu YK, Liu X, Ertl RF, Kohyama T, Wen FQ, Wang H, et al. Retinoic acid attenuates cytokine-driven fibroblast degradation of extracellular matrix in three-dimensional culture. Am J Respir Cell Mol Biol. 2001;25(5):620-7.
126. Rogers LRK, Verlinde M, Mias GI. Gene expression microarray public dataset reanalysis in chronic obstructive pulmonary disease. PLoS One. 2019;14(11):e0224750.
127. Winter J, Jung S, Keller S, Gregory RI, Diederichs S. Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol. 2009;11(3):228-34.
128. Baker M. MicroRNA profiling: separating signal from noise. Nat Methods. 2010;7(9):687-92.
129. Schembri F, Sridhar S, Perdomo C, Gustafson AM, Zhang X, Ergun A, et al. MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. Proc Natl Acad Sci U S A. 2009;106(7):2319-24.
130. Pandit KV, Corcoran D, Yousef H, Yarlagadda M, Tzouvelekis A, Gibson KF, et al. Inhibition and role of let-7d in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2010;182(2):220-9.
131. Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell. 2006;9(3):189-98.
132. Ezzie ME, Crawford M, Cho JH, Orellana R, Zhang S, Gelinas R, et al. Gene expression networks in COPD: microRNA and mRNA regulation. Thorax. 2012;67(2):122-31.

133. Kara M, Kirkil G, Kalemci S. Differential Expression of MicroRNAs in Chronic Obstructive Pulmonary Disease. Adv Clin Exp Med. 2016;25(1):21-6.
134. Soeda S, Ohyashiki JH, Ohtsuki K, Umezu T, Setoguchi Y, Ohyashiki K. Clinical relevance of plasma miR-106b levels in patients with chronic obstructive pulmonary disease. Int J Mol Med. 2013;31(3):533-9.
135. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 2015;4:1521.
136. Gan Q, Chepelev I, Wei G, Tarayrah L, Cui K, Zhao K, et al. Dynamic regulation of alternative splicing and chromatin structure in Drosophila gonads revealed by RNA-seq. Cell Res. 2010;20(7):763-83.
137. Vu TN, Deng W, Trac QT, Calza S, Hwang W, Pawitan Y. A fast detection of fusion genes from paired-end RNA-seq data. BMC Genomics. 2018;19(1):786.
138. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, et al. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. 2006;34(Database issue):D354-7.
139. Morrow JD, Chase RP, Parker MM, Glass K, Seo M, Divo M, et al. RNA-sequencing across three matched tissues reveals shared and tissue-specific gene expression and pathway signatures of COPD. Respir Res. 2019;20(1):65.
140. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. Genome Biol. 2016;17:13.
141. Campbell JD, McDonough JE, Zeskind JE, Hackett TL, Pechkovsky DV, Brandsma CA, et al. A gene expression signature of emphysema-related lung destruction and its reversal by the tripeptide GHK. Genome Med. 2012;4(8):67.
142. Faner R, Cruz T, Casserras T, Lopez-Giraldo A, Noell G, Coca I, et al. Network Analysis of Lung Transcriptomics Reveals a Distinct B-Cell Signature in Emphysema. Am J Respir Crit Care Med. 2016;193(11):1242-53.
143. Morrow JD, Zhou X, Lao T, Jiang Z, DeMeo DL, Cho MH, et al. Functional interactors of three genome-wide association study genes are differentially expressed in severe chronic obstructive pulmonary disease lung tissue. Sci Rep. 2017;7:44232.
144. Morrow JD, Qiu W, Chhabra D, Rennard SI, Belloni P, Belousov A, et al. Identifying a gene expression signature of frequent COPD exacerbations in peripheral blood using network methods. BMC Med Genomics. 2015;8:1.
145. Obeidat M, Nie Y, Chen V, Shannon CP, Andiappan AK, Lee B, et al. Network-based analysis reveals novel gene signatures in peripheral blood of patients with chronic obstructive pulmonary disease. Respir Res. 2017;18(1):72.
146. Reinhold D, Morrow JD, Jacobson S, Hu J, Ringel B, Seibold MA, et al. Meta-analysis of peripheral blood gene expression modules for COPD phenotypes. PLoS One. 2017;12(10):e0185682.
147. Xue J, Schmidt SV, Sander J, Draffehn A, Krebs W, Quester I, et al. Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. Immunity. 2014;40(2):274-88.
148. Ghosh AJ, Hobbs BD, Yun JH, Saferali A, Moll M, Xu Z, et al. Lung tissue shows divergent gene expression between chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis. Respir Res. 2022;23(1):97.
149. Du M, Garcia JG, Christie JD, Xin J, Cai G, Meyer NJ, et al. Integrative omics provide biological and clinical insights into acute respiratory distress syndrome. Intensive care medicine. 2021;47(7):761-71.
150. Kan M, Shumyatcher M, Himes BE. Using omics approaches to understand pulmonary diseases. Respiratory research. 2017;18(1):1-20.

151. Ruan P, Todd JL, Zhao H, Liu Y, Vinisko R, Soellner JF, et al. Integrative multi-omics analysis reveals novel idiopathic pulmonary fibrosis endotypes associated with disease progression. Respiratory Research. 2023;24(1):1-12.
152. Li CX, Gao J, Sköld CM, Wheelock ÅM. miRNA–mRNA–protein dysregulated network in COPD in women. Frontiers in Genetics. 2022;13:1010048.
153. Tasena H, Faiz A, Timens W, Noordhoek J, Hylkema MN, Gosens R, et al. microRNA–mRNA regulatory networks underlying chronic mucus hypersecretion in COPD. European Respiratory Journal. 2018;52(3).
154. Konigsberg IR, Borie R, Walts AD, Cardwell J, Rojas M, Metzger F, et al. Molecular signatures of idiopathic pulmonary fibrosis. American Journal of Respiratory Cell and Molecular Biology. 2021;65(4):430-41.
155. Zheng P, Sun S, Wang J, Cheng ZJ, Lei KC, Xue M, et al. Integrative omics analysis identifies biomarkers of idiopathic pulmonary fibrosis. Cellular and Molecular Life Sciences. 2022;79(1):66.
156. Sheppard D. Aspen Lung Conference 2010: systems biology of lung diseases--progress in the omics era. Proc Am Thorac Soc. 2011;8(2):199-202.
157. Kaminsky DA, Irvin CG, Sterk PJ. Complex systems in pulmonary medicine: a systems biology approach to lung disease. J Appl Physiol (1985). 2011;110(6):1716-22.
158. Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. Cell. 2008;134(5):714-7.
159. Li CX, Gao J, Skold CM, Wheelock AM. miRNA-mRNA-protein dysregulated network in COPD in women. Front Genet. 2022;13:1010048.
160. Zhuang Y, Hobbs BD, Hersh CP, Kechris K. Identifying miRNA-mRNA Networks Associated With COPD Phenotypes. Front Genet. 2021;12:748356.
161. O'Leary L, Sevinç K, Papazoglou IM, Tildy B, Detillieux K, Halayko AJ, et al. Airway smooth muscle inflammation is regulated by micro RNA-145 in COPD. Febs Letters. 2016;590(9):1324-34.
162. Kusko RL, Brothers JF, 2nd, Tedrow J, Pandit K, Huleihel L, Perdomo C, et al. Integrated Genomics Reveals Convergent Transcriptomic Networks Underlying Chronic Obstructive Pulmonary Disease and Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2016;194(8):948-60.
163. Du Y, Ding Y, Chen X, Mei Z, Ding H, Wu Y, et al. MicroRNA-181c inhibits cigarette smoke–induced chronic obstructive pulmonary disease by regulating CCN1 expression. Respiratory Research. 2017;18(1):1-8.
164. Jin Y, Kim HP, Ifedigbo E, Lau LF, Choi AM. Cyr61 protects against hyperoxia-induced cell death via Akt pathway in pulmonary epithelial cells. American journal of respiratory cell and molecular biology. 2005;33(3):297-302.
165. Moon H-G, Zheng Y, An CH, Kim Y-K, Jin Y. CCN1 secretion induced by cigarette smoking extracts augments IL-8 release from bronchial epithelial cells. PloS one. 2013;8(7):e68199.
166. Paschalaki KE, Zampetaki A, Baker JR, Birrell MA, Starke RD, Belvisi MG, et al. Downregulation of MicroRNA-126 Augments DNA Damage Response in Cigarette Smokers and Patients with Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2018;197(5):665-8.
167. Ishino Y, Krupovic M, Forterre P. History of CRISPR-Cas from encounter with a mysterious repeated sequence to genome editing technology. Journal of bacteriology. 2018;200(7):10.1128/jb. 00580-17.
168. Mojica FJ, Díez-Villaseñor Cs, García-Martínez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. Journal of molecular evolution. 2005;60:174-82.

169. Mojica FJ, Rodriguez-Valera F. The discovery of CRISPR in archaea and bacteria. The FEBS journal. 2016;283(17):3162-9.
170. Jansen R, Embden JDv, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. Molecular microbiology. 2002;43(6):1565-75.
171. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315(5819):1709-12.
172. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science. 2008;321(5891):960-4.
173. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339(6121):819-23.
174. McGinn J, Marraffini LA. Molecular mechanisms of CRISPR-Cas spacer acquisition. Nat Rev Microbiol. 2019;17(1):7-12.
175. Jiang F, Doudna JA. CRISPR-Cas9 Structures and Mechanisms. Annu Rev Biophys. 2017;46:505-29.
176. Dong C, Fontana J, Patel A, Carothers JM, Zalatan JG. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. Nat Commun. 2018;9(1):2489.
177. Komor AC, Badran AH, Liu DR. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell. 2017;169(3):559.
178. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. Science. 2016;351(6268):84-8.
179. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. science. 2012;337(6096):816-21.
180. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proceedings of the National Academy of Sciences. 2012;109(39):E2579-E86.
181. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339(6121):819-23.
182. JE CG. RNA-guided human genome engineering via Cas9. Science 339: 823826Mao Y, Zhang H, Xu N, Zhang B, Gao F, Zhu JK (2013) Application of the CRISPR-Cas system for efficient genome engineering in plants. Mol Plant. 2013;6:20082011Miao.
183. Nishiga M, Qi LS, Wu JC. CRISPRi/a screening with human iPSCs. Pluripotent Stem-Cell Derived Cardiomyocytes: Springer; 2021. p. 261-81.
184. Joung J, Konermann S, Gootenberg JS, Abudayyeh OO, Platt RJ, Brigham MD, et al. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. Nature protocols. 2017;12(4):828-63.
185. Joung J, Engreitz JM, Konermann S, Abudayyeh OO, Verdine VK, Aguet F, et al. Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. Nature. 2017;548(7667):343-6.
186. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science. 2014;343(6166):84-7.
187. Nishiga M, Liu C, Qi LS, Wu JC. The use of new CRISPR tools in cardiovascular research and medicine. Nature Reviews Cardiology. 2022:1-17.
188. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. Nature methods. 2014;11(8):783-4.
189. Xu CL, Ruan MZ, Mahajan VB, Tsang SH. Viral delivery systems for CRISPR. Viruses. 2019;11(1):28.
190. Lino CA, Harper JC, Carney JP, Timlin JA. Delivering CRISPR: a review of the challenges and approaches. Drug delivery. 2018;25(1):1234-57.

191. Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system. Science. 2014;343(6166):80-4.
192. Yaniz-Galende E, Hajjar R. Stem cell and gene therapy for cardiac regeneration. Cardiac Regeneration and Repair: Elsevier; 2014. p. 347-79.
193. Pauwels K, Gijsbers R, Toelen J, Schambach A, Willard-Gallo K, Verheust C, et al. State-of-the-art lentiviral vectors for research use: risk assessment and biosafety recommendations. Current gene therapy. 2009;9(6):459-74.
194. Tiscornia G, Singer O, Verma IM. Production and purification of lentiviral vectors. Nature protocols. 2006;1(1):241-5.
195. Agrotis A, Ketteler R. A new age in functional genomics using CRISPR/Cas9 in arrayed library screening. Frontiers in genetics. 2015;6:300.
196. Yeung AT, Choi YH, Lee AH, Hale C, Ponstingl H, Pickard D, et al. A genome-wide knockout screen in human macrophages identified host factors modulating Salmonella infection. MBio. 2019;10(5):10.1128/mbio. 02169-19.
197. Yau EH, Rana TM. Next-generation sequencing of genome-wide CRISPR screens. Next Generation Sequencing: Methods and Protocols. 2018:203-16.
198. Slesarev A, Viswanathan L, Tang Y, Borgschulte T, Achtien K, Razafsky D, et al. CRISPR/Cas9 targeted CAPTURE of mammalian genomic regions for characterization by NGS. Scientific reports. 2019;9(1):3587.
199. Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome biology. 2014;15(12):1-12.
200. Li W, Köster J, Xu H, Chen C-H, Xiao T, Liu JS, et al. Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. Genome biology. 2015;16:1-13.
201. Miles LA, Garippa RJ, Poirier JT. Design, execution, and analysis of pooled in vitro CRISPR/Cas9 screens. The FEBS journal. 2016;283(17):3170-80.
202. Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, et al. Identification and characterization of essential genes in the human genome. Science. 2015;350(6264):1096-101.
203. Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. Cell. 2015;163(6):1515-26.
204. Tzelepis K, Koike-Yusa H, De Braekeleer E, Li Y, Metzakopian E, Dovey OM, et al. A CRISPR dropout screen identifies genetic vulnerabilities and therapeutic targets in acute myeloid leukemia. Cell reports. 2016;17(4):1193-205.
205. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a cancer dependency map. Cell. 2017;170(3):564-76. e16.
206. Chen L, Alexe G, Dharia NV, Ross L, Iniguez AB, Conway AS, et al. CRISPR-Cas9 screen reveals a MYCN-amplified neuroblastoma dependency on EZH2. The Journal of clinical investigation. 2018;128(1):446-62.
207. Westbrook TF, Martin ES, Schlabach MR, Leng Y, Liang AC, Feng B, et al. A genetic screen for candidate tumor suppressors identifies REST. Cell. 2005;121(6):837-48.
208. Luo B, Cheung HW, Subramanian A, Sharifnia T, Okamoto M, Yang X, et al. Highly parallel identification of essential genes in cancer cells. Proceedings of the National Academy of Sciences. 2008;105(51):20380-5.
209. Evers B, Jastrzebski K, Heijmans JP, Grernrum W, Beijersbergen RL, Bernards R. CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. Nature biotechnology. 2016;34(6):631-3.

210. Orchard RC, Wilen CB, Doench JG, Baldridge MT, McCune BT, Lee Y-CJ, et al. Discovery of a proteinaceous cellular receptor for a norovirus. Science. 2016;353(6302):933-6.
211. Marceau CD, Puschnik AS, Majzoub K, Ooi YS, Brewer SM, Fuchs G, et al. Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens. Nature. 2016;535(7610):159-63.
212. Savidis G, McDougall WM, Meraner P, Perreira JM, Portmann JM, Trincucci G, et al. Identification of Zika virus and dengue virus dependency factors using functional genomics. Cell reports. 2016;16(1):232-46.
213. Zhang R, Miner JJ, Gorman MJ, Rausch K, Ramage H, White JP, et al. A CRISPR screen defines a signal peptide processing pathway required by flaviviruses. Nature. 2016;535(7610):164-8.
214. Haga K, Fujimoto A, Takai-Todaka R, Miki M, Doan YH, Murakami K, et al. Functional receptor molecules CD300lf and CD300ld within the CD300 family enable murine noroviruses to infect cells. Proceedings of the National Academy of Sciences. 2016;113(41):E6248-E55.
215. Park RJ, Wang T, Koundakjian D, Hultquist JF, Lamothe-Molina P, Monel B, et al. A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. Nature genetics. 2017;49(2):193-203.
216. Ma H, Dang Y, Wu Y, Jia G, Anaya E, Zhang J, et al. A CRISPR-based screen identifies genes essential for West-Nile-virus-induced cell death. Cell reports. 2015;12(4):673-83.
217. Puschnik AS, Majzoub K, Ooi YS, Carette JE. A CRISPR toolbox to study virus–host interactions. Nature Reviews Microbiology. 2017;15(6):351-64.
218. Hou P, Wu C, Wang Y, Qi R, Bhavanasi D, Zuo Z, et al. A genome-wide CRISPR screen identifies genes critical for resistance to FLT3 inhibitor AC220. Cancer research. 2017;77(16):4402-13.
219. Koike-Yusa H, Li Y, Tan E-P, Velasco-Herrera MDC, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nature biotechnology. 2014;32(3):267-73.
220. Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature. 2015;517(7536):583-8.
221. Horlbeck MA, Gilbert LA, Villalta JE, Adamson B, Pak RA, Chen Y, et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. elife. 2016;5:e19760.
222. Wu Q, Tian Y, Zhang J, Tong X, Huang H, Li S, et al. In vivo CRISPR screening unveils histone demethylase UTX as an important epigenetic regulator in lung tumorigenesis. Proceedings of the National Academy of Sciences. 2018;115(17):E3978-E86.
223. Liao S, Davoli T, Leng Y, Li MZ, Xu Q, Elledge SJ. A genetic interaction analysis identifies cancer drivers that modify EGFR dependency. Genes & Development. 2017;31(2):184-96.
224. Han J, Perez JT, Chen C, Li Y, Benitez A, Kandasamy M, et al. Genome-wide CRISPR/Cas9 screen identifies host factors essential for influenza virus replication. Cell reports. 2018;23(2):596-607.
225. Kim HS, Lee K, Bae S, Park J, Lee C-K, Kim M, et al. CRISPR/Cas9-mediated gene knockout screens and target identification via whole-genome sequencing uncover host genes required for picornavirus infection. Journal of Biological Chemistry. 2017;292(25):10664-71.
226. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, et al. Genome-scale CRISPR-mediated control of gene repression and activation. Cell. 2014;159(3):647-61.

227. Unniyampurath U, Pilankatta R, Krishnan MN. RNA interference in the age of CRISPR: will CRISPR interfere with RNAi? International journal of molecular sciences. 2016;17(3):291.
228. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. cell. 2016;167(7):1853-66. e17.
229. Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, et al. Pooled CRISPR screening with single-cell transcriptome readout. Nature methods. 2017;14(3):297-301.
230. Agustí A, Vogelmeier C, Faner R. COPD 2020: changes and challenges. American Physiological Society Bethesda, MD; 2020. p. L879-L83.
231. Diaz-Guzman E, Mannino DM. Epidemiology and prevalence of chronic obstructive pulmonary disease. Clinics in chest medicine. 2014;35(1):7-16.
232. Cheyne L, Irvin-Sellers MJ, White J. Tiotropium versus ipratropium bromide for chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews. 2015(9).
233. Salvi S. Tobacco smoking and environmental risk factors for chronic obstructive pulmonary disease. Clinics in chest medicine. 2014;35(1):17-27.
234. Budden KF, Shukla SD, Rehman SF, Bowerman KL, Keely S, Hugenholtz P, et al. Functional effects of the microbiota in chronic respiratory disease. The lancet Respiratory medicine. 2019;7(10):907-20.
235. Chotirmall SH, Gellatly SL, Budden KF, Mac Aogain M, Shukla SD, Wood DL, et al. Microbiomes in respiratory health and disease: an Asia-Pacific perspective. Respirology. 2017;22(2):240-50.
236. Dua K, Malyla V, Singhvi G, Wadhwa R, Krishna RV, Shukla SD, et al. Increasing complexity and interactions of oxidative stress in chronic respiratory diseases: an emerging need for novel drug delivery systems. Chemico-biological interactions. 2019;299:168-78.
237. Tse HN, Tseng CZS. Update on the pathological processes, molecular biology, and clinical utility of N-acetylcysteine in chronic obstructive pulmonary disease. International journal of chronic obstructive pulmonary disease. 2014;9:825.
238. Chiba T, Chihara J, Furue M. Role of the Arylhydrocarbon Receptor (AhR) in the Pathology of Asthma and COPD. Journal of allergy. 2012;2012.
239. Thome T, Miguez K, Willms AJ, Burke SK, Chandran V, de Souza AR, et al. Chronic aryl hydrocarbon receptor activity phenocopies smoking-induced skeletal muscle impairment. Journal of Cachexia, Sarcopenia and Muscle. 2022;13(1):589-604.
240. de Souza AR, Zago M, Pollock SJ, Sime PJ, Phipps RP, Baglole CJ. Genetic ablation of the aryl hydrocarbon receptor causes cigarette smoke-induced mitochondrial dysfunction and apoptosis. Journal of Biological Chemistry. 2011;286(50):43214-28.
241. Hansbro PM, Hamilton MJ, Fricker M, Gellatly SL, Jarnicki AG, Zheng D, et al. Importance of mast cell Prss31/transmembrane tryptase/tryptase-γ in lung function and experimental chronic obstructive pulmonary disease and colitis. Journal of Biological Chemistry. 2014;289(26):18214-27.
242. Tay HL, Kaiko GE, Plank M, Li J, Maltby S, Essilfie A-T, et al. Antagonism of miR-328 increases the antimicrobial function of macrophages and neutrophils and rapid clearance of non-typeable Haemophilus influenzae (NTHi) from infected lung. PLoS pathogens. 2015;11(4):e1004549.
243. Zhang J, Xu Z, Kong L, Gao H, Zhang Y, Zheng Y, et al. miRNA-486-5p promotes COPD progression by targeting HAT1 to regulate the TLR4-triggered inflammatory response of alveolar macrophages. International Journal of Chronic Obstructive Pulmonary Disease. 2020;15:2991.

244. Kim RY, Rae B, Neal R, Donovan C, Pinkerton J, Balachandran L, et al. Elucidating novel disease mechanisms in severe asthma. Clinical & Translational Immunology. 2016;5(7):e91.
245. Liu G, Jarnicki AG, Paudel KR, Lu W, Wadhwa R, Philp AM, et al. Adverse roles of mast cell chymase-1 in COPD. European Respiratory Journal. 2022;60(6).
246. Kim RY, Sunkara KP, Bracke KR, Jarnicki AG, Donovan C, Hsu AC, et al. A microRNA-21–mediated SATB1/S100A9/NF-κB axis promotes chronic obstructive pulmonary disease pathogenesis. Science Translational Medicine. 2021;13(621):eaav7223.
247. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004;20(3):307-15.
248. Dessau R, Pipper CB. "R"--project for statistical computing. Ugeskrift for laeger. 2008;170(5):328-30.
249. Allaire J. RStudio: integrated development environment for R. Boston, MA. 2012;770(394):165-71.
250. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome biology. 2004;5(10):1-16.
251. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118-27.
252. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28(6):882-3.
253. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research. 2015;43(7):e47-e.
254. Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature. 2005;433(7027):769-73.
255. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC bioinformatics. 2013;14(1):7.
256. Skerrett-Byrne DA, Bromfield EG, Murray HC, Jamaluddin MFB, Jarnicki AG, Fricker M, et al. Time-resolved proteomic profiling of cigarette smoke-induced experimental chronic obstructive pulmonary disease. Respirology. 2021.
257. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic acids research. 2021;49(D1):D605-D12.
258. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic acids research. 2019.
259. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, et al. g: Profiler—a web server for functional interpretation of gene lists (2016 update). Nucleic acids research. 2016;44(W1):W83-W9.
260. Shen H. How to synthesize your gRNAs for CRISPR. 2016.
261. Brinkman EK, van Steensel B. Rapid Quantitative Evaluation of CRISPR Genome Editing by TIDE and TIDER. CRISPR Gene Editing: Springer; 2019. p. 29-44.
262. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic acids research. 2014;42(22):e168-e.
263. Hsiau T, Maures T, Waite K, Yang J, Kelso R, Holden K, et al. Inference of CRISPR edits from Sanger trace data. BioRxiv. 2018:251082.

264. Roginsky J. Analyzing CRISPR Editing Results: Synthego Developed a Tool Called ICE to Be More Efficient Than Other Methods. Genetic Engineering & Biotechnology News. 2018;38(11):S24-S6.
265. Prism G. Prism 8 for windows. GraphPad Software Inc; 2019.
266. Larigot L, Juricek L, Dairou J, Coumoul X. AhR signaling pathways and regulatory functions. Biochimie open. 2018;7:1-9.
267. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, et al. Pathway enrichment analysis and visualization of omics data using g: Profiler, GSEA, Cytoscape and EnrichmentMap. Nature protocols. 2019;14(2):482-517.
268. Danielson Pá. The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. Current drug metabolism. 2002;3(6):561-97.
269. Zuber R, Anzenbacherová E, Anzenbacher P. Cytochromes P450 and experimental models of drug metabolism. Journal of cellular and molecular medicine. 2002;6(2):189-98.
270. Androutsopoulos VP, Tsatsakis AM, Spandidos DA. Cytochrome P450 CYP1A1: wider roles in cancer progression and prevention. BMC cancer. 2009;9(1):1-17.
271. Tian L-X, Tang X, Zhu J-Y, Luo L, Ma X-Y, Cheng S-W, et al. Cytochrome P450 1A1 enhances inflammatory responses and impedes phagocytosis of bacteria in macrophages during sepsis. Cell Communication and Signaling. 2020;18:1-16.
272. Sekine H, Mimura J, Yamamoto M, Fujii-Kuriyama Y. Unique and overlapping transcriptional roles of arylhydrocarbon receptor nuclear translocator (Arnt) and Arnt2 in xenobiotic and hypoxic responses. Journal of Biological Chemistry. 2006;281(49):37507-16.
273. Dougherty EJ, Pollenz RS. Analysis of Ah receptor-ARNT and Ah receptor-ARNT2 complexes in vitro and in cell culture. Toxicological sciences. 2008;103(1):191-206.
274. Hankinson O. Why does ARNT2 behave differently from ARNT? Toxicological sciences. 2008;103(1):1-3.
275. Zong D, Li J, Cai S, He S, Liu Q, Jiang J, et al. Notch1 regulates endothelial apoptosis via the ERK pathway in chronic obstructive pulmonary disease. American Journal of Physiology-Cell Physiology. 2018;315(3):C330-C40.
276. Quillard T, Coupel S, Coulon F, Fitau J, Chatelais M, Cuturi M, et al. Impaired Notch4 activity elicits endothelial cell activation and apoptosis: implication for transplant arteriosclerosis. Arteriosclerosis, thrombosis, and vascular biology. 2008;28(12):2258-65.
277. Bensellam M, Montgomery MK, Luzuriaga J, Chan JY, Laybutt DR. Inhibitor of differentiation proteins protect against oxidative stress by regulating the antioxidant–mitochondrial response in mouse beta cells. Diabetologia. 2015;58(4):758-70.
278. Sauler M, McDonough J, Adams T, Kothapalli N, Schupp J, Nouws J, et al. Single-cell RNA sequencing identifies aberrant transcriptional profiles of cellular populations and altered alveolar niche signalling networks in Chronic Obstructive Pulmonary Disease (COPD). medRxiv. 2020.
279. Yin L, Liu M-x, Li W, Wang F-y, Tang Y-h, Huang C-x. Over-expression of inhibitor of differentiation 2 attenuates post-infarct cardiac fibrosis through inhibition of TGF-β1/SMAD3/HIF-1α/Il-11 signaling pathway. Frontiers in pharmacology. 2019;10:1349.
280. George T, Chakraborty M, Giembycz MA, Newton R. A bronchoprotective role for Rgs2 in a murine model of lipopolysaccharide-induced airways inflammation. Allergy, Asthma & Clinical Immunology. 2018;14(1):1-14.
281. Newton R, Giembycz MA. Understanding how long-acting β2-adrenoceptor agonists enhance the clinical efficacy of inhaled corticosteroids in asthma–an update. British journal of pharmacology. 2016;173(24):3405-30.
282. Newton R. Regulators of G-protein signaling as asthma therapy? : American Thoracic Society; 2018. p. 7-9.

283. Nishina T, Komazawa-Sakon S, Yanaka S, Piao X, Zheng D-M, Piao J-H, et al. Interleukin-11 links oxidative stress and compensatory proliferation. Science signaling. 2012;5(207):ra5-ra.
284. Vannahme C, Smyth N, Miosge N, Gosling S, Frie C, Paulsson M, et al. Characterization of SMOC-1, a novel modular calcium-binding protein in basement membranes. Journal of Biological Chemistry. 2002;277(41):37977-86.
285. Gersdorff N, Müller M, Schall A, Miosge N. Secreted modular calcium-binding protein-1 localization during mouse embryogenesis. Histochemistry and cell biology. 2006;126(6):705-12.
286. Bradshaw AD. Diverse biological functions of the SPARC family of proteins. The international journal of biochemistry & cell biology. 2012;44(3):480-8.
287. Choi Y-A, Lim J, Kim KM, Acharya B, Cho J-Y, Bae Y-C, et al. Secretome analysis of human BMSCs and identification of SMOC1 as an important ECM protein in osteoblast differentiation. Journal of proteome research. 2010;9(6):2946-56.
288. Wang Y, Wu X. SMOC1 silencing suppresses the angiotensin II-induced myocardial fibrosis of mouse myocardial fibroblasts via affecting the BMP2/Smad pathway. Oncology Letters. 2018;16(3):2903-10.
289. Bekki K, Vogel H, Li W, Ito T, Sweeney C, Haarmann-Stemmann T, et al. The aryl hydrocarbon receptor (AhR) mediates resistance to apoptosis induced in breast cancer cells. Pesticide biochemistry and physiology. 2015;120:5-13.
290. Lu Z, Van Eeckhoutte HP, Liu G, Nair PM, Jones B, Gillis CM, et al. Necroptosis signaling promotes inflammation, airway remodeling, and emphysema in chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine. 2021;204(6):667-81.
291. De Smet EG, Mestdagh P, Vandesompele J, Brusselle GG, Bracke KR. Non-coding RNAs in the pathogenesis of COPD. Thorax. 2015;70(8):782-91.
292. O'Brien J, Hayder H, Zayed Y, Peng C. Overview of microRNA biogenesis, mechanisms of actions, and circulation. Frontiers in endocrinology. 2018;9:402.
293. Dua K, Hansbro NG, Foster PS, Hansbro PM. MicroRNAs as therapeutics for future drug delivery systems in treatment of lung diseases. Drug delivery and translational research. 2017;7(1):168-78.
294. Foster PS, Plank M, Collison A, Tay HL, Kaiko GE, Li J, et al. The emerging role of micro RNA s in regulating immune and inflammatory responses in the lung. Immunological reviews. 2013;253(1):198-215.
295. Conickx G, Mestdagh P, Avila Cobos F, Verhamme FM, Maes T, Vanaudenaerde BM, et al. MicroRNA profiling reveals a role for microRNA-218-5p in the pathogenesis of chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine. 2017;195(1):43-56.
296. Ge L, Habiel DM, Hansbro PM, Kim RY, Gharib SA, Edelman JD, et al. miR-323a-3p regulates lung fibrosis by targeting multiple profibrotic pathways. JCI insight. 2016;1(20).
297. Hansbro PM, Kim RY, Starkey MR, Donovan C, Dua K, Mayall JR, et al. Mechanisms and treatments for severe, steroid-resistant allergic airway disease and asthma. Immunological reviews. 2017;278(1):41-62.
298. Kim RY, Horvat JC, Pinkerton JW, Starkey MR, Essilfie AT, Mayall JR, et al. MicroRNA-21 drives severe, steroid-insensitive experimental asthma by amplifying phosphoinositide 3-kinase–mediated suppression of histone deacetylase 2. Journal of Allergy and Clinical Immunology. 2017;139(2):519-32.
299. Savarimuthu Francis SM, Davidson MR, Tan ME, Wright CM, Clarke BE, Duhig EE, et al. MicroRNA-34c is associated with emphysema severity and modulates SERPINE1 expression. BMC genomics. 2014;15(1):1-8.

300. Ezzie ME, Crawford M, Cho J-H, Orellana R, Zhang S, Gelinas R, et al. Gene expression networks in COPD: microRNA and mRNA regulation. Thorax. 2012;67(2):122-31.
301. Izzotti A, Calin GA, Arrigo P, Steele VE, Croce CM, De Flora S. Downregulation of microRNA expression in the lungs of rats exposed to cigarette smoke. The FASEB Journal. 2009;23(3):806-12.
302. Advani J, Subbannayya Y, Patel K, Khan AA, Patil AH, Jain AP, et al. Long-term cigarette smoke exposure and changes in MiRNA expression and proteome in non-small-cell lung cancer. Omics: a journal of integrative biology. 2017;21(7):390-403.
303. Conickx G, Avila Cobos F, van den Berge M, Faiz A, Timens W, Hiemstra PS, et al. microRNA profiling in lung tissue and bronchoalveolar lavage of cigarette smoke-exposed mice and in COPD patients: a translational approach. Scientific reports. 2017;7(1):1-14.
304. Osei ET, Florez-Sampedro L, Tasena H, Faiz A, Noordhoek JA, Timens W, et al. miR-146a-5p plays an essential role in the aberrant epithelial–fibroblast cross-talk in COPD. European Respiratory Journal. 2017;49(5).
305. Wang C, Feng D, Dong S, He R, Fan B. Dysregulated circulating microRNA-126 in chronic obstructive pulmonary disease: linkage with acute exacerbation risk, severity degree, and inflammatory cytokines. Journal of clinical laboratory analysis. 2022;36(3):e24204.
306. Cañas JA, Rodrigo-Muñoz JM, Sastre B, Gil-Martinez M, Redondo N, Del Pozo V. MicroRNAs as potential regulators of immune response networks in asthma and chronic obstructive pulmonary disease. Frontiers in Immunology. 2021;11:608666.
307. Min H, Yoon S. Got target?: computational methods for microRNA target prediction and their extension. Experimental & molecular medicine. 2010;42(4):233-44.
308. Faiz A, Steiling K, Roffel MP, Postma DS, Spira A, Lenburg ME, et al. Effect of long-term corticosteroid treatment on microRNA and gene-expression profiles in COPD. European Respiratory Journal. 2019;53(4).
309. Dang X, Qu X, Wang W, Liao C, Li Y, Zhang X, et al. Bioinformatic analysis of microRNA and mRNA Regulation in peripheral blood mononuclear cells of patients with chronic obstructive pulmonary disease. Respiratory Research. 2017;18(1):1-13.
310. Bowerman KL, Rehman SF, Vaughan A, Lachner N, Budden KF, Kim RY, et al. Disease-associated gut microbiome and metabolome changes in patients with chronic obstructive pulmonary disease. Nature communications. 2020;11(1):1-15.
311. Haw TJ, Starkey MR, Pavlidis S, Fricker M, Arthurs AL, Nair PM, et al. Toll-like receptor 2 and 4 have opposing roles in the pathogenesis of cigarette smoke-induced chronic obstructive pulmonary disease. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2018;314(2):L298-L317.
312. Starkey MR, Plank MW, Casolari P, Papi A, Pavlidis S, Guo Y, et al. IL-22 and its receptors are increased in human and experimental COPD and contribute to pathogenesis. European Respiratory Journal. 2019;54(1).
313. Donovan C, Starkey MR, Kim RY, Rana BM, Barlow JL, Jones B, et al. Roles for T/B lymphocytes and ILC2s in experimental chronic obstructive pulmonary disease. Journal of leukocyte biology. 2019;105(1):143-50.
314. Hansbro PM, Kim RY, Beckett EL, Jarnicki AG, Sunkara KP. Lung miRNA expression profiling in a mouse model of chronic obstructive pulmonary disease. 2021.
315. Dessau RB, Pipper CB. "R"--project for statistical computing. Ugeskrift for laeger. 2008;170(5):328-30.
316. Blighe K, Lewis M, Lun A, Blighe MK. Package 'PCAtools.'. PCAtools; 2019.
317. Lopez-Romero P, Lopez-Romero MP, Biobase I, biocViews Microarray A, OneChannel P. Package 'AgiMicroRna'. 2013.

318. van den Berge M, Steiling K, Timens W, Hiemstra PS, Sterk PJ, Heijink IH, et al. Airway gene expression in COPD is dynamic with inhaled corticosteroid treatment and reflects biological pathways associated with disease activity. Thorax. 2014;69(1):14-23.
319. Revelle W. An overview of the psych package. 2011.
320. Ru Y, Kechris KJ, Tabakoff B, Hoffman P, Radcliffe RA, Bowler R, et al. The multiMiR R package and database: integration of microRNA–target interactions along with their disease and drug associations. Nucleic acids research. 2014;42(17):e133-e.
321. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA–target interactions. Nucleic acids research. 2009;37(suppl_1):D105-D10.
322. Hsu S-D, Lin F-M, Wu W-Y, Liang C, Huang W-C, Chan W-L, et al. miRTarBase: a database curates experimentally validated microRNA–target interactions. Nucleic acids research. 2011;39(suppl_1):D163-D9.
323. Vergoulis T, Vlachos IS, Alexiou P, Georgakilas G, Maragkakis M, Reczko M, et al. TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. Nucleic acids research. 2012;40(D1):D222-D9.
324. Maragkakis M, Vergoulis T, Alexiou P, Reczko M, Plomaritou K, Gousis M, et al. DIANA-microT Web server upgrade supports Fly and Worm miRNA target prediction and bibliographic miRNA to disease association. Nucleic acids research. 2011;39(suppl_2):W145-W8.
325. Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M. Inference of miRNA targets using evolutionary conservation and pathway analysis. BMC bioinformatics. 2007;8:1-22.
326. Griffiths-Jones S, Saini HK, Van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic acids research. 2007;36(suppl_1):D154-D8.
327. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA. org resource: targets and expression. Nucleic acids research. 2008;36(suppl_1):D149-D53.
328. Wang X. miRDB: a microRNA target prediction and functional annotation database with a wiki interface. Rna. 2008;14(6):1012-7.
329. Anders G, Mackowiak SD, Jens M, Maaskola J, Kuntzagk A, Rajewsky N, et al. doRiNA: a database of RNA interactions in post-transcriptional regulation. Nucleic acids research. 2012;40(D1):D180-D6.
330. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. Nature genetics. 2007;39(10):1278-84.
331. Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Molecular cell. 2007;27(1):91-105.
332. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic acids research. 2009;37(suppl_1):D98-D104.
333. Rukov JL, Wilentzik R, Jaffe I, Vinther J, Shomron N. Pharmaco-miR: linking microRNAs and drug effects. Briefings in bioinformatics. 2014;15(4):648-59.
334. Ruepp A, Kowarsch A, Schmidl D, Buggenthin F, Brauner B, Dunger I, et al. PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. Genome biology. 2010;11:1-11.
335. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic acids research. 2019;47(W1):W191-W8.
336. Plank MW, Maltby S, Tay HL, Stewart J, Eyers F, Hansbro PM, et al. MicroRNA expression is altered in an ovalbumin-induced asthma model and targeting miR-155 with antagomirs reveals cellular specificity. PloS one. 2015;10(12):e0144810.

337. Tantisuwat A, Thaveeratitham P. Effects of smoking on chest expansion, lung function, and respiratory muscle strength of youths. Journal of physical therapy science. 2014;26(2):167-70.
338. Tommola M, Ilmarinen P, Tuomisto LE, Haanpää J, Kankaanranta T, Niemelä O, et al. The effect of smoking on lung function: a clinical study of adult-onset asthma. European Respiratory Journal. 2016;48(5):1298-306.
339. Jaakkola JJ, Hernberg S, Lajunen TK, Sripaijboonkij P, Malmberg LP, Jaakkola MS. Smoking and lung function among adults with newly onset asthma. BMJ open respiratory research. 2019;6(1):e000377.
340. Hancox RJ, Gray AR, Poulton R, Sears MR. The effect of cigarette smoking on lung function in young adults with asthma. American journal of respiratory and critical care medicine. 2016;194(3):276-84.
341. Isabel U, Alberto C, María QJ, Nerea M, Xavier B, Jordi S. Smoking habit, respiratory symptoms and lung function in young adults. The European Journal of Public Health. 2005;15(2):160-5.
342. Izzotti A, Calin GA, Arrigo P, Steele VE, Croce CM, De Flora S. Downregulation of microRNA expression in the lungs of rats exposed to cigarette smoke. The FASEB Journal. 2009;23(3):806.
343. Izzotti A, Calin GA, Steele VE, Cartiglia C, Longobardi M, Croce CM, et al. Chemoprevention of cigarette smoke–induced alterations of microRNA expression in rat lungs. Cancer Prevention Research. 2010;3(1):62-72.
344. Schembri F, Sridhar S, Perdomo C, Gustafson AM, Zhang X, Ergun A, et al. MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. Proceedings of the National Academy of Sciences. 2009;106(7):2319-24.
345. Vucic EA, Thu KL, Pikor LA, Enfield KS, Yee J, English JC, et al. Smoking status impacts microRNA mediated prognosis and lung adenocarcinoma biology. BMC cancer. 2014;14:1-14.
346. Zhang M, Lu Y, Liu L, Zhang X, Ning J. Role and mechanism of miR-181a-5p in mice with chronic obstructive pulmonary disease by regulating HMGB1 and the NF-κB pathway. Cells Tissues Organs. 2022:1-13.
347. Steiling K, Van Den Berge M, Hijazi K, Florido R, Campbell J, Liu G, et al. A dynamic bronchial airway gene expression signature of chronic obstructive pulmonary disease and lung function impairment. American journal of respiratory and critical care medicine. 2013;187(9):933-42.
348. De Smet EG, Van Eeckhoutte HP, Cobos FA, Blomme E, Verhamme FM, Provoost S, et al. The role of miR-155 in cigarette smoke-induced pulmonary inflammation and COPD. Mucosal immunology. 2020;13(3):423-36.
349. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews genetics. 2009;10(1):57-63.
350. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends in genetics. 2008;24(3):133-41.
351. Patel RR, Ryu JH, Vassallo R. Cigarette smoking and diffuse lung disease. Drugs. 2008;68(11):1511-27.
352. Stämpfli MR, Anderson GP. How cigarette smoke skews immune responses to promote infection, lung disease and cancer. Nature Reviews Immunology. 2009;9(5):377-84.
353. Vassallo R, Tamada K, Lau JS, Kroening PR, Chen L. Cigarette smoke extract suppresses human dendritic cell function leading to preferential induction of Th-2 priming. The Journal of Immunology. 2005;175(4):2684-91.

354. D'hulst A, Vermaelen K, Brusselle G, Joos G, Pauwels R. Time course of cigarette smoke-induced pulmonary inflammation in mice. European Respiratory Journal. 2005;26(2):204-13.
355. Lu L-M, Zavitz CC, Chen B, Kianpour S, Wan Y, Stampfli MR. Cigarette smoke impairs NK cell-dependent tumor immune surveillance. The Journal of Immunology. 2007;178(2):936-43.
356. Israel-Assayag E, Dakhama A, Lavigne S, Laviolette M, Cormier Y. Expression of costimulatory molecules on alveolar macrophages in hypersensitivity pneumonitis. American journal of respiratory and critical care medicine. 1999;159(6):1830-4.
357. Hou W, Hu S, Li C, Ma H, Wang Q, Meng G, et al. Cigarette smoke induced lung barrier dysfunction, EMT, and tissue remodeling: a possible link between COPD and lung cancer. BioMed research international. 2019;2019.
358. Zong D, Liu X, Li J, Ouyang R, Chen P. The role of cigarette smoke-induced epigenetic alterations in inflammation. Epigenetics & Chromatin. 2019;12(1):1-25.
359. Baraldo S, Turato G, Saetta M. Pathophysiology of the small airways in chronic obstructive pulmonary disease. Respiration. 2012;84(2):89-97.
360. Leopold PL, O'Mahony MJ, Lian XJ, Tilley AE, Harvey B-G, Crystal RG. Smoking is associated with shortened airway cilia. PloS one. 2009;4(12):e8157.
361. MAESTRELLI P, SAETTA M, MAPP CE, FABBRI LM. Remodeling in response to infection and injury: airway inflammation and hypersecretion of mucus in smoking subjects with chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine. 2001;164(supplement_2):S76-S80.
362. van Eeden SF, Hogg JC. Immune-modulation in chronic obstructive pulmonary disease: current concepts and future strategies. Respiration. 2020;99(7):550-65.
363. Domej W, Oettl K, Renner W. Oxidative stress and free radicals in COPD–implications and relevance for treatment. International journal of chronic obstructive pulmonary disease. 2014;9:1207.
364. Yoshida T, Tuder RM. Pathobiology of cigarette smoke-induced chronic obstructive pulmonary disease. Physiological reviews. 2007;87(3):1047-82.
365. Zinellu E, Zinellu A, Fois AG, Carru C, Pirina P. Circulating biomarkers of oxidative stress in chronic obstructive pulmonary disease: a systematic review. Respiratory research. 2016;17(1):1-11.
366. Zeng H, Kong X, Zhang H, Chen Y, Cai S, Luo H, et al. Inhibiting DNA methylation alleviates cigarette smoke extract-induced dysregulation of Bcl-2 and endothelial apoptosis. Tobacco induced diseases. 2020;18.
367. Demedts IK, Demoor T, Bracke KR, Joos GF, Brusselle GG. Role of apoptosis in the pathogenesis of COPD and pulmonary emphysema. Respiratory research. 2006;7(1):1-10.
368. Song Q, Chen P, Liu X-M. The role of cigarette smoke-induced pulmonary vascular endothelial cell apoptosis in COPD. Respiratory Research. 2021;22(1):1-15.
369. Wu L, Ge Y, Yuan Y, Li H, Sun H, Xu C, et al. Genome-wide CRISPR screen identifies MTA3 as an inducer of gemcitabine resistance in pancreatic ductal adenocarcinoma. Cancer Letters. 2022;548:215864.
370. Huang S, Ma Z, Zhou Q, Wang A, Gong Y, Li Z, et al. Genome-wide CRISPR/Cas9 library screening identified that dusp4 deficiency induces lenvatinib resistance in hepatocellular carcinoma. International Journal of Biological Sciences. 2022;18(11):4357.
371. Szlachta K, Kuscu C, Tufan T, Adair SJ, Shang S, Michaels AD, et al. CRISPR knockout screening identifies combinatorial drug targets in pancreatic cancer and models cellular drug response. Nature communications. 2018;9(1):1-13.
372. Doench JG. Am I ready for CRISPR? A user's guide to genetic screens. Nature Reviews Genetics. 2018;19(2):67-80.

373. Hart T, Tong AHY, Chan K, Van Leeuwen J, Seetharaman A, Aregger M, et al. Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. G3: Genes, Genomes, Genetics. 2017;7(8):2719-27.
374. Ludwig ML, Michmerhuizen NM, Hoesli RC, Mann JE, Devenport SN, Kulkarni AS, et al. 15 Generation and Utilization of CRISPR/Cas9 Screening Libraries in Mammalian Cells. Genome editing and engineering: From TALENs, ZFNs and CRISPRs to molecular surgery. 2018:223.
375. Shalem O, Sanjana NE, Zhang F. High-throughput functional genomics using CRISPR–Cas9. Nature Reviews Genetics. 2015;16(5):299-311.
376. Grodzki M, Bluhm AP, Schaefer M, Tagmount A, Russo M, Sobh A, et al. Genome-scale CRISPR screens identify host factors that promote human coronavirus infection. Genome Medicine. 2022;14(1):1-18.
377. Wang B, Wang M, Zhang W, Xiao T, Chen C-H, Wu A, et al. Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. Nature protocols. 2019;14(3):756-80.
378. Xie Z, Bailey A, Kuleshov MV, Clarke DJ, Evangelista JE, Jenkins SL, et al. Gene set knowledge discovery with enrichr. Current protocols. 2021;1(3):e90.
379. Chen E. Kuleshov et al., 2016 http://amp. pharm. mssm. edu. Enrichr; 2013.
380. Croft D, O'kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database of reactions, pathways and biological processes. Nucleic acids research. 2010;39(suppl_1):D691-D7.
381. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The reactome pathway knowledgebase. Nucleic acids research. 2018;46(D1):D649-D55.
382. Killian T, Gatto L. Exploiting the DepMap cancer dependency data using the Depmap R package. F1000Research. 2021;10(416):416.
383. Kang N, Chen P, Chen Y, Zeng H, He X, Zhu Y. PRMT6 mediates CSE induced inflammation and apoptosis. International immunopharmacology. 2015;24(1):95-101.
384. Chang SS, Jiang WW, Smith I, Glazer C, Sun WY, Mithani S, et al. Chronic cigarette smoke extract treatment selects for apoptotic dysfunction and mitochondrial mutations in minimally transformed oral keratinocytes. International journal of cancer. 2010;126(1):19-27.
385. Inukai R, Mori K, Kuwata K, Suzuki C, Maki M, Takahara T, et al. The novel ALG-2 target protein CDIP1 promotes cell death by interacting with ESCRT-I and VAPA/B. International Journal of Molecular Sciences. 2021;22(3):1175.
386. He T-S, Ji W, Zhang J, Lu J, Liu X. ALG-2 couples T cell activation and apoptosis by regulating proteasome activity and influencing MCL1 stability. Cell death & disease. 2020;11(1):1-15.
387. Huang G, Li H, Zhang H. Abnormal expression of mitochondrial ribosomal proteins and their encoding genes with cell apoptosis and diseases. International journal of molecular sciences. 2020;21(22):8879.
388. Navratilova Z, Kolek V, Petrek M. Matrix metalloproteinases and their inhibitors in chronic obstructive pulmonary disease. Archivum immunologiae et therapiae experimentalis. 2016;64:177-93.
389. Mahor D, Kumari V, Vashisht K, Galgalekar R, Samarth RM, Mishra PK, et al. Elevated serum matrix metalloprotease (MMP-2) as a candidate biomarker for stable COPD. BMC pulmonary medicine. 2020;20:1-9.
390. Yang L, Tan W, Yang X, You Y, Wang J, Wen G, et al. Sorting nexins: A novel promising therapy target for cancerous/neoplastic diseases. Journal of Cellular Physiology. 2021;236(5):3317-35.
391. Vieira N, Deng F-M, Liang F-X, Liao Y, Chang J, Zhou G, et al. SNX31: a novel sorting nexin associated with the uroplakin-degrading multivesicular bodies in terminally differentiated urothelial cells. PloS one. 2014;9(6):e99644.

392. Liao Y, Tham DK, Liang F-X, Chang J, Wei Y, Sudhir P-R, et al. Mitochondrial lipid droplet formation as a detoxification mechanism to sequester and degrade excessive urothelial membranes. Molecular biology of the cell. 2019;30(24):2969-84.
393. Yim WW-Y, Mizushima N. Lysosome biology in autophagy. Cell discovery. 2020;6(1):6.
394. Nyunoya T, Mebratu Y, Contreras A, Delgado M, Chand HS, Tesfaigzi Y. Molecular processes that drive cigarette smoke–induced epithelial cell fate of the lung. American journal of respiratory cell and molecular biology. 2014;50(3):471-82.
395. Nakahira K, Cloonan SM, Mizumura K, Choi AM, Ryter SW. Autophagy: a crucial moderator of redox balance, inflammation, and apoptosis in lung disease. Antioxidants & redox signaling. 2014;20(3):474-94.
396. Guan X, Lu J, Sun F, Li Q, Pang Y. The molecular evolution and functional divergence of lamprey programmed cell death genes. Frontiers in Immunology. 2019;10:1382.
397. Johnstone RW, Gerber M, Landewe T, Tollefson A, Wold WS, Shilatifard A. Functional analysis of the leukemia protein ELL: evidence for a role in the regulation of cell growth and survival. Molecular and Cellular Biology. 2001;21(5):1672-81.
398. Zuo ZH, Yu YP, Martin A, Luo JH. Cellular stress response 1 down-regulates the expression of epidermal growth factor receptor and platelet-derived growth factor receptor through inactivation of splicing factor 3A3. Molecular carcinogenesis. 2017;56(2):315-24.
399. Burrell JA, Stephens JM. KAT8, lysine acetyltransferase 8, is required for adipocyte differentiation in vitro. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease. 2021;1867(6):166103.
400. Tyteca S, Legube G, Trouche D. To die or not to die: a HAT trick. Molecular cell. 2006;24(6):807-8.
401. Sykes SM, Mellert HS, Holbert MA, Li K, Marmorstein R, Lane WS, et al. Acetylation of the p53 DNA-binding domain regulates apoptosis induction. Molecular cell. 2006;24(6):841-51.
402. Mellert HS, McMahon SB. hMOF, a KAT (8) with many lives. Molecular cell. 2009;36(2):174-5.
403. Zhou T, Zhong Y, Hu Y, Sun C, Wang Y, Wang G. PM2. 5 downregulates miR-194-3p and accelerates apoptosis in cigarette-inflamed bronchial epithelium by targeting death-associated protein kinase 1. International journal of chronic obstructive pulmonary disease. 2018;13:2339.
404. Park H-R, Vallarino J, O'Sullivan M, Wirth C, Panganiban RA, Webb G, et al. Electronic cigarette smoke reduces ribosomal protein gene expression to impair protein synthesis in primary human airway epithelial cells. Scientific Reports. 2021;11(1):1-11.
405. Liu Y, Deisenroth C, Zhang Y. RP–MDM2–p53 pathway: Linking ribosomal biogenesis and tumor surveillance. Trends in cancer. 2016;2(4):191-204.
406. Dong J, Liao W, Peh HY, Tan WD, Zhou S, Wong WF. Ribosomal protein S3 gene silencing protects against cigarette smoke-induced acute lung injury. Molecular Therapy-Nucleic Acids. 2018;12:370-80.
407. Chu Y, Li D, Zhang H, Ding J, Xu P, Qiu X. PIG3 suppresses gastric cancer proliferation by regulating p53-mediated apoptosis. Journal of Biological Regulators and Homeostatic Agents. 2018;32(5):1185-9.
408. Karimian A, Ahmadi Y, Yousefi B. Multiple functions of p21 in cell cycle, apoptosis and transcriptional regulation after DNA damage. DNA repair. 2016;42:63-71.
409. Cho J, Park J, Shin SC, Kim J-H, Kim EE, Song EJ. Ribosomal protein S2 interplays with MDM2 to induce p53. Biochemical and biophysical research communications. 2020;523(2):542-7.

410. Li H, Zhang H, Huang G, Bing Z, Xu D, Liu J, et al. Loss of RPS27a expression regulates the cell cycle, apoptosis, and proliferation via the RPL11-MDM2-p53 pathway in lung adenocarcinoma cells. Journal of Experimental & Clinical Cancer Research. 2022;41(1):1-20.
411. Zhang X, Wang W, Wang H, Wang M, Xu W, Zhang R. Identification of ribosomal protein S25 (RPS25)–MDM2–p53 regulatory feedback loop. Oncogene. 2013;32(22):2782-91.
412. Ju J, He Y. PRMT5 promotes inflammation of cigarette smoke extract-induced bronchial epithelial cells by up-regulation of CXCL10. Allergologia et immunopathologia. 2021;49(5):131-6.
413. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4(2):249-64.
414. Lives B. GraphPad Prism. 2014.
415. Pellegrini R. Edit single bases with Benchling. 2016.

Appendix

Chapter 2: Longitudinal analysis of genome-wide expression profiling of cigarette smoke-induced experimental COPD

Methodology

COPD mouse model

All experimental mice, including BALB/c, inhaled the smoke of 12 cigarettes (3R4F reference cigarettes, University of Kentucky, Lexington, Ky) twice per day for 5 days a week for up to 12 weeks using a custom-designed and purpose-built exposure system. Each exposure lasted for 75 minutes. According to the defined protocol, time was defined as a categorical variable with 4 weeks as baseline, 6, 8, and 12 weeks.

RNA extraction and Transcriptome profiling

Total RNA from the whole lung homogenates was extracted with TRIzol® and quantitated using a NanoDropTM ND1000 spectrophotometer. The RNA Integrity Number (RIN) of each sample of BALB/c mice was measured using an Agilent RNA6000 Nano LabChip Kit and Agilent 2100 Bioanalyser system according to the manufacturer's instructions. Briefly, the chip was primed by adding 9μL of the gel-dye mix to each of the two wells on the chip after 1μL of dye concentration had been added to 65μL of column-filtered gel matrix. First, 5μL of nano marker (buffer) was added to sample and ladder wells, which were then supplemented with 1μL of heat-denatured RNA samples (70°C for 2 minutes) and ladder to their corresponding wells. Loaded chips were vortexed before the run. A total of 4 samples having RIN >8 were selected in each group for expression profiling. Gene expression profiling for the time-course

experiment was performed on the Affymetrix Mouse 430 plate array platform at each time point.

Data processing and statistical analysis

Data was analyzed using R statistical software (v3.5.2) (247, 248) with RStudio version 1.3.1 (249) and Bioconductor (250). A preliminary exploratory analysis was performed to identify the outliers and other problems through principal component analysis (PCA) using the "PCAtools" package (316). To obtain an estimate of the normalized signal for each gene, gene expression data were pre-processed for background correction, normalization, and expression calculation through the Robust Multi-Array Average (RMA) algorithm (413) in *affy* package (38). The measured intensities were log2 transformed. Differential expressions of genes were analyzed with the "limma" package (253, 254) of R and corrected for multiple testing using the Benjamini–Hochberg (BH) false discovery rate (FDR). Other statistical testing methods were used GraphPad Prism (v8) (414).

Batch effect correction

Since the BALB/c mice RNA samples from 8 weeks timepoint run on different machines led to the technical variation or batch effect in the data. To correct the batch effect from our data and remove the potentially unwanted source of interpretation, we used the ComBat (251) function implemented in the sva package (https://bioconductor.org/packages/sva) (252). The function adjusts for known batches using an empirical Bayesian framework (251).

Characterization of functionally enriched DEGs

Pathway analysis was performed unbiasedly using the STRING database and g:Profiler (https://biit.cs.ut.ee/gprofiler/gost). The significant DEGs with FDR<0.05 and FC >|2| during longitudinal study were provided, as an 'ordered query' in g:Profiler. '*Mus musculus* (Mouse)'

was selected as the organism, and the databases included in the g:Profiler platform were Gene Ontology, KEGG, Reactome, and WikiPathways. The top pathways associated with the differentially expressed genes were further analyzed. We also chose Gene Ontology (GO), Reactome, and WikiPathways from the STRING database for pathway analysis. DEGs provided to STRING were also observed in clusters to observe the closeness among them. The algorithm used was the K-Means clustering algorithm.

Protein expression profile of longitudinally differential expressed genes

To investigate the relationship between DEGs with protein abundance, we compared the fold change (FC) expression of membrane-enriched proteins and soluble proteins with fold change expression values of top candidates. CS-induced experimental COPD proteome was profiled using the iTRAQ8plex system *by David et al.,* in 2021 (256).

Four mice's lungs from the normal air or CS-exposed groups were perfused with tris-buffered saline supplemented with protease and phosphatase inhibitors (Roche, Complete EDTA free) at each time point, and the lungs were then extracted for proteome analysis. Using commercial desalting columns (Oasis, Waters), all peptide solutions were cleaned and desalted. Chemical isobaric tag-based approaches were utilized to label 200 µg of the peptide using quantitative fluorescent peptide quantification (Qubit protein assay kit, Thermo Fisher Scientific, Carlsbad, CA, USA), in compliance with the manufacturer's instructions (iTRAQ, SCIEX). The effectiveness of isobaric tag labeling and digestion was assessed using nano-liquid chromatography-tandem mass spectrometry (nLC-MS/MS). Before high-resolution nLC-MS/MS, proteome populations were desalted using a modified StageTip microcolumn and exposed to offline hydrophilic interaction liquid chromatography (HILIC). A Q-Exactive Plus hybrid quadrupole-Orbitrap MS system (Thermo Fisher Scientific) connected to a Dionex Ultimate 3000RSLC nanoflow HPLC system (Thermo Fisher Scientific) was used to perform

reverse phase nLC-MS/MS. The Acclaim PepMap100 C18 trap column (Thermo Fisher Scientific) was used to load samples for pre-concentration and online desalting. Next, separation was accomplished with an EASY-Spray PepMap C18 (Thermo Fisher Scientific). Full MS/data-dependent acquisition MS/MS mode (data-dependent acquisition) was used when operating the Q-Exactive Plus MS System. The complete MS was acquired using the Orbitrap mass analyzer (256).

All raw files were searched using a database using Proteome Discoverer 2.1 (Thermo Fisher Scientific). The Swiss_Prot and Uniprot_mouse databases (containing 25,041 sequences, retrieved on July 11, 2017) were searched against using Mascot 2.2.3 and SEQUEST HT. Every iTRAQ 8plex was examined separately. The Peptide and Protein Quantifier node was used to perform normalization. The normalized abundances are then scaled using the channel average by Proteome Discoverer 2.1 after normalization. Using scaled and normalized abundances, iTRAQ quantification ratios were produced by dividing each smoke-exposed mouse by the average of the control mice at each time point (256).

Cell culture

The normal human bronchial epithelium (BEAS-2B) cells were cultured in DMEM media supplemented with 5% Fetal Bovine Serum (FBS) and 1% Penicillin [100U/ml]-streptomycin [100mg/ml] antibody. Cells were incubated at 37^{o}C in a gas mixture of 5% CO_2/95% air. The cell number was determined using a hemocytometer. After reaching a confluence of 80%, the cells were trypsinized and passaged in the ratio of 1:10 concentration from the previously confluent flask.

CRISPR-cas9 vector design and construction for gene modification in cell lines

Paired guided RNA (gRNA) was designed against human (ENSG00000172379) *ARNT2* sequence using Benchling (415) online software (version 2018, San Fransisco, CA, USA). CRISPR gRNA (sequence: GAAGGGCAGCAGTCATCCATG and PAM: AGG) with the best on-target (score = 74.7) and off-target (score = 60.8) targeted at exon 6 of positive strand of ARNT2 at 80508212position was chosen to target.

For the human ARNT2 sequence, the paired gRNA primers designed were gRNA_F (5'-CACCGAAGGGCAGCAGTCATCCATG-3') and gRNA_R (5'-aaacCATGGATGACTGCTGCCCTTC-3'). CRISPR gRNA sequences were ligated into the pX458 backbone vector (plasmid 48138, Addgene). Following the construction of the CRISPR gRNA plasmid, the *Escherichia coli* DH5α strain (NEB 5-alpha competent *E. coli*) (New England BioLabs, Ipswich, MA, USA) was transformed with the constructed plasmid. The heat-shock plasmid-containing bacteria were subsequently screened based on the expression of the ampicillin-resistant gene. Selected bacteria colonies were then processed for plasmid purification using QIAprep spin miniprep kits (50) (QIAGEN, cat # 27104). Plasmid DNA was quantified by nanodrop (Thermofisher, platform release 1.4.177, software version 1.4.0.177) and validated using Sanger sequencing.

Transfection and clonal isolation

BEAS-2B cells were seeded in 12-well plates at the concentration of 5 x10^4 /ml and 10 x10^4/ml cells, respectively, with complete growth media 24 hours before transfection. The cells were then transfected with the *ARNT2*-targeting CRISPR-Cas9 constructed plasmid using lipofectamine 3000 reagents and Opti-MEM I medium. After 4 hours of incubation, the transfecting medium was removed and replaced with complete growth media at 37^{o}C for 24 hours before sorting. The individual EGFP$^+$cell per well was sorted into a 96-well plate with a

FACS AriaII flow cytometer. When cells reached confluence, the EGFP+ single-cell clones were transferred to a larger plate for genotyping and characterization (supplementary figure 2).

Sequencing and analysis

The isolated DNA from CRISPR-Cas9 targeted BEAS-2B clones were then purified using a QIAamp DNA mini kit (QIAGEN, Cat No. ID: 51304). Purified DNA was then quantified, following the PCR amplification with *ARNT2*_F (5'-AGCAAAGCAGAGACAGCAAGT-3') and *ARNT2*_R (5'-ATACAGTGGACAAGAGGGTGT-3') primers using Hi-fidelity PCR master mix (Thermofisher, AB0794B). Amplified PCR product bands were run in 2% agarose gel and visualized with gel doc (Syngene Ingenius 3 Invitrogen). Followed by PCR product purification using a MiniElute Reaction cleanup kit (50) (QIAGEN cat #28204) and Sanger sequenced with a forward primer (5'- TTTGCACAGCCACACTTGTC -3') to confirm the knockouts. Clonal allelic analysis of individual BEAS-2B clones was performed using TIDE (261, 262) and ICE (263, 264).

cDNA synthesis and qRT-PCR

RNA was isolated from WT and KO cells exposed to CSE using TRIzol Reagent and was reverse transcribed to cDNA. The gene expression was then quantified by normalizing to *GAPDH*. The list of primer sequences used is in Table S1.

Cell viability and apoptotic study

To investigate the cell viability upon CS exposure, 10,000 cells were seeded in 96-well plates. Seeded cells were exposed with 1-,2.5-,5-,10-, and 100% CSE for about 24 hours. MTT assay was then performed to measure the cell viability. For this, 10ul of yellow tetrazolium salt (MTT) solution (5mg/ml) was added to each well. The cells were incubated for about 3-4 hours

at 37ºC. MTT was removed, and the precipitants were solubilized in 100ul DMSO (Dimethyl sulfoxide). Measurements were taken at 570nm through a microplate reader.

For annexin V-FITC apoptotic assay, 1x10^6 cells were cultured in 6-well plates. Cells were starved overnight and exposed to 5% and 7.5% CSE for 24 hours. The supernatant was collected, and cells were washed with 1xPBS twice before trypsinization. After neutralizing the trypsin, cells were collected along with the supernatant collected previously. Cells were washed twice with 1xPBS and suspended in 1X binding buffer at a concentration of ~2 x 10^6 cells/ml. For analyzing the cells, 2x10^5 cells were incubated with 2.5ul of Annexin V-FITC (BD Biosciences, 556420) and 2.5 µl of 7-AAD (7-Aminoactinomycin D) viability staining solution. *Cells were gently* vortexed and incubated at room temperature for 15-20 minutes in the dark. Subsequently, cells were analyzed by LSR Fortessa X20 flow cytometry. For 10x annexin V binding buffer, 0.1M HEPES, 1.4M NaCl, and 25mM $CaCl_2$ were dissolved in 10ml miliQ H_2O and filtered with a 0.22um filter.

DNAse-treatment and pooling strategy for RNA-Seq

Amplification Grade DNAse I (Deoxyribonuclease I) (Sigma-Aldrich) is an endonuclease used to eliminate the DNA from RNA preparations. DNase I reaction buffer was used to remove DNA from RNA samples at room temperature for 15 minutes. Stop solution was then added to inactivate the DNase I. RNA was diluted in nuclease-free water and quantification was performed using RNA Nano 6000 Assay Kit and Agilent 2100 Bioanalyzer and NanoDropTM, ND1000 spectrophotometer. The ratios of absorbance estimated the quality of total RNA at 260 to 280 nm and 260 to 230 nm. Total RNA from all 3 biological replicates of each experimental group was pooled for sequencing. For example, (WT_control =WT1_control + WT2_control + WT3_control; WT_CSE = WT1_CSE + WT2_CSE+ WT3_CSE; KO_control

= KO1_contol + KO2_control + KO3_control; KO_CSE = KO1_CSE + KO2_CSE+ KO3_CSE).

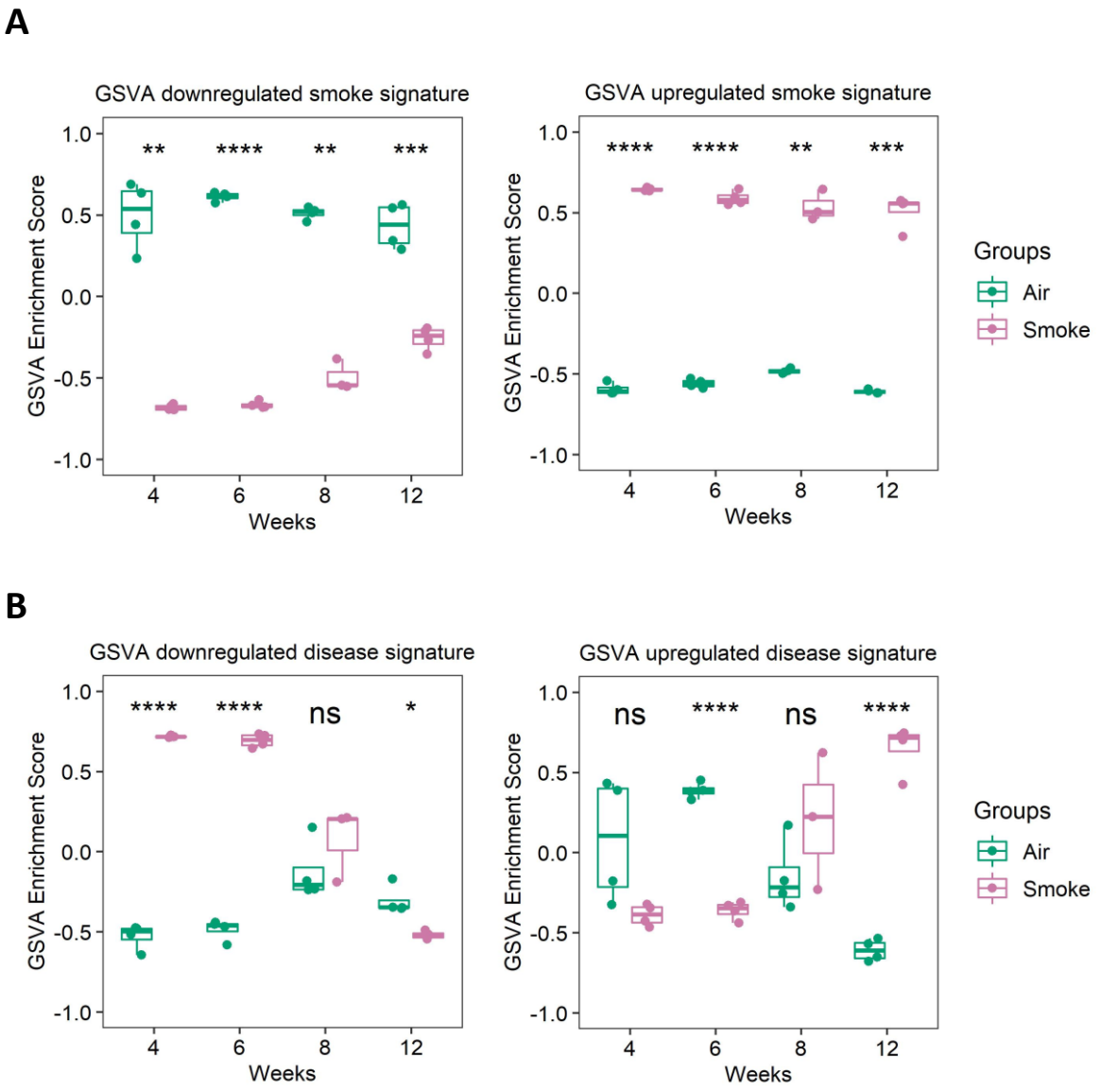

Figure S2.1: Gene set variation analysis of smoke-affected and disease-related genes through GSVA: ***A)*** *Enrichment score (ES) for genes affected during initial smoke exposure and* ***B)*** *Enrichment score (ES) for genes affected during longitudinal smoke exposure at each timepoint of 4,6,8 and 12 weeks. *p <0.05, **p <0.01, ***p < 0.001, ****p < 0.0001.*

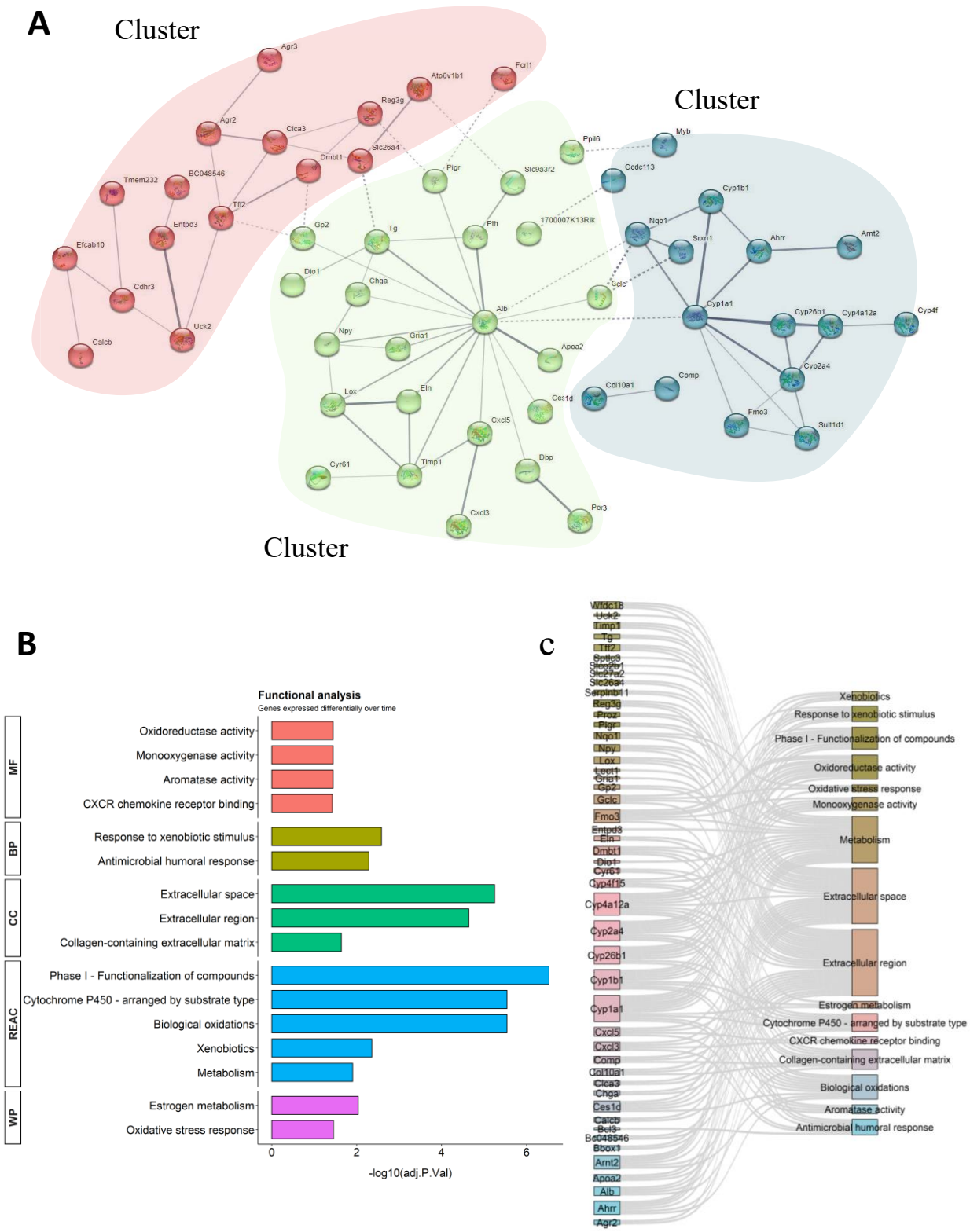

Figure S2.2: STRING interaction network and functional analysis of dysregulated genes identified during longitudinal analysis in CS-induced COPD mouse model: ***A)*** *Protein-protein interaction network was divided into 3 k-means clusters. Nodes represent protein-coding genes and lines represent linkage between edges. Nodes without any connections were removed from the*

interaction network. Isolated nodes were removed from the interaction network. ***B)*** *The bar plot depicts the functional enrichment gene ontology (GO) terms with the categories: molecular function (MF), cellular component (CC), biological process (BP), and associated critical pathways retrieved from Reactome, and WikiPathways (WP).* ***C)*** *The Sanky plot highlights the DEGs enriched in identified pathways and GO terms with FDR < 0.05.*

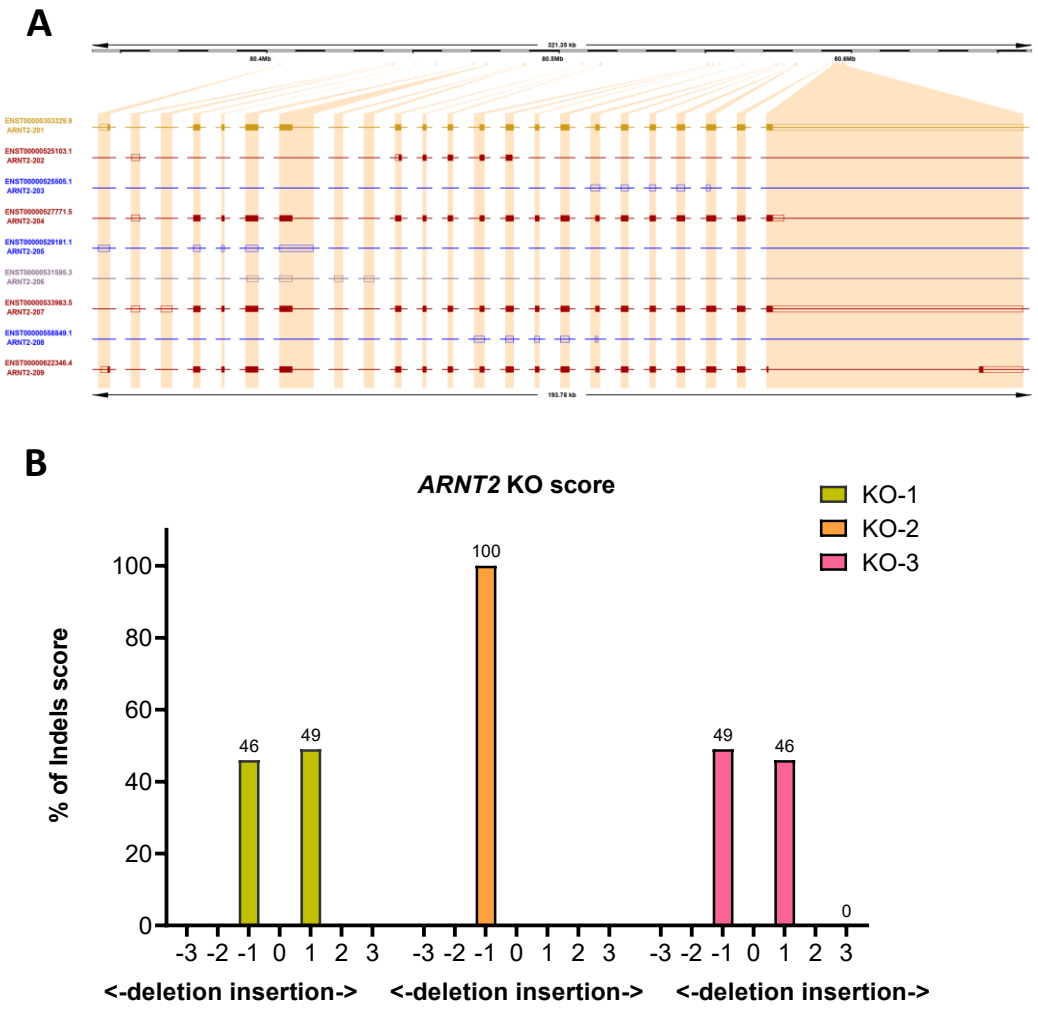

Figure S2.3: Human *ARNT2* isoforms extracted from ENSEMBL to ensure complete ablation of the gene: ***A)*** *Yellow and red lines represent the protein-coding isoform. The topmost is given with the RefSeq identifier NM_014862.4, While blue lines represent retained biotype and the purple line represents the lncRNA.* ***B)*** *Bar plot indicates % of sequences with generated alteration in each knockout.*

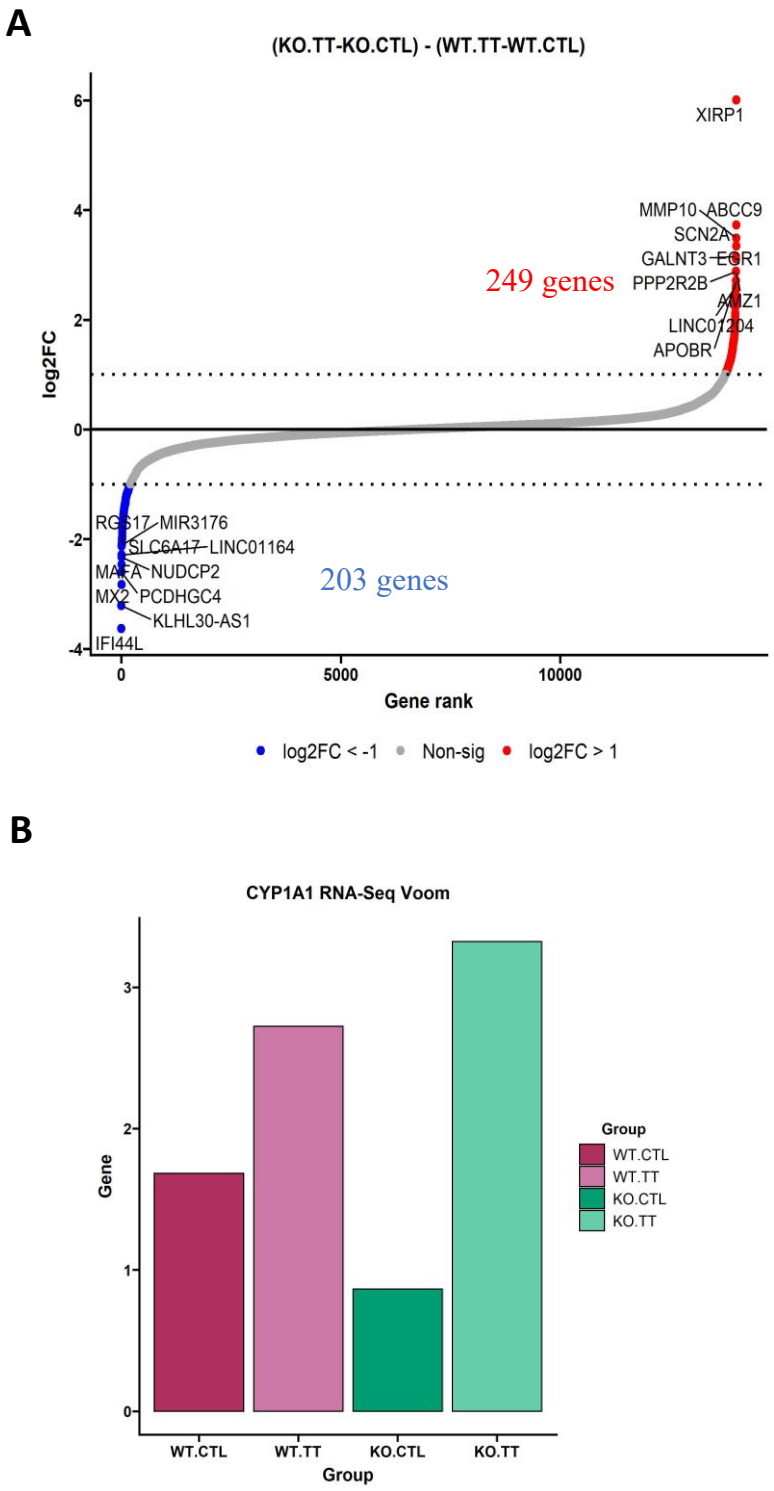

Figure S2.4: Distribution of *observed genes* during RNA-sequencing of ARNT2 KO CSE exposed cells. ***A)*** *The total number of genes with logFC > |1|.* ***B)*** *Distribution of normalized counts of CYP1A1 observed during RNA-sequencing of ARNT2 KO CSE exposed cells*

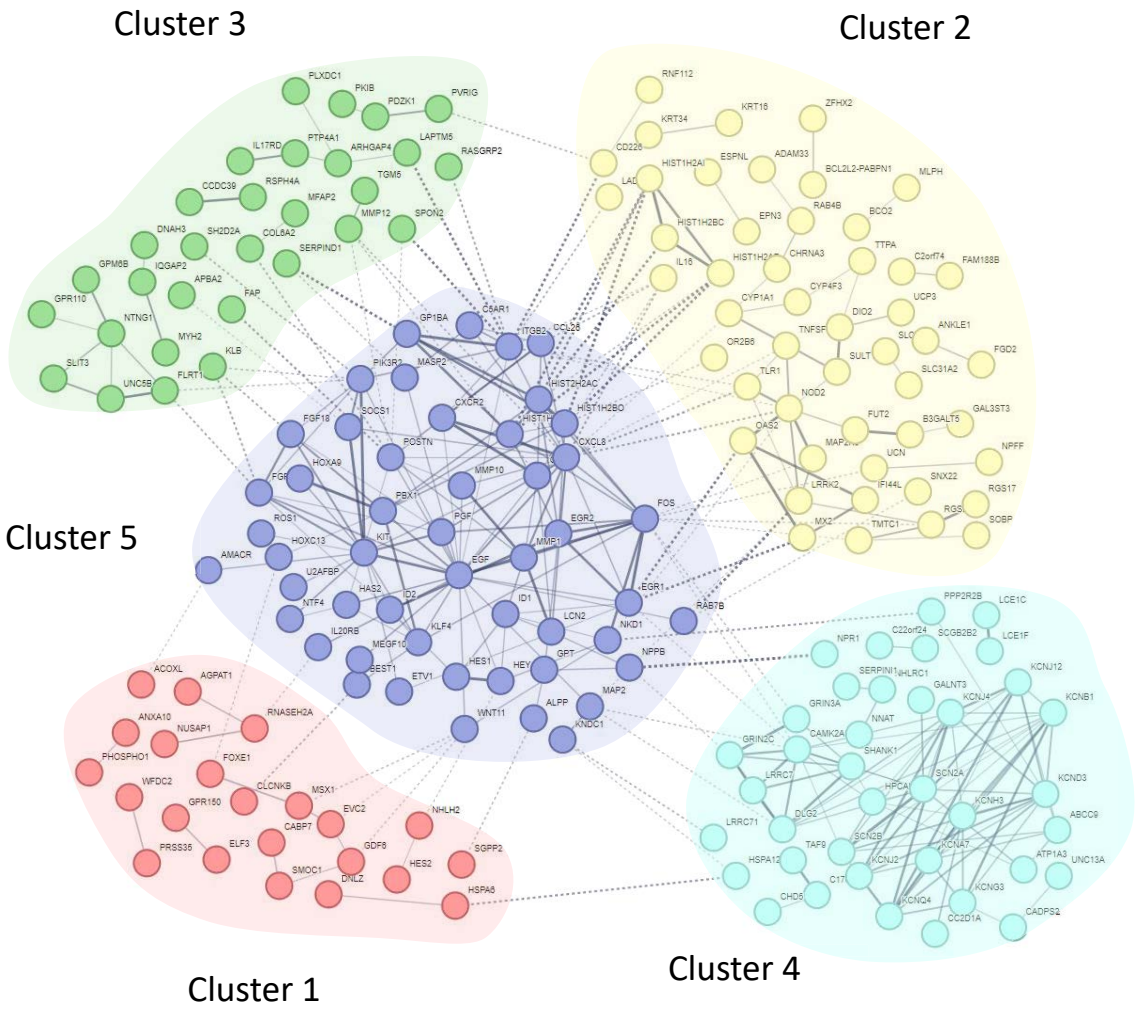

Figure S2.5: Using STRING, the cluster analysis of identified dysregulated genes from RNA sequencing of *ARNT2* KO CSE exposed cells: *The identified clusters are colored differently. The solid line indicates the interaction within the same cluster, while the dotted line indicates a connection within a different cluster.*

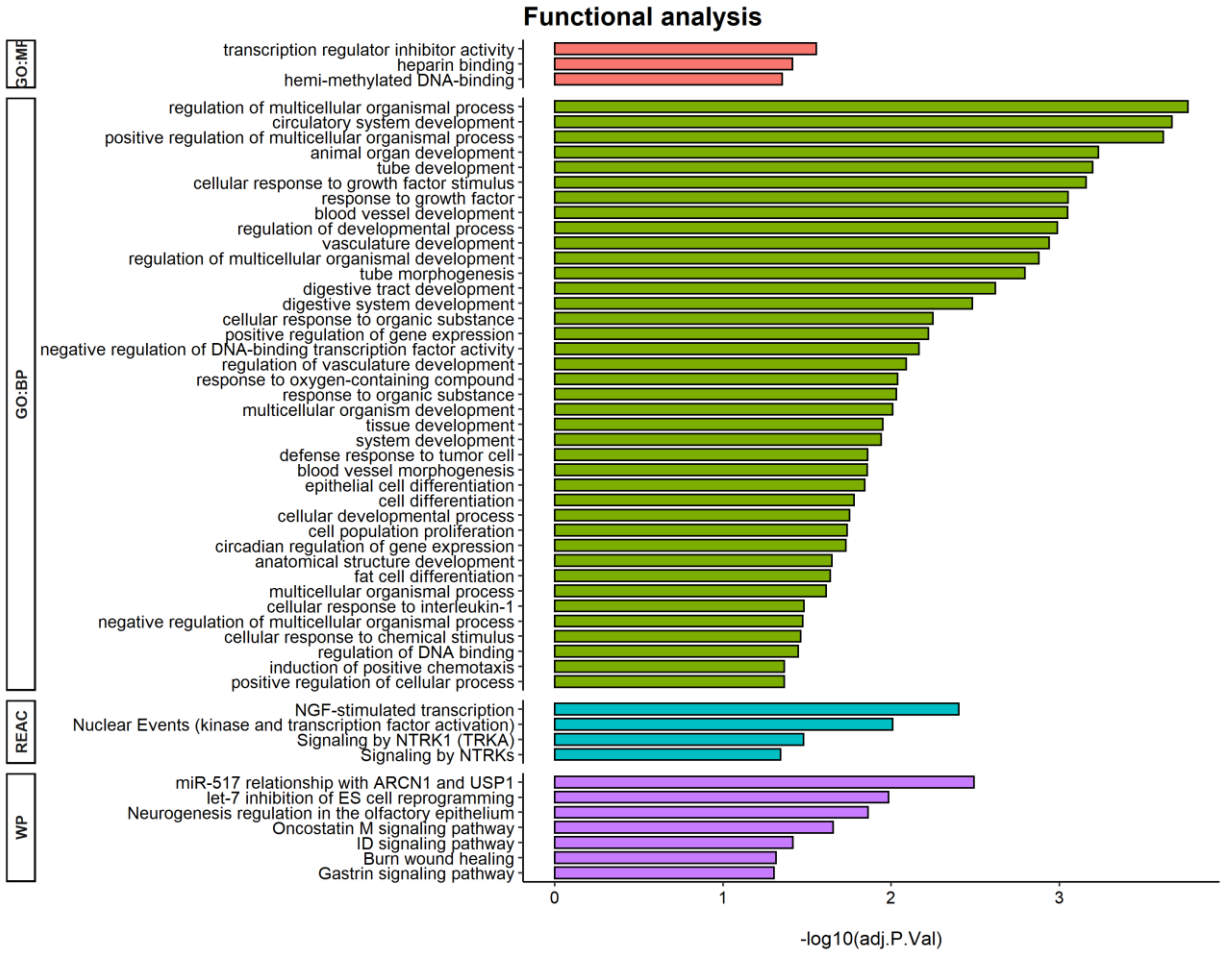

Figure S2.6: Functional enrichment analysis potential candidates from *ARNT2* KO CSE exposed cells through g:GOSt module of g:profiler: *The bar plot depicts the functional enrichment gene ontology (GO) terms with the categories: molecular function (MF), biological process (BP), and associated critical pathways retrieved from Reactome, KEGG, and WikiPathways (WP). The y-axis shows the source of the functional category, and the x-axis represents the enrichment significance in terms of -log$_{10}$(adj.p).*

Chapter 3: Longitudinal evaluation of whole lung miRNAs and their interactions with gene expression in experimental COPD

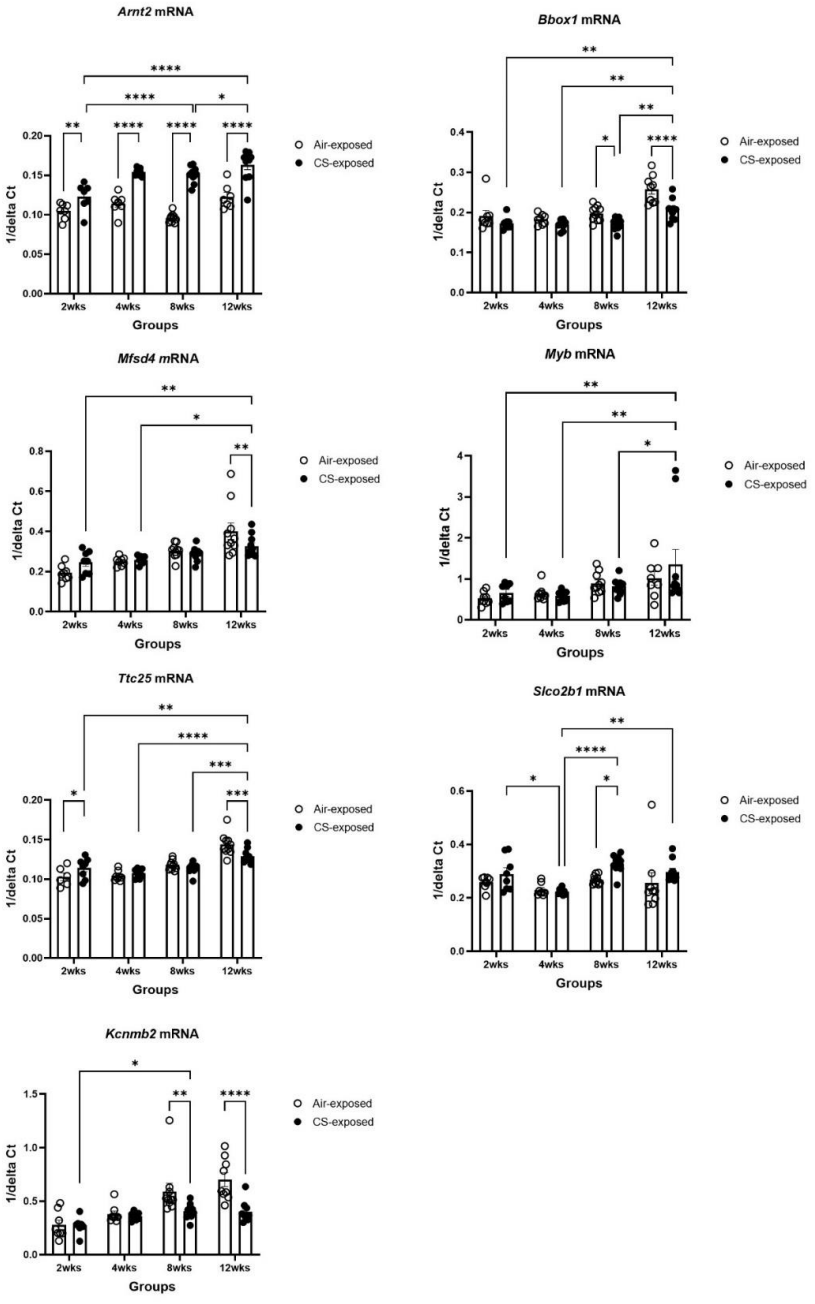

Figure S3.1: The relative expression of genes in whole lungs of C57BL/6 mice: *Expression of genes were normalized to the Hprt dis represented in terms of 1/delta Ct values on the y-axis. Data is shown in terms of means ± standard error mean (SEM). *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001.*

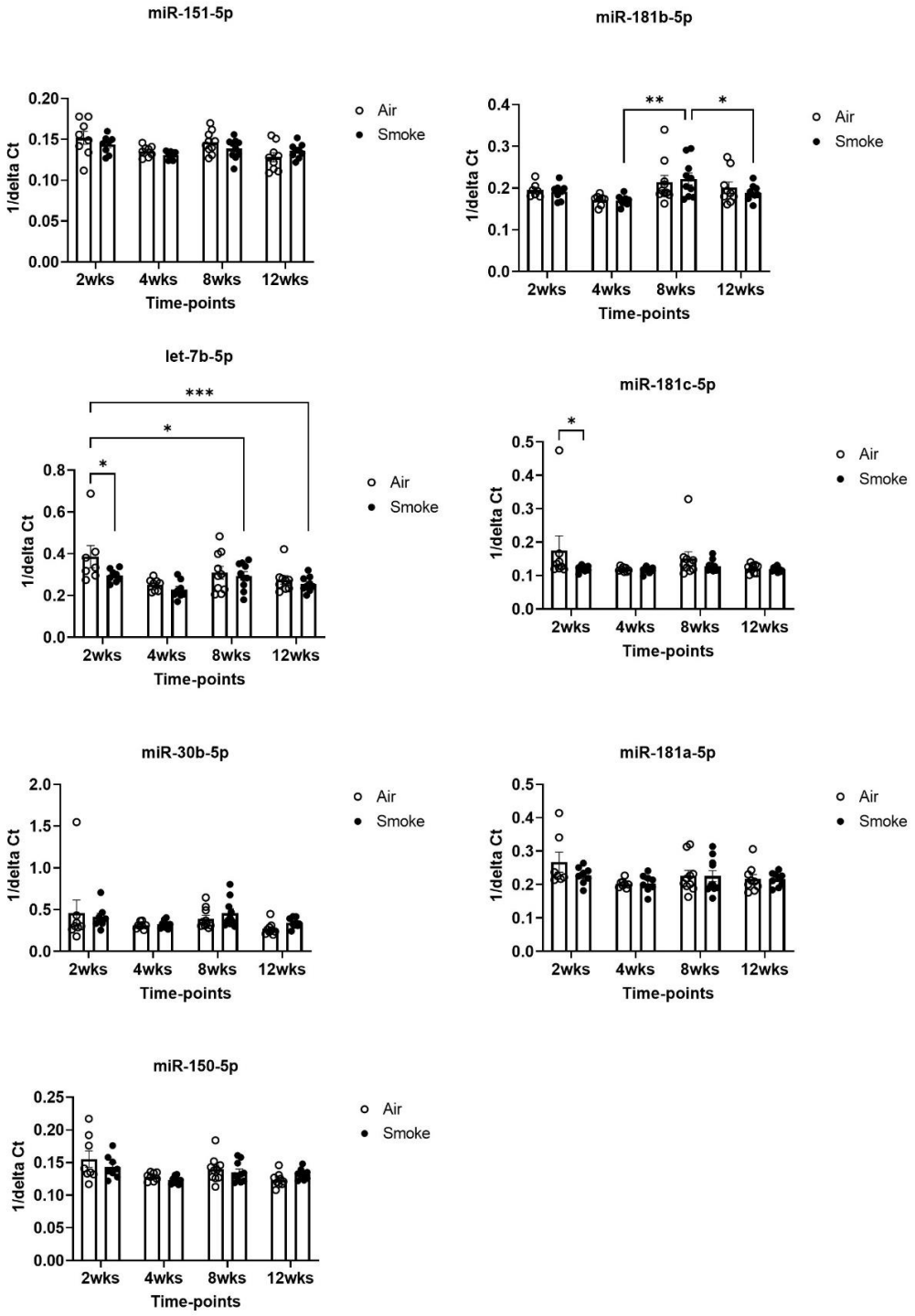

Figure S3.2: The relative expression of miRNAs in whole lungs of C57BL/6 mice: *Normalization of the miRNAs was performed to the geomteric mean of controls including snoRNA U49 and snRNA U6. Expression is represented in terms of 1/delta Ct values on the y-axis. Data is shown in terms of means ± standard error mean (SEM). *p < 0.05, **p < 0.01, ***p < 0.001.*

Chapter 4: Genome-wide CRISPR-Cas9 Knock-out library screening to identify protective factors against cigarette smoke

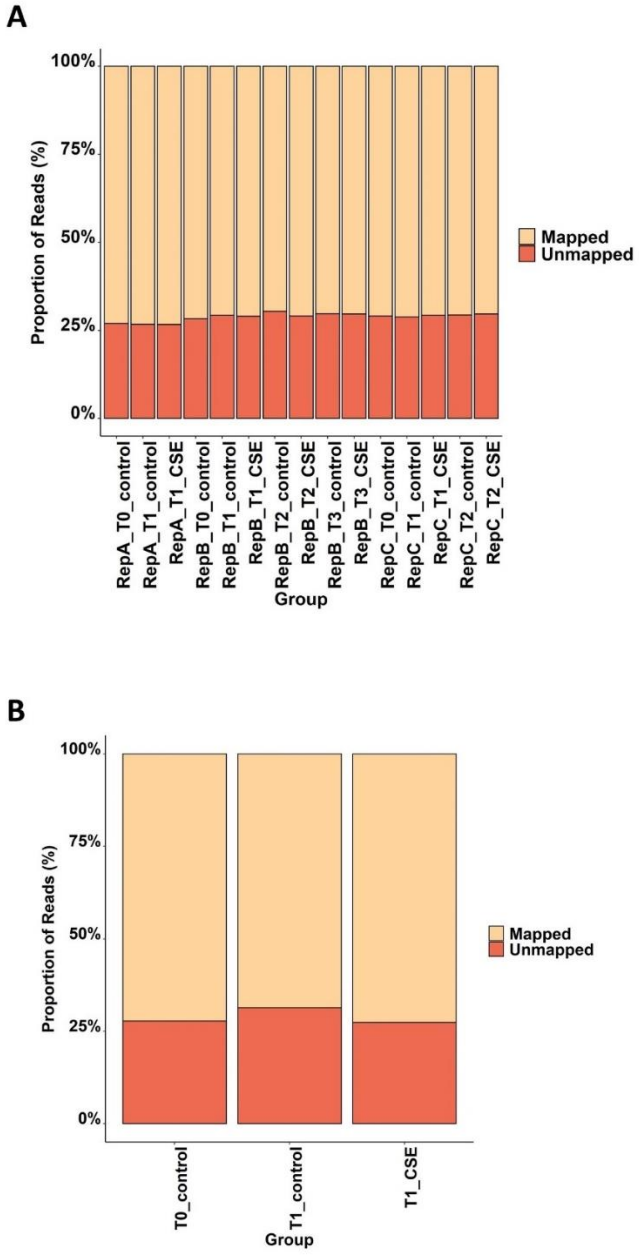

Figure S4.1: Percentage of mapped and unmapped reads after alignment of guides sequences to the TKOv3 reference library: ***A)*** *A549 and* ***B)*** *BCi NS1.1.*

Supplementary tables

Table S2.1: Primers used for quantitative PCR.

Targets	Species	Forward Sequence (For)	Reverse Sequence (Rev)	Annealing Temperature	Design
GAPDH	*Homo Sapiens*	ACAGTCAGCCGCATCTTCTT	GACCAAATCCGTTGACTCCGA	55-65	Custom
CYP1A1	*Homo Sapiens*	GATCAACCATGACCAGAAGCTAT	ACCTTGTCGATAGCACCATCA	60.5	Custom
ID1	*Homo Sapiens*	ACAGTCAGCCGCATCTTCTT	AAATCTGAGAAGCACCAAAC	56	Predesign-KiCqStart
ID2	*Homo Sapiens*	ATATCAGCATCCTGTCCTTG	CCGTGAATTTGTTGTTGTTG	59	Predesign-KiCqStart
HES1	*Homo Sapiens*	GCCTATTATGGAGAAAAGACG	CTATCTTTCTTCAGAGCATCC	57	Predesign-KiCqStart
KCNJ12	*Homo Sapiens*	CCACTACTAATTCAGCCTTG	GGCTCCTCTTGAGTTCTATC	57	Predesign-KiCqStart
SMOC1	*Homo Sapiens*	ATACTTACACTGGGTACTGC	TGAACCTGAACATACAGGAG	59	Predesign-KiCqStart
RGS2	*Homo Sapiens*	AATATGGTCTTGCTGCATTC	TTTTCCTTGCTTTTGAGGAC	56	Predesign-KiCqStart
IL11RA	*Homo Sapiens*	CTACAGGAAGAAGACAGTCC	CACCTCAGTCACATTAATCC	57	Predesign-KiCqStart

Table S2.2: Differentially expressed genes identified during initial cigarette smoke (CS) exposure of 4 weeks in mice.

Ids	logFC	AveExpr	t	P.Value	adj.P.Val	B	Legend	Symbol
1422217_PM_a_at	7.42970147	7.13007239	33.4655537	1.69332E-24	7.64383E-20	42.3836536	upregulated	Cyp1a1
1419725_PM_at	4.84681647	6.200532033	29.9421117	3.75266E-23	8.46994E-19	40.1211287	upregulated	Slc26a4
1450652_PM_at	2.71584868	10.24919876	27.6278429	3.48203E-22	4.08533E-18	38.4092352	upregulated	Ctsk
1419728_PM_at	5.37705812	5.864903083	27.5889794	3.62007E-22	4.08533E-18	38.3787642	upregulated	Cxcl5
1448303_PM_at	3.39962112	8.393182139	26.2786182	1.38461E-21	1.25005E-17	37.3153589	upregulated	Gpnmb
1457644_PM_s_at	3.63251073	8.008983735	25.4403387	3.37511E-21	2.53926E-17	36.596568	upregulated	Cxcl1
1449227_PM_at	3.3524003	8.482928118	24.6349697	8.15163E-21	4.69738E-17	35.875793	upregulated	Ch25h
1449153_PM_at	4.15390122	8.163479859	24.6160575	8.32481E-21	4.69738E-17	35.8584981	upregulated	Mmp12
1425151_PM_a_at	2.38932839	7.009753761	23.7850101	2.12816E-20	1.06741E-16	35.0811466	upregulated	Noxo1
1447845_PM_s_at	3.09789488	8.04108469	22.4422941	1.03442E-19	4.24499E-16	33.749634	upregulated	Vnn1
1419128_PM_at	1.88977437	8.723767285	20.193369	1.78503E-18	6.71485E-15	31.2870523	upregulated	Itgax
1416613_PM_at	3.08964459	4.782394307	20.0135641	2.26832E-18	7.87649E-15	31.0764021	upregulated	Cyp1b1
1417936_PM_at	2.35199027	9.928293975	19.955943	2.45033E-18	7.90074E-15	31.0084379	upregulated	Ccl9
1423933_PM_a_at	1.33455358	10.9352456	19.6059591	3.93229E-18	1.18338E-14	30.5907822	upregulated	1600029D21Rik
1417268_PM_at	2.23502239	9.410343103	18.9297786	1.00155E-17	2.82569E-14	29.7596128	upregulated	Cd14
1423627_PM_at	2.7130704	6.905959617	18.2729559	2.55364E-17	6.78081E-14	28.9203	upregulated	Nqo1
1455660_PM_at	1.58638824	8.81266793	18.0994499	3.28555E-17	8.23962E-14	28.6931252	upregulated	Csf2rb
1450826_PM_a_at	4.84204193	8.921543241	17.8848039	4.50031E-17	1.0692E-13	28.4088393	upregulated	Saa3
1419764_PM_at	1.46123809	11.76838383	17.7017344	5.90025E-17	1.33172E-13	28.163496	upregulated	Chi3l3
1443163_PM_at	2.51271418	6.259257906	16.749997	2.50764E-16	5.14534E-13	26.8437641	upregulated	Slc39a2
1417898_PM_a_at	-1.82551868	10.23935917	-16.5163288	3.61448E-16	7.09396E-13	26.5079827	downregulated	Gzma
1427221_PM_at	2.53055057	5.842211978	16.448119	4.02477E-16	7.57009E-13	26.4090613	upregulated	Slc6a20a
1434798_PM_at	1.8402108	9.816649785	16.1955472	6.01219E-16	1.08558E-12	26.0391599	upregulated	Atp6v0d2
1426808_PM_at	1.05231866	12.14566349	15.9754316	8.56511E-16	1.48707E-12	25.7120903	upregulated	Lgals3
1449360_PM_at	1.60152287	7.451047286	15.8710597	1.0144E-15	1.69597E-12	25.5554503	upregulated	Csf2rb2
1427994_PM_at	1.49552384	7.740639766	15.8154985	1.11041E-15	1.79018E-12	25.4716529	upregulated	Cd300lf

1434484_PM_at	2.69020164	9.0730259	15.7631842	1.20938E-15	1.88251E-12	25.3924893	upregulated	1100001G20Rik
1427313_PM_at	2.1367897	6.211523508	15.6982121	1.34509E-15	2.02396E-12	25.2938148	upregulated	Ptgir
1427747_PM_a_at	2.36882294	10.82723767	15.4455795	2.04085E-15	2.8652E-12	24.9063421	upregulated	Lcn2
1450060_PM_at	2.39864894	9.324321933	15.4299392	2.09458E-15	2.8652E-12	24.8821537	upregulated	Pigr
1455342_PM_at	1.64763094	5.637403117	15.1839483	3.16092E-15	4.19668E-12	24.4986065	upregulated	A230083H22Rik
1420249_PM_s_at	1.6427378	11.4563247	15.1479951	3.35834E-15	4.3314E-12	24.4420543	upregulated	Ccl6
1436530_PM_at	2.56457186	9.66862613	15.0839049	3.7424E-15	4.69265E-12	24.340929	upregulated	OTTMUSG00000000971
1424296_PM_at	1.43198784	10.34302084	14.83802	5.68888E-15	6.94058E-12	23.9491732	upregulated	Gclc
1453234_PM_at	2.05287211	4.567183322	14.7818818	6.26441E-15	7.44163E-12	23.858881	upregulated	1300002K09Rik
1438148_PM_at	2.44819297	4.129499957	14.7587851	6.51832E-15	7.54471E-12	23.8216398	upregulated	Cxcl3
1418645_PM_at	3.06641109	6.35167995	14.6187609	8.3023E-15	9.14083E-12	23.5947036	upregulated	Hal
1419413_PM_at	2.04924153	7.13423444	14.4966922	1.02666E-14	1.07777E-11	23.3952283	upregulated	Ccl17
1460197_PM_a_at	2.35495613	8.945905311	14.4690633	1.07741E-14	1.10535E-11	23.3498657	upregulated	Steap4
1416121_PM_at	1.85974998	8.142041728	14.1343012	1.94436E-14	1.90805E-11	22.7938916	upregulated	Lox
1448118_PM_a_at	1.16090336	11.38138506	14.1159747	2.00884E-14	1.91983E-11	22.763113	upregulated	Ctsd
1451798_PM_at	1.85791252	6.396709254	14.1069446	2.04142E-14	1.91983E-11	22.7479342	upregulated	Il1rn
1449015_PM_at	2.43134655	11.43581375	13.8118848	3.46822E-14	3.19508E-11	22.2471406	upregulated	Retnla
1449984_PM_at	2.85282494	4.692669698	13.7760505	3.70093E-14	3.34128E-11	22.1856774	upregulated	Cxcl2
1421366_PM_at	2.09301045	5.116904051	13.5155789	5.95645E-14	5.27216E-11	21.7346635	upregulated	Clec5a
1420699_PM_at	1.23068334	11.38639312	13.4685772	6.49525E-14	5.6385E-11	21.6524758	upregulated	Clec7a
1449164_PM_at	1.96134844	10.0157016	13.4199668	7.10547E-14	6.05185E-11	21.5672146	upregulated	Cd68
1451139_PM_at	1.2121268	8.629258677	13.313273	8.6607E-14	7.23986E-11	21.3791431	upregulated	Slc39a4
1449254_PM_at	1.42244504	10.64645068	13.2378501	9.96836E-14	8.18149E-11	21.2454151	upregulated	Spp1
1420401_PM_a_at	1.74628661	6.273757338	13.0380845	1.45081E-13	1.16948E-10	20.8880758	upregulated	AB182283
1419100_PM_at	2.44557201	6.698912931	13.0118032	1.52471E-13	1.20749E-10	20.8407216	upregulated	Serpina3n
1434046_PM_at	2.11811533	4.803200573	12.8253608	2.17352E-13	1.69164E-10	20.5024748	upregulated	AA467197
1436779_PM_at	1.34689303	8.207653874	12.7229756	2.64487E-13	2.02359E-10	20.3149902	upregulated	Cybb
1457753_PM_at	1.35460859	6.812955439	12.6842343	2.8496E-13	2.12623E-10	20.2437249	upregulated	Tlr13
1416013_PM_at	1.24054513	9.209868342	12.6799509	2.87322E-13	2.12623E-10	20.2358345	upregulated	Pld3
1420499_PM_at	1.34068187	8.074488561	12.5014208	4.06008E-13	2.95329E-10	19.9050206	upregulated	Gch1

1419609_PM_at	1.46071868	7.236597512	12.493685	4.12169E-13	2.95329E-10	19.8905998	upregulated	Ccr1
1420589_PM_at	2.04211745	4.666482664	12.4522489	4.46848E-13	3.15174E-10	19.8132338	upregulated	Has3
1422053_PM_at	2.43628905	6.566700016	12.4173578	4.78371E-13	3.32218E-10	19.747927	upregulated	Inhba
1420723_PM_at	2.0036728	5.648019985	12.3770869	5.17616E-13	3.48742E-10	19.6723674	upregulated	Vnn3
1450513_PM_at	1.44744334	5.942125755	12.3449216	5.51333E-13	3.65996E-10	19.6118747	upregulated	Cd33
1417556_PM_at	-1.71107163	8.416561866	-12.2984368	6.041E-13	3.89567E-10	19.5242289	downregulated	Fabp1
1421529_PM_a_at	1.20154466	9.371776182	12.0033435	1.08529E-12	6.75187E-10	18.9616481	upregulated	Txnrd1
1434129_PM_s_at	1.69655402	7.359730872	12.0003196	1.09188E-12	6.75187E-10	18.9558275	upregulated	Lhfpl2
1451563_PM_at	-2.01277988	6.112106689	-11.9693405	1.16182E-12	7.08726E-10	18.8961299	downregulated	Emr4
1451537_PM_at	1.07855876	11.68669721	11.9592299	1.18563E-12	7.13607E-10	18.8766206	upregulated	Chi3l1
1431609_PM_a_at	1.32063069	8.816354185	11.8343481	1.52471E-12	9.05619E-10	18.6345937	upregulated	Acp5
1433864_PM_at	1.28585288	7.784234671	11.7834527	1.69017E-12	9.90857E-10	18.5353934	upregulated	Lrp12
1452483_PM_a_at	1.06674765	7.82578259	11.7113283	1.95686E-12	1.13249E-09	18.3942543	upregulated	Cd44
1455544_PM_at	2.31388343	6.549192742	11.6629361	2.15972E-12	1.23408E-09	18.299186	upregulated	Zranb3
1451680_PM_at	1.74940234	5.741105252	11.5843241	2.53654E-12	1.43127E-09	18.1441136	upregulated	Srxn1
1420330_PM_at	1.64400841	5.379966906	11.5627494	2.65133E-12	1.47625E-09	18.1014164	upregulated	Clec4e
1438796_PM_at	1.67861625	5.192128251	11.557211	2.68165E-12	1.47625E-09	18.0904462	upregulated	Nr4a3
1448894_PM_at	1.44801587	8.921393563	11.4806088	3.13956E-12	1.68718E-09	17.9383103	upregulated	Akr1b8
1434645_PM_at	1.47960289	5.175795504	11.4293488	3.49022E-12	1.85355E-09	17.8360835	upregulated	C530008M17Rik
1449203_PM_at	2.01463075	6.710216673	11.3650408	3.98779E-12	2.09317E-09	17.7073548	upregulated	Slco1a5
1422029_PM_at	2.4020698	6.09028947	11.3530354	4.08847E-12	2.12135E-09	17.6832635	upregulated	Ccl20
1449896_PM_at	1.36619658	6.161410124	11.2857291	4.70337E-12	2.38556E-09	17.547853	upregulated	Mlph
1422062_PM_at	1.81494258	6.529526607	11.2743331	4.81654E-12	2.41582E-09	17.5248676	upregulated	Msr1
1449498_PM_at	2.62333114	6.88761495	11.130122	6.51614E-12	3.19723E-09	17.2325296	upregulated	Marco
1449305_PM_at	1.49681987	6.228020169	11.1049889	6.87032E-12	3.33476E-09	17.1813018	upregulated	F10
1460245_PM_at	-1.61887044	7.047769138	-10.9077109	1.04367E-11	4.90753E-09	16.7762986	downregulated	Klrd1
1418126_PM_at	-1.3772391	9.512957441	-10.8899393	1.08398E-11	4.99307E-09	16.7395607	downregulated	Ccl5
1455490_PM_at	2.07624658	9.98677874	10.891651	1.08003E-11	4.99307E-09	16.7431011	upregulated	NA
1450678_PM_at	1.22761424	10.53074915	10.8248495	1.2458E-11	5.68049E-09	16.6046461	upregulated	Itgb2
1418936_PM_at	1.13127902	8.128131606	10.8095039	1.28745E-11	5.81166E-09	16.572756	upregulated	Maff

1435313_PM_at	1.63316966	8.092610204	10.7431689	1.48457E-11	6.5701E-09	16.4345413	upregulated	Cd200r2
1438855_PM_x_at	1.3124914	8.902882217	10.6905491	1.66283E-11	7.21749E-09	16.3244833	upregulated	Tnfaip2
1431808_PM_a_at	1.4527962	8.657285489	10.6510643	1.81092E-11	7.71198E-09	16.2416535	upregulated	Itih4
1422010_PM_at	1.51335942	6.5934027	10.5713336	2.15266E-11	8.99754E-09	16.0737564	upregulated	Tlr7
1417262_PM_at	1.5464252	8.513109633	10.5484824	2.26233E-11	9.36918E-09	16.0254778	upregulated	Ptgs2
1450430_PM_at	1.01475587	10.25136395	10.4493523	2.8087E-11	1.14223E-08	15.815222	upregulated	Mrc1
1419004_PM_s_at	1.01805366	10.12655892	10.3954628	3.16084E-11	1.27396E-08	15.7003616	upregulated	Bcl2a1a
1421679_PM_a_at	1.13729051	7.825888254	10.3797776	3.27162E-11	1.30694E-08	15.6668556	upregulated	Cdkn1a
1431320_PM_a_at	1.30175228	6.475784039	10.3713211	3.33299E-11	1.31978E-08	15.6487772	upregulated	Myo5a
1419721_PM_at	1.48208611	7.604077547	10.3543127	3.46003E-11	1.35817E-08	15.6123873	upregulated	Gpr109a
1417870_PM_x_at	1.01911693	10.58959907	10.2905672	3.98207E-11	1.52335E-08	15.4756499	upregulated	Ctsz
1421792_PM_s_at	2.06605281	5.985626278	10.1505681	5.43163E-11	2.04614E-08	15.1733902	upregulated	Trem2
1423596_PM_at	1.11114947	7.872543939	10.0898639	6.21909E-11	2.32013E-08	15.0414917	upregulated	Nek6
1420394_PM_s_at	1.62083068	10.11183593	10.0759486	6.41555E-11	2.37381E-08	15.0111847	upregulated	Gp49a
1419321_PM_at	1.66980287	8.390805762	10.0494053	6.80813E-11	2.49858E-08	14.9533007	upregulated	F7
1437250_PM_at	1.68492796	7.486567144	10.0078732	7.47251E-11	2.7203E-08	14.8625347	upregulated	Mreg
1420438_PM_at	1.9329381	4.37231702	9.9959303	7.67561E-11	2.77188E-08	14.83639	upregulated	Orm2
1419534_PM_at	1.8848199	6.628560035	9.98396993	7.88468E-11	2.82478E-08	14.8101874	upregulated	Olr1
1424214_PM_at	1.55379496	6.045322019	9.93917063	8.72112E-11	3.09984E-08	14.7118655	upregulated	9130213B05Rik
1426008_PM_a_at	1.518119	5.53688861	9.93455983	8.81222E-11	3.10775E-08	14.7017303	upregulated	Slc7a2
1435495_PM_at	1.37444474	4.616505716	9.83774865	1.09678E-10	3.80845E-08	14.4882451	upregulated	Adora1
1451054_PM_at	3.24032228	5.583635007	9.82911453	1.11846E-10	3.85408E-08	14.4691422	upregulated	Orm1
1420804_PM_s_at	1.71668696	8.045260598	9.76684082	1.28843E-10	4.37301E-08	14.3310553	upregulated	Clec4d
1435644_PM_at	1.14653964	7.271745415	9.73492279	1.38559E-10	4.63312E-08	14.2600706	upregulated	Sh3pxd2b
1419537_PM_at	1.32926895	7.360549292	9.66042791	1.64269E-10	5.45239E-08	14.0938442	upregulated	Tcfec
1417314_PM_at	2.13532699	8.194614484	9.63023178	1.76038E-10	5.8004E-08	14.0262448	upregulated	Cfb
1418133_PM_at	1.32470286	6.298326137	9.53769272	2.17784E-10	7.02214E-08	13.8182871	upregulated	Bcl3
1423944_PM_at	1.17059932	5.204601796	9.46689637	2.56483E-10	8.09645E-08	13.6583827	upregulated	Hpx
1437540_PM_at	1.43433304	6.736549807	9.45446847	2.63972E-10	8.27498E-08	13.6302401	upregulated	Mcoln3
1420413_PM_at	1.27498024	4.087341376	9.43922964	2.73461E-10	8.51333E-08	13.5957028	upregulated	Slc7a11

1416646_PM_at	1.23175618	5.268351772	9.41507495	2.89224E-10	8.82154E-08	13.5408919	upregulated	Afp
1419627_PM_s_at	1.41785554	9.996851588	9.37799892	3.15252E-10	9.4872E-08	13.4566016	upregulated	Clec4n
1424754_PM_at	1.41963989	6.791399363	9.30787205	3.71241E-10	1.08819E-07	13.2966459	upregulated	Ms4a7
1418847_PM_at	1.14946967	7.686123954	9.28858752	3.88355E-10	1.12377E-07	13.2525382	upregulated	Arg2
1426037_PM_a_at	1.61040394	4.681625557	9.17824267	5.03077E-10	1.41396E-07	12.9991553	upregulated	Rgs16
1422089_PM_at	-1.96496789	6.535188187	-9.10446969	5.98652E-10	1.63933E-07	12.8288007	downregulated	Ncr1
1419532_PM_at	1.54110509	5.781365268	9.07821186	6.36995E-10	1.72055E-07	12.7679831	upregulated	Il1r2
1424733_PM_at	1.12606292	7.616918086	9.07067814	6.48453E-10	1.73206E-07	12.7505158	upregulated	P2ry14
1417990_PM_at	1.31902655	5.270706289	9.04277374	6.92766E-10	1.83954E-07	12.6857492	upregulated	Ppp1r14d
1449328_PM_at	1.10251473	7.902888931	8.92692809	9.12524E-10	2.36169E-07	12.415705	upregulated	Ly75
1426316_PM_at	1.01182804	6.008772636	8.89503008	9.84753E-10	2.52572E-07	12.3410192	upregulated	6330416G13Rik
1424404_PM_at	1.04823349	6.991715542	8.84716666	1.10428E-09	2.81628E-07	12.2286857	upregulated	0610040J01Rik
1417925_PM_at	1.45309855	6.018885782	8.78780671	1.27339E-09	3.1758E-07	12.0889265	upregulated	Ccl22
1420671_PM_x_at	-1.28694347	7.111156011	-8.77518066	1.31265E-09	3.25573E-07	12.0591359	downregulated	Ms4a4c
1450753_PM_at	-1.39383143	7.546038325	-8.76709431	1.33845E-09	3.28972E-07	12.0400449	downregulated	Nkg7
1448025_PM_at	1.00986333	8.97487185	8.74084879	1.42579E-09	3.47902E-07	11.9780193	upregulated	668101
1434980_PM_at	1.17071105	6.866100049	8.7163514	1.51259E-09	3.6319E-07	11.9200386	upregulated	Pik3r5
1435767_PM_at	-1.00655431	10.37345302	-8.68604441	1.62749E-09	3.88711E-07	11.8481923	downregulated	Scn3b
1418248_PM_at	1.09588136	6.736925707	8.67797248	1.65957E-09	3.94287E-07	11.8290353	upregulated	Gla
1417290_PM_at	1.3603784	8.094031798	8.63019856	1.86314E-09	4.38042E-07	11.715469	upregulated	Lrg1
1442590_PM_at	1.04267966	5.502851108	8.57239568	2.14398E-09	4.98873E-07	11.5776395	upregulated	Tnfrsf22
1418752_PM_at	2.62259875	7.601892486	8.5617385	2.20031E-09	5.09355E-07	11.5521772	upregulated	Aldh3a1
1417483_PM_at	1.11169676	8.531990009	8.55740389	2.22365E-09	5.12132E-07	11.5418164	upregulated	Nfkbiz
1419561_PM_at	1.50644287	6.007965113	8.53798879	2.33134E-09	5.29978E-07	11.4953777	upregulated	Ccl3
1449279_PM_at	1.34033854	9.24745446	8.53505049	2.3481E-09	5.29978E-07	11.488345	upregulated	Gpx2
1421074_PM_at	1.27284335	6.13377562	8.52260769	2.42044E-09	5.43587E-07	11.4585508	upregulated	Cyp7b1
1455577_PM_at	1.47720421	5.739674054	8.46109682	2.81297E-09	6.25518E-07	11.3109498	upregulated	Ccl28
1422191_PM_at	1.82408973	6.58215778	8.45501312	2.85517E-09	6.3179E-07	11.2963231	upregulated	Cd200r1
1457700_PM_at	1.24087204	5.281692587	8.42235243	3.09306E-09	6.77785E-07	11.2177117	upregulated	Cyp4f39
1453851_PM_a_at	1.07801434	7.651378671	8.32674064	3.91275E-09	8.45097E-07	10.9867417	upregulated	Gadd45g

1418174_PM_at	-1.41597538	8.906292312	-8.3050944	4.12732E-09	8.87196E-07	10.9342772	downregulated	Dbp
1435513_PM_at	1.35187	5.324801348	8.28482414	4.33914E-09	9.2393E-07	10.8850897	upregulated	Htr2c
1445642_PM_at	1.32843941	4.151342791	8.22373686	5.04718E-09	1.04552E-06	10.7365183	upregulated	Lemd1
1456792_PM_at	1.24004094	5.929516816	8.22188054	5.07046E-09	1.04552E-06	10.7319956	upregulated	LOC100045317
1419057_PM_at	1.4276327	5.806322041	8.21188812	5.19764E-09	1.05688E-06	10.7076421	upregulated	Slc5a1
1433907_PM_at	-1.23762899	6.641057867	-8.20903266	5.23458E-09	1.05961E-06	10.7006803	downregulated	Pknox2
1438194_PM_at	1.23541073	4.991516861	8.20426815	5.29681E-09	1.06743E-06	10.6890617	upregulated	2900019G14Rik
1433678_PM_at	-1.03621029	6.773338859	-8.16845674	5.78942E-09	1.15637E-06	10.6016343	downregulated	Pld4
1427683_PM_at	1.04202382	6.335888488	8.12408177	6.46534E-09	1.26892E-06	10.4930603	upregulated	Egr2
1451610_PM_at	1.26708872	9.598644883	8.11736962	6.57438E-09	1.28474E-06	10.4766143	upregulated	Cxcl17
1418747_PM_at	1.0080221	7.281414626	8.11098297	6.67988E-09	1.29415E-06	10.4609602	upregulated	Sfpi1
1420796_PM_at	1.71593015	3.991777192	8.05255725	7.72896E-09	1.46594E-06	10.3175009	upregulated	Ahrr
1417896_PM_at	1.18908401	6.473064605	7.97242535	9.44757E-09	1.76228E-06	10.120002	upregulated	Tjp3
1448620_PM_at	1.16677825	9.715494099	7.95927617	9.7648E-09	1.81396E-06	10.087512	upregulated	Fcgr3
1422557_PM_s_at	1.11639501	11.0030019	7.93395176	1.04067E-08	1.89423E-06	10.0248739	upregulated	Mt1
1428942_PM_at	1.55597186	8.764018673	7.8503759	1.28475E-08	2.30922E-06	9.81755475	upregulated	Mt2
1430570_PM_at	1.32671871	4.983946789	7.79159876	1.49075E-08	2.61843E-06	9.67120297	upregulated	Kynu
1435903_PM_at	1.08982786	7.27121873	7.77505842	1.55458E-08	2.69905E-06	9.62993728	upregulated	Cd300a
1418809_PM_at	1.01531106	6.3785179	7.7697308	1.57573E-08	2.72528E-06	9.61663808	upregulated	Pira2
1449091_PM_at	-1.33993336	6.590265212	-7.75592845	1.63189E-08	2.77552E-06	9.58216655	downregulated	Cldn8
1460227_PM_at	1.37816003	6.81317828	7.75342065	1.64231E-08	2.77661E-06	9.57590063	upregulated	Timp1
1427196_PM_at	1.1273464	5.816859893	7.72754276	1.7539E-08	2.91366E-06	9.51119545	upregulated	Cntd1
1420703_PM_at	1.03473201	6.641692266	7.64358731	2.17208E-08	3.52697E-06	9.30067951	upregulated	Csf2ra
1419222_PM_at	-1.0111386	5.925992778	-7.6412206	2.18524E-08	3.53562E-06	9.29473198	downregulated	Tbxa2r
1428572_PM_at	1.16075198	9.248244342	7.62337243	2.28714E-08	3.67452E-06	9.24985659	upregulated	Basp1
1417601_PM_at	1.14124689	5.628653976	7.57093048	2.6155E-08	4.11381E-06	9.11776866	upregulated	Rgs1
1437060_PM_at	1.34635679	4.219802732	7.55706567	2.71008E-08	4.24777E-06	9.08278865	upregulated	Olfm4
1425062_PM_at	-1.41311533	5.562972259	-7.54532674	2.79288E-08	4.3327E-06	9.05315316	downregulated	Fcrl1
1424509_PM_at	2.10443364	6.223370898	7.54544377	2.79205E-08	4.3327E-06	9.05344869	upregulated	Cd177
1460273_PM_a_at	1.00758523	6.452372837	7.44731072	3.59313E-08	5.47964E-06	8.80503475	upregulated	Naip2

1435172_PM_at	-1.24026986	5.322097878	-7.44124708	3.64971E-08	5.52858E-06	8.78964603	downregulated	Eomes
1421037_PM_at	1.01725595	4.047275448	7.42022653	3.85299E-08	5.79759E-06	8.73626354	upregulated	Npas2
1443946_PM_s_at	1.5800777	9.995247246	7.40421453	4.01553E-08	5.93097E-06	8.695564	upregulated	AU040972
1423465_PM_at	1.02489746	6.524562059	7.37661781	4.31226E-08	6.27934E-06	8.62534458	upregulated	Frrs1
1418932_PM_at	1.37779458	5.131332978	7.28499948	5.46735E-08	7.83497E-06	8.39156045	upregulated	LOC100046232
1425890_PM_at	1.28280301	6.036739543	7.23675211	6.19769E-08	8.7978E-06	8.26804198	upregulated	Ly6i
1429171_PM_a_at	1.03937074	4.130867838	7.22970556	6.31237E-08	8.9325E-06	8.24997894	upregulated	Ncapg
1420421_PM_s_at	1.20365762	5.384746935	7.2030758	6.76562E-08	9.51424E-06	8.18166357	upregulated	Klrb1b
1424451_PM_at	-1.21731866	7.731954034	-7.19213585	6.9613E-08	9.72879E-06	8.15357434	downregulated	Acaa1a
1450201_PM_at	1.38102942	4.575011499	7.18353808	7.11911E-08	9.88811E-06	8.131489	upregulated	Proz
1428720_PM_s_at	-1.41452583	9.475671434	-7.14777558	7.81557E-08	1.07235E-05	8.03953232	downregulated	2010309G21Rik
1419060_PM_at	-1.03558062	6.687273573	-7.13945883	7.98723E-08	1.08733E-05	8.01812605	downregulated	Gzmb
1452367_PM_at	1.01298531	5.236236204	7.13899168	7.99699E-08	1.08733E-05	8.01692342	upregulated	Coro2a
1453418_PM_at	1.01867734	3.862528326	7.12374785	8.32206E-08	1.12139E-05	7.97766615	upregulated	Col24a1
1440837_PM_at	-1.51878616	6.710193832	-7.08540686	9.20037E-08	1.23239E-05	7.87880892	downregulated	H2-Ob
1453416_PM_at	1.17360322	5.092436384	7.04641308	1.01905E-07	1.349E-05	7.77809668	upregulated	Gas2l3
1429918_PM_at	1.08940349	4.547561992	6.98953019	1.18328E-07	1.5305E-05	7.63087374	upregulated	Arhgap20
1448239_PM_at	1.38010993	6.863690677	6.96643587	1.25741E-07	1.60795E-05	7.57099876	upregulated	Hmox1
1427429_PM_at	1.16454815	6.771259991	6.94599697	1.32695E-07	1.67786E-05	7.51795921	upregulated	Csf2
1419276_PM_at	1.20354804	5.594039671	6.92116755	1.4167E-07	1.77151E-05	7.45346465	upregulated	Enpp1
1457117_PM_at	1.27743685	4.947143923	6.91691747	1.43268E-07	1.78161E-05	7.4424183	upregulated	Nfe2l2
1429863_PM_at	1.0861717	5.624141462	6.90683118	1.47131E-07	1.82328E-05	7.41619526	upregulated	Lonrf3
1441445_PM_at	-1.09420855	7.46272023	-6.7859417	2.02582E-07	2.37526E-05	7.10105255	downregulated	Per3
1423226_PM_at	-1.47419267	8.507159269	-6.78362249	2.03832E-07	2.38373E-05	7.09499163	downregulated	Ms4a1
1440865_PM_at	-1.54750105	6.293138413	-6.75282904	2.21189E-07	2.54712E-05	7.0144649	downregulated	Ifitm6
1429444_PM_at	1.55006185	6.892365719	6.62540232	3.10518E-07	3.444E-05	6.68021627	upregulated	Rasl11a
1447918_PM_x_at	-1.56449351	7.002880657	-6.61945272	3.15488E-07	3.47352E-05	6.66457081	downregulated	Igl-V1
1425958_PM_at	1.04790401	4.534257679	6.584108	3.46722E-07	3.73792E-05	6.57155562	upregulated	Il1f9
1425099_PM_a_at	1.08575947	5.173749692	6.58385753	3.46955E-07	3.73792E-05	6.57089602	upregulated	Arntl
1421944_PM_a_at	-1.11938374	4.504006719	-6.57894263	3.51543E-07	3.77813E-05	6.55795209	downregulated	Asgr1

1451828_PM_a_at	1.29892736	6.551202938	6.54426109	3.85716E-07	4.07766E-05	6.46654931	upregulated	Acsl4
1424162_PM_at	1.16206858	5.88420489	6.53793205	3.92307E-07	4.128E-05	6.44985698	upregulated	Trim29
1416871_PM_at	1.19049224	6.438963306	6.48494863	4.52157E-07	4.64939E-05	6.30997201	upregulated	Adam8
1441899_PM_x_at	-1.38196193	5.16175878	-6.47951281	4.58799E-07	4.69629E-05	6.29560598	downregulated	Bcan
1452040_PM_a_at	1.08539327	5.349349449	6.40261347	5.64083E-07	5.63346E-05	6.09208991	upregulated	Cdca3
1419307_PM_at	-1.31337367	5.716728647	-6.36966375	6.16395E-07	6.08855E-05	6.00472911	downregulated	Tnfrsf13c
1456328_PM_at	-1.47165126	7.211448215	-6.30810629	7.27662E-07	7.07918E-05	5.84127319	downregulated	Bank1
1432032_PM_a_at	1.08958236	4.407140211	6.27158238	8.03083E-07	7.72963E-05	5.74414223	upregulated	Artn
1424367_PM_a_at	1.02596206	6.862109883	6.26881782	8.09104E-07	7.77101E-05	5.73678586	upregulated	Homer2
1460407_PM_at	-1.15282053	5.410624421	-6.21045051	9.47457E-07	8.91024E-05	5.58133202	downregulated	Spib
1429262_PM_at	1.2759698	5.833897675	6.14882216	1.11964E-06	0.000102936	5.41691003	upregulated	Rassf6
1451584_PM_at	1.02868813	6.334525708	6.14346237	1.13604E-06	0.000104231	5.40259702	upregulated	Havcr2
1459838_PM_s_at	1.14505055	4.919850134	6.14242778	1.13923E-06	0.000104312	5.39983396	upregulated	Btbd11
1435331_PM_at	-1.28310491	6.311250224	-6.13557707	1.1606E-06	0.000105626	5.381536	downregulated	LOC100048304
1424118_PM_a_at	1.00379348	6.582734764	6.12616538	1.19062E-06	0.000107923	5.35639224	upregulated	Spc25
1427398_PM_at	1.15597571	5.018701585	5.96165675	1.86258E-06	0.000161071	4.91592162	upregulated	Muc4
1418480_PM_at	-1.15821761	7.206485743	-5.95477281	1.89787E-06	0.000163808	4.89745185	downregulated	Ppbp
1456123_PM_at	-1.13580501	4.137224076	-5.94246538	1.96265E-06	0.000167796	4.86442356	downregulated	AI481121
1420380_PM_at	1.04656675	4.499847042	5.90216525	2.19082E-06	0.000184163	4.75621121	upregulated	Ccl2
1419907_PM_s_at	-1.27021964	4.558817212	-5.88236804	2.31253E-06	0.000190841	4.70301828	downregulated	Fcrla
1449456_PM_a_at	-1.28186522	5.510703408	-5.86139917	2.44889E-06	0.000198823	4.64665345	downregulated	Cma1
1424304_PM_at	1.18801841	5.020008585	5.85382387	2.50013E-06	0.000200879	4.62628497	upregulated	Tpcn2
1420579_PM_s_at	1.04829654	7.986170038	5.73990937	3.41485E-06	0.000260388	4.31963838	upregulated	Cftr
1426278_PM_at	-1.39805804	10.32056381	-5.72454839	3.5617E-06	0.000270671	4.27824082	downregulated	Ifi27
1460482_PM_at	1.18414004	4.361939588	5.71535143	3.65265E-06	0.000275726	4.25345022	upregulated	3110047P20Rik
1424046_PM_at	1.12488326	5.237100353	5.69622132	3.84939E-06	0.000288647	4.20187291	upregulated	Bub1
1417640_PM_at	-1.18940797	8.460749311	-5.68587527	3.96021E-06	0.000295942	4.1739722	downregulated	Cd79b
1431780_PM_at	1.03167875	6.000150822	5.67920005	4.03341E-06	0.000299461	4.15596848	upregulated	1700021K14Rik
1450110_PM_at	1.30673432	5.945933378	5.66103528	4.23958E-06	0.000311122	4.10696724	upregulated	Adh7
1438377_PM_x_at	1.05780056	4.685301612	5.58449432	5.23154E-06	0.000370152	3.90035514	upregulated	Slc13a3

1434500_PM_at	1.02425259	7.250195802	5.56605238	5.5036E-06	0.000385774	3.85054362	upregulated	Ttyh2
1429022_PM_at	1.10161557	4.813441584	5.50933762	6.43282E-06	0.000437985	3.69729363	upregulated	Adcyap1r1
1421816_PM_at	1.18196911	7.443811345	5.47809571	7.01051E-06	0.000471882	3.61283738	upregulated	LOC630729
1417958_PM_at	1.25817675	6.173057596	5.45411386	7.48911E-06	0.000497156	3.54799161	upregulated	Tspan1
1449393_PM_at	-1.24366655	4.844610819	-5.44192021	7.74492E-06	0.000509641	3.51501583	downregulated	LOC100046930
1454366_PM_at	1.1200333	4.71755721	5.42453475	8.1249E-06	0.000530776	3.46799451	upregulated	4833419A21Rik
1422789_PM_at	-1.22885342	7.412571399	-5.41140866	8.42415E-06	0.000544806	3.43248948	downregulated	Aldh1a2
1452934_PM_at	1.29472841	4.838920717	5.35094165	9.95219E-06	0.000621372	3.26889602	upregulated	Tmc5
1425854_PM_x_at	-1.01954663	9.00897681	-5.29743716	1.15349E-05	0.000698924	3.12410388	downregulated	LOC665506
1426817_PM_at	1.03183863	7.830573067	5.2436505	1.33808E-05	0.00079164	2.9785301	upregulated	Mki67
1451679_PM_at	-1.02995705	5.057864369	-5.2271995	1.40024E-05	0.000820889	2.93400418	downregulated	6530401D17Rik
1422286_PM_a_at	1.1044996	6.429356887	5.18623521	1.56792E-05	0.000903928	2.82313315	upregulated	Tgif1
1453568_PM_at	-1.15098409	5.721552153	-5.1519167	1.72379E-05	0.000978789	2.73025544	downregulated	Dapl1
1423147_PM_at	1.08257735	5.252451055	5.12907933	1.83603E-05	0.001030847	2.66845512	upregulated	Mat1a
1416930_PM_at	-1.85641173	5.596001638	-5.12753577	1.84387E-05	0.001033965	2.66427829	downregulated	Ly6d
1440927_PM_x_at	-1.47777634	4.240478377	-5.01865755	2.49091E-05	0.001318195	2.36974893	downregulated	Apol11b
1446748_PM_at	1.07500718	6.442781762	5.00319074	2.59965E-05	0.001366133	2.32792859	upregulated	2010007H06Rik
1422914_PM_at	1.07280988	5.516889052	4.9608557	2.92219E-05	0.001499578	2.21349237	upregulated	Sp5
1449851_PM_at	-1.19388139	8.569232014	-4.9508486	3.0041E-05	0.001532296	2.18644971	downregulated	Per1
1429889_PM_at	-1.23858602	6.725834184	-4.87543504	3.69981E-05	0.001817337	1.9827667	downregulated	Faim3
1425917_PM_at	-1.17682782	5.027501231	-4.86251083	3.83425E-05	0.001861096	1.94788181	downregulated	H28
1460302_PM_at	1.09630263	6.930798438	4.84573306	4.01607E-05	0.001930664	1.90260605	upregulated	Thbs1
1424525_PM_at	-1.00035354	7.29886183	-4.79896129	4.56962E-05	0.002122194	1.77645803	downregulated	Grp
1426170_PM_a_at	-1.14740373	5.986959593	-4.77900177	4.82842E-05	0.002209338	1.72265817	downregulated	Cd8b1
1439163_PM_at	-1.27503113	8.204757438	-4.77613043	4.86683E-05	0.002221505	1.71492032	downregulated	Zbtb16
1447806_PM_s_at	-1.18975047	6.343308238	-4.72534214	5.59891E-05	0.002488928	1.57812936	downregulated	Srpk3
1452318_PM_a_at	1.53408372	8.629648986	4.7231346	5.6331E-05	0.002500334	1.57218709	upregulated	Hspa1b
1429272_PM_a_at	-1.02834219	4.662584877	-4.64275355	7.03053E-05	0.00296603	1.35602897	downregulated	666661
1432088_PM_at	1.00153219	5.936548222	4.60034141	7.90177E-05	0.003240329	1.24215756	upregulated	Veph1
1455530_PM_at	-1.28621409	6.751805029	-4.59866871	7.93825E-05	0.003242901	1.23766938	downregulated	Ighv14-2

1435188_PM_at	-1.26971979	5.883418915	-4.59353595	8.05123E-05	0.003273709	1.22389845	downregulated	Gm129
1442023_PM_at	-1.05152244	4.551840533	-4.49515331	0.000105535	0.003993254	0.96036651	downregulated	A530030E21Rik
1448201_PM_at	-1.28117216	6.680598462	-4.46339219	0.000115157	0.004257413	0.87547526	downregulated	Sfrp2
1449191_PM_at	-1.0783329	4.687409058	-4.35440156	0.000155277	0.005416822	0.58494542	downregulated	Wfdc12
1419684_PM_at	1.20817658	7.162028384	4.35021431	0.000157068	0.005466619	0.57380963	upregulated	Ccl8
1452260_PM_at	-1.38368177	5.777078475	-4.28544378	0.000187518	0.006196744	0.4018206	downregulated	Cidec
1416464_PM_at	-1.11059932	7.078885829	-4.11757285	0.000296295	0.008793594	-0.04136228	downregulated	Slc4a1
1418608_PM_at	-1.50771987	5.834907302	-4.08098185	0.000327237	0.00946305	-0.13741279	downregulated	Calml3
1429106_PM_at	1.03351997	7.405978391	3.949171	0.000467389	0.012359947	-0.48155799	upregulated	4921509J17Rik
1431530_PM_a_at	1.04064458	5.829962885	3.61267205	0.001147814	0.023262192	-1.34442453	upregulated	Tspan5
1422230_PM_s_at	-1.01333563	10.13092976	-3.45455147	0.001738297	0.030929621	-1.7404344	downregulated	Cyp2a4
1426464_PM_at	-1.12058994	5.65430695	-3.39427084	0.002033413	0.034237333	-1.8895523	downregulated	Nr1d1
1450790_PM_at	-3.10699584	5.053419207	-3.38334097	0.00209187	0.034839502	-1.91647352	downregulated	Tg
1450281_PM_a_at	-1.14000919	3.8059619	-3.30926922	0.00253295	0.0391441	-2.0979423	downregulated	1700021K02Rik
1417976_PM_at	-1.74214211	6.075613863	-3.28389731	0.002703669	0.040738723	-2.15969899	downregulated	Ada
1416794_PM_at	1.03592224	6.136819883	3.25732624	0.002894295	0.042696529	-2.22414764	upregulated	Arl6ip2

Table S2.3: Differentially expressed genes identified during longitudinal cigarette smoke (CS) exposure in mice from 4 to 12 weeks.

Ids	logFC	AveExpr	t	P.Value	adj.P.Val	B	Legend	Symbol
1422217_PM_a_at	-5.91033306	7.13007239	-18.8245069	1.16E-17	5.24E-13	25.2926461	downregulated	Cyp1a1
1416613_PM_at	-2.50748124	4.782394307	-11.4852012	3.11E-12	5.50E-08	16.4205184	downregulated	Cyp1b1
1419728_PM_at	-3.14398254	5.864903083	-11.4065953	3.66E-12	5.50E-08	16.2908206	downregulated	Cxcl5
1423627_PM_at	-2.25552258	6.905959617	-10.7418708	1.49E-11	1.68E-07	15.1575585	downregulated	Nqo1
1420401_PM_a_at	-1.70046777	6.273757338	-8.97742269	8.09E-10	7.30E-06	11.8170157	downregulated	AB182283
1420855_PM_at	-1.76893457	7.028175365	-8.26551091	4.55E-09	3.42E-05	10.3271202	downregulated	Eln
1424296_PM_at	-1.1087974	10.34302084	-8.12407312	6.47E-09	4.17E-05	10.0213521	downregulated	Gclc
1434239_PM_at	-1.46552491	5.20810425	-7.72237647	1.78E-08	0.000100274	9.13558638	downregulated	Rrp12

1416121_PM_at	-1.33452412	8.142041728	-7.17185022	7.34E-08	0.000331312	7.88118963	downregulated	Lox
1422230_PM_s_at	2.93188411	10.13092976	7.06757063	9.64E-08	0.000395621	7.63853672	upregulated	Cyp2a4
1424631_PM_a_at	2.19360575	4.27849415	6.92059108	1.42E-07	0.000465457	7.29391517	upregulated	Ighg
1453121_PM_at	1.54361366	6.535958679	6.91882918	1.43E-07	0.000465457	7.28976588	upregulated	Tekt4
1450521_PM_a_at	1.22668097	5.776843013	6.71456955	2.45E-07	0.000563733	6.8058928	upregulated	Tcrg-V4
1455531_PM_at	1.54700362	4.351752377	6.70345124	2.52E-07	0.000563733	6.77939595	upregulated	Mfsd4
1434046_PM_at	1.56221933	4.803200573	6.6887813	2.62E-07	0.000563733	6.74441043	upregulated	AA467197
1442005_PM_at	-1.00335133	7.703486973	-6.8117336	1.89E-07	0.000563733	7.03676159	downregulated	AW987390
1438148_PM_at	-1.59263804	4.129499957	-6.78901939	2.01E-07	0.000563733	6.98290275	downregulated	Cxcl3
1439492_PM_at	1.2026569	5.466249911	6.62442612	3.11E-07	0.000611028	6.59060672	upregulated	Sptlc3
1419618_PM_at	1.20331879	5.838775592	6.44369448	5.05E-07	0.00091204	6.15590585	upregulated	Bbox1
1429918_PM_at	-1.42192566	4.547561992	-6.45091196	4.95E-07	0.00091204	6.17334164	downregulated	Arhgap20
1448872_PM_at	4.05837777	8.889848706	6.40767955	5.56E-07	0.000966101	6.06880996	upregulated	Reg3g
1438862_PM_at	-1.12545727	4.00680658	-6.37611602	6.06E-07	0.000980288	5.99235411	downregulated	A630005I04Rik
1452418_PM_at	-1.47385755	9.985275099	-6.37472444	6.08E-07	0.000980288	5.98898063	downregulated	1200016E24Rik
1429032_PM_at	1.05385727	6.472350614	6.33374375	6.79E-07	0.001056985	5.88953578	upregulated	Micalcl
1441776_PM_at	-1.13308996	6.251683653	-6.31180198	7.20E-07	0.001084048	5.83621277	downregulated	Tspan11
1418174_PM_at	1.49873155	8.906292312	6.2158099	9.34E-07	0.001317288	5.60230393	upregulated	Dbp
1422734_PM_a_at	1.17217727	5.983234378	6.13505559	1.16E-06	0.001543083	5.40475868	upregulated	Myb
1446591_PM_at	1.10875106	8.157062139	6.04030131	1.50E-06	0.001793613	5.1721145	upregulated	Gm1574
1424600_PM_at	1.68052632	5.203702005	6.0012305	1.67E-06	0.001832893	5.07592974	upregulated	Abp1
1422921_PM_at	1.04258274	6.942554502	5.9576368	1.88E-06	0.001931941	4.96844019	upregulated	Vpreb3
1419373_PM_at	1.47875209	4.799639691	5.89370476	2.24E-06	0.002200152	4.81048801	upregulated	Atp6v1b1
1451680_PM_at	-1.25453942	5.741105252	-5.87422141	2.36E-06	0.002223732	4.76227995	downregulated	Srxn1
1448973_PM_at	2.24629233	8.254712583	5.86320564	2.44E-06	0.002244918	4.73500888	upregulated	Sult1d1
1429130_PM_at	1.52571858	5.968258554	5.67172916	4.12E-06	0.003439828	4.25940726	upregulated	Ttc25
1458176_PM_at	1.31776779	7.714256499	5.62797858	4.64E-06	0.003675064	4.15034988	upregulated	Per3
1437096_PM_at	1.10775707	6.840627131	5.57361036	5.39E-06	0.004124222	4.01464521	upregulated	Ttc29
1433933_PM_s_at	1.05675106	7.759454177	5.56001641	5.60E-06	0.004140917	3.98068442	upregulated	Slco2b1
1436675_PM_at	1.1936605	7.005575088	5.51272442	6.37E-06	0.004528892	3.86244965	upregulated	Wdr63

1449091_PM_at	1.34580206	6.590265212	5.50828983	6.45E-06	0.004528892	3.85135589	upregulated	Cldn8
1418133_PM_at	-1.0755232	6.298326137	-5.47557379	7.06E-06	0.004756218	3.76947714	downregulated	Bcl3
1453310_PM_at	1.25479916	7.036506355	5.44440055	7.69E-06	0.005060009	3.69140456	upregulated	Ppil6
1428705_PM_at	1.19194974	8.398198121	5.43385585	7.92E-06	0.005106674	3.66498393	upregulated	1700007K13Rik
1441109_PM_at	1.12820092	6.236811656	5.42541398	8.11E-06	0.00515322	3.64382799	upregulated	5430414B12Rik
1419725_PM_at	-1.23568122	6.200532033	-5.39780635	8.75E-06	0.005483341	3.57461627	downregulated	Slc26a4
1439575_PM_at	1.42029805	5.473699505	5.36022647	9.70E-06	0.005761785	3.48034562	upregulated	E130009J12Rik
1428962_PM_at	1.26218421	5.740383761	5.33663122	1.04E-05	0.005991497	3.42112331	upregulated	1700013F07Rik
1452935_PM_at	1.06645524	7.827036835	5.32613326	1.07E-05	0.006089452	3.39476654	upregulated	EG546143
1434639_PM_at	-1.22730272	5.617136821	-5.28898177	1.18E-05	0.00642156	3.30145588	downregulated	Klhl29
1450060_PM_at	1.15751594	9.324321933	5.26513491	1.26E-05	0.006619121	3.24153357	upregulated	Pigr
1417160_PM_s_at	3.66104603	4.941055462	5.24965466	1.32E-05	0.006622567	3.20262391	upregulated	Expi
1416316_PM_at	1.8598432	6.080561455	5.24584343	1.33E-05	0.006622567	3.19304309	upregulated	Slc27a2
1453558_PM_at	1.12982076	6.509288534	5.23417285	1.37E-05	0.006739486	3.16370215	upregulated	4930504H06Rik
1460107_PM_at	1.35581434	6.985951769	5.20255661	1.50E-05	0.007121959	3.08419415	upregulated	1700129I04Rik
1418287_PM_a_at	5.88627427	5.241569259	5.12423463	1.86E-05	0.008316499	2.88711205	upregulated	Dmbt1
1435239_PM_at	-1.05025621	5.528932649	-5.10221824	1.98E-05	0.008666357	2.83168658	downregulated	Gria1
1427221_PM_at	1.10655355	5.842211978	5.08579273	2.07E-05	0.008754246	2.79032999	upregulated	Slc6a20a
1449316_PM_at	1.12439395	5.131296591	5.06888063	2.17E-05	0.008897696	2.74774344	upregulated	Cyp4f15
1422448_PM_at	2.96699842	7.781812666	5.03282797	2.40E-05	0.009484659	2.65694518	upregulated	Tff2
1431844_PM_at	1.28831763	6.857386722	4.89581782	3.50E-05	0.012333685	2.31180132	upregulated	Kcnmb2
1449452_PM_a_at	2.8003278	4.840379163	4.88848696	3.57E-05	0.012392324	2.29333397	upregulated	Gp2
1449486_PM_at	-1.0681598	5.83432254	-4.88909478	3.56E-05	0.012392324	2.29486513	downregulated	Ces1
1419345_PM_at	7.78656939	5.27903931	4.84522103	4.02E-05	0.013447842	2.18434968	upregulated	Pth
1449525_PM_at	1.10259493	8.731547631	4.83699965	4.11E-05	0.013457549	2.16364259	upregulated	Fmo3
1426007_PM_a_at	1.03342186	7.123456977	4.83280345	4.16E-05	0.013503851	2.15307398	upregulated	Ubxn10
1419127_PM_at	1.23794869	5.200692211	4.82074865	4.30E-05	0.013651402	2.12271388	upregulated	Npy
1436503_PM_at	3.90583409	7.402865985	4.80370184	4.51E-05	0.013871564	2.07978493	upregulated	BC048546
1422253_PM_at	1.14149803	3.728667389	4.79607038	4.61E-05	0.013871564	2.06056815	upregulated	Col10a1
1428914_PM_at	-1.02430747	9.527636523	-4.77080439	4.94E-05	0.014477059	1.9969534	downregulated	2310014D11Rik

1419268_PM_at	2.57808851	6.339736507	4.75540646	5.15E-05	0.014630418	1.95819077	upregulated	Agr2
1439740_PM_s_at	-1.08680993	6.616582167	-4.74848463	5.25E-05	0.014682914	1.94076752	downregulated	Uck2
1434028_PM_at	1.34915828	4.722381477	4.69176395	6.14E-05	0.016026724	1.79803986	upregulated	Arnt2
1418932_PM_at	-1.2457527	5.131332978	-4.65759685	6.75E-05	0.017309419	1.71211098	downregulated	LOC100046232
1424509_PM_at	1.83437233	6.223370898	4.65073982	6.88E-05	0.017482668	1.69487067	upregulated	Cd177
1417459_PM_at	6.11337172	5.206740085	4.645828	6.97E-05	0.017482668	1.68252216	upregulated	Dcpp1
1422729_PM_at	-1.04429309	4.544936261	-4.64105522	7.06E-05	0.017519441	1.67052408	downregulated	Pcdhb10
1417950_PM_a_at	1.17379722	3.754372011	4.61834127	7.52E-05	0.017831938	1.61343656	upregulated	Apoa2
1424214_PM_at	-1.0177816	6.045322019	-4.60358364	7.83E-05	0.017854721	1.5763571	downregulated	9130213B05Rik
1427123_PM_s_at	-1.13890047	7.173225185	-4.57871233	8.39E-05	0.01874153	1.51388793	downregulated	Copg2as2
1419527_PM_at	1.30012726	4.221632681	4.5575033	8.89E-05	0.019004754	1.46064	upregulated	Comp
1447526_PM_at	1.14101787	4.246739921	4.53176887	9.54E-05	0.019838015	1.39606069	upregulated	D5Ertd255e
1417991_PM_at	1.02369839	3.987545967	4.50640421	0.000102321	0.020807969	1.33244391	upregulated	Dio1
1460227_PM_at	-1.13123774	6.81317828	-4.50020797	0.000104079	0.020881026	1.31690871	downregulated	Timp1
1450790_PM_at	5.81525283	5.053419207	4.47773821	0.000110708	0.021540737	1.26059148	upregulated	Tg
1418149_PM_at	2.60475556	4.148123101	4.47014685	0.00011304	0.021668039	1.24157174	upregulated	Chga
1425260_PM_at	1.33978364	4.455479597	4.46936978	0.000113282	0.021668039	1.23962505	upregulated	Alb
1439147_PM_at	1.08551473	7.541185148	4.44440716	0.000121319	0.022914031	1.17710961	upregulated	1110049B09Rik
1452976_PM_a_at	-1.02269242	6.398931684	-4.42949671	0.000126386	0.023286576	1.13978806	downregulated	Slc9a3r2
1420796_PM_at	-1.33165408	3.991777192	-4.41886409	0.000130128	0.023496394	1.1131834	downregulated	Ahrr
1442174_PM_at	-1.08488742	7.952219004	-4.37999231	0.000144764	0.025427144	1.01598857	downregulated	Tspan18
1434777_PM_at	1.0413483	6.328737166	4.33605961	0.000163274	0.027331512	0.90627877	upregulated	Mycl1
1431780_PM_at	1.10736382	6.000150822	4.3104044	0.000175148	0.028646167	0.84228547	upregulated	1700021K14Rik
1430921_PM_at	1.1917917	5.237630973	4.29798178	0.0001812	0.029196002	0.8113195	upregulated	9130015G15Rik
1424352_PM_at	1.44133156	4.380628279	4.29623507	0.000182067	0.029196002	0.80696655	upregulated	Cyp4a12a
1423719_PM_at	2.31674024	7.582853991	4.28080894	0.000189908	0.029511823	0.76853539	upregulated	LOC632073
1445918_PM_at	1.09935288	6.338047362	4.26810491	0.000196616	0.030086183	0.73690225	upregulated	Tmem2
1438394_PM_x_at	5.82994723	5.705239986	4.25392904	0.000204376	0.030548868	0.70162221	upregulated	Krt4
1435246_PM_at	-1.10619644	5.908705738	-4.24451046	0.000209699	0.03095029	0.67819253	downregulated	Paqr8
1425221_PM_at	1.06875102	7.00080389	4.23322438	0.000216258	0.031612884	0.65012874	upregulated	Agr3

1436285_PM_at	1.00770835	6.832993376	4.21878542	0.000224946	0.032133786	0.61424365	upregulated	Ccdc113
1417514_PM_at	-1.01481162	5.471503747	-4.19910849	0.000237342	0.032965646	0.56537516	downregulated	Ssx2ip
1460258_PM_at	1.65162993	3.867183213	4.17605435	0.000252724	0.034144092	0.50817147	upregulated	Lect1
1422639_PM_at	-1.36461924	6.064424571	-4.17508847	0.00025339	0.034144092	0.5057761	downregulated	Calcb
1449451_PM_at	2.3312619	4.572084263	4.15450828	0.000267988	0.035166397	0.45476237	upregulated	Serpinb11
1458506_PM_at	1.10021633	5.858174734	4.14860461	0.000272326	0.035464913	0.44013729	upregulated	Gm626
1416306_PM_at	1.47484186	3.669867115	4.14289793	0.000276586	0.03563878	0.42600401	upregulated	Clca3
1420470_PM_at	2.59819755	4.249477459	4.1420812	0.000277201	0.03563878	0.42398158	upregulated	LOC100039077
1418608_PM_at	2.16068575	5.834907302	4.13542943	0.00028226	0.035817524	0.40751304	upregulated	Calml3
1457823_PM_at	1.07560738	7.040147571	4.13528102	0.000282374	0.035817524	0.40714565	upregulated	Cyr61
1443946_PM_s_at	1.24689688	9.995247246	4.13157927	0.00028523	0.036065975	0.39798311	upregulated	AU040972
1460011_PM_at	2.04225082	5.120237585	3.99194635	0.000416431	0.044039695	0.05362022	upregulated	Cyp26b1
1434049_PM_at	-1.0757668	4.529073083	-3.98280445	0.000426842	0.044146315	0.03116599	downregulated	Entpd3
1434152_PM_at	1.19072362	4.206114122	3.98107701	0.000428837	0.044196669	0.02692442	upregulated	666661
1425062_PM_at	1.05344684	5.562972259	3.977389	0.000433129	0.04433529	0.01787031	upregulated	Fcrl1
1457541_PM_at	1.11249943	5.732883642	3.96975883	0.000442142	0.044915563	-0.0008556	upregulated	Akap14
1450201_PM_at	-1.06805029	4.575011499	-3.92836805	0.000494337	0.047403716	-0.10228556	downregulated	Proz

Table S2.4: List of enriched pathways and gene ontology (GO) terms identified by STRING analysis.

Source	#term ID	term description	observed gene count	background gene count	strength	false discovery rate	matching proteins in your network (labels)
GO:BP	GO:0009410	Response to xenobiotic stimulus	7	86	1.26	0.0026	Ahrr, Cyp1b1, Cyp1a1, Gclc, Cyp26b1, Arnt2, Cyp2a4
GO:BP	GO:0019730	Antimicrobial humoral response	7	107	1.16	0.0052	Wfdc18, Cxcl5, Cxcl3, Npy, Reg3g, Dmbt1, Bcl3
GO:CC	GO:0005615	Extracellular space	25	1424	0.59	5.62E-06	Wfdc18, Slc26a4, Comp, Apoa2, Timp1, Agr2, Chga, Tff2, Serpinb11, Pigr, Clca3, Alb, Cxcl5, Cxcl3, Npy, Calcb, Gp2, Proz, Ces1d, BC048546, Tg, Cyp4a12a, Dmbt1, Col10a1, Lox

GO:CC	GO:0005576	Extracellular region	30	2229	0.48	2.28E-05	Wfdc18, Slc26a4, Comp, Apoa2, Timp1, Eln, Agr2, Chga, Lect1, Tff2, Serpinb11, Pigr, Cyr61, Clca3, Alb, Cxcl5, Cxcl3, Npy, Reg3g, Calcb, Gp2, Proz, Ces1d, Gria1, BC048546, Tg, Cyp4a12a, Dmbt1, Col10a1, Lox
GO:CC	GO:0062023	Collagen-containing extracellular matrix	9	359	0.75	0.023	Comp, Timp1, Eln, Cyr61, Alb, Reg3g, Dmbt1, Col10a1, Lox
GO:MF	GO:0004497	Monooxygenase activity	6	106	1.1	0.0361	Cyp1b1, Fmo3, Cyp1a1, Cyp26b1, Cyp4a12a, Cyp2a4
GO:MF	GO:0016705	Oxidoreductase activity	7	169	0.96	0.0361	Cyp1b1, Fmo3, Cyp1a1, Bbox1, Cyp26b1, Cyp4a12a, Cyp2a4
GO:MF	GO:0070330	Aromatase activity	4	32	1.44	0.0361	Cyp1b1, Cyp1a1, Cyp4a12a, Cyp2a4
GO:MF	GO:0016709	Oxidoreductase activity	4	41	1.34	0.0376	Fmo3, Cyp1a1, Cyp26b1, Cyp4a12a
GO:MF	GO:0045236	CXCR chemokine receptor binding	3	13	1.71	0.0376	Tff2, Cxcl5, Cxcl3
REAC	MMU-211945	Phase I - Functionalization of compounds	10	112	1.3	2.96E-07	Ahrr, Cyp1b1, Fmo3, Ces1d, Cyp1a1, Cyp26b1, Cyp4a12a, Arnt2, Cyp2a4, Cyp4f15
REAC	MMU-211859	Biological oxidations	11	213	1.06	2.90E-06	Ahrr, Cyp1b1, Fmo3, Ces1d, Cyp1a1, Gclc, Cyp26b1, Cyp4a12a, Arnt2, Cyp2a4, Cyp4f15
REAC	MMU-211897	Cytochrome P450 - arranged by substrate type	8	75	1.38	2.90E-06	Ahrr, Cyp1b1, Cyp1a1, Cyp26b1, Cyp4a12a, Arnt2, Cyp2a4, Cyp4f15
REAC	MMU-211981	Xenobiotics	4	27	1.52	0.0044	Ahrr, Cyp1a1, Arnt2, Cyp2a4
REAC	MMU-1430728	Metabolism	21	1779	0.42	0.0124	Nqo1, Apoa2, Ahrr, Cyp1b1, Uck2, Fmo3, Alb, Slco2b1, Ces1d, Cyp1a1, Gclc, Entpd3, Bbox1, Sptlc3, Slc27a2, Cyp26b1, Dio1, Cyp4a12a, Arnt2, Cyp2a4, Cyp4f15
WP	WP1264	Estrogen metabolism	3	13	1.71	0.0092	Nqo1, Cyp1b1, Cyp1a1
WP	WP412	Oxidative stress response	3	28	1.38	0.0352	Nqo1, Cyp1a1, Gclc

Table S2.5: List of functionally enriched disease related differentially expressed genes identified by using g:GOSt module of g:Profiler.

source	Term name	Term id	Adjusted p.value	-log10 (adj.p.value)	Term size	Query size	Intersection size	Effective domain size	intersections
GO:MF	monooxygenase activity	GO:0004497	1.84E-05	4.735333276	160	78	8	25299	CYP1A1, CYP1B1, CYP2A4, MICALCL, CYP4F15, FMO3, CYP4A12A, CYP26B1

GO:MF	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen	GO:0016709	2.01E-05	4.696196782	62	78	6	25299	CYP1A1, MICALCL, CYP4F15, FMO3, CYP4A12A, CYP26B1
GO:MF	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	GO:0016705	2.28E-05	4.641701995	233	78	9	25299	CYP1A1, CYP1B1, CYP2A4, BBOX1, MICALCL, CYP4F15, FMO3, CYP4A12A, CYP26B1
GO:MF	hydroperoxy icosatetraenoate dehydratase activity	GO:0106256	0.000104705	3.980034077	7	3	2	25299	CYP1A1, CYP1B1
GO:MF	aromatase activity	GO:0070330	0.000167741	3.775361303	36	10	3	25299	CYP1A1, CYP1B1, CYP2A4
GO:MF	oxidoreductase activity	GO:0016491	0.000212781	3.672068061	787	24	8	25299	CYP1A1, CYP1B1, NQO1, LOX, CYP2A4, BBOX1, MICALCL, SRXN1
GO:MF	estrogen 16-alpha-hydroxylase activity	GO:0101020	0.000453636	3.343292114	14	3	2	25299	CYP1A1, CYP1B1
GO:MF	heme binding	GO:0020037	0.000811614	3.09065063	182	78	7	25299	CYP1A1, CYP1B1, CYP2A4, 1700013F07RIK, CYP4F15, CYP4A12A, CYP26B1
GO:MF	tetrapyrrole binding	GO:0046906	0.001119176	2.95110161	191	78	7	25299	CYP1A1, CYP1B1, CYP2A4, 1700013F07RIK, CYP4F15, CYP4A12A, CYP26B1
GO:MF	steroid hydroxylase activity	GO:0008395	0.001222488	2.912755301	69	10	3	25299	CYP1A1, CYP1B1, CYP2A4
GO:MF	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen	GO:0016712	0.001321424	2.878957727	80	68	5	25299	CYP1A1, CYP1B1, CYP2A4, CYP4F15, CYP4A12A
GO:MF	iron ion binding	GO:0005506	0.002372847	2.624730247	214	78	7	25299	CYP1A1, CYP1B1, CYP2A4, BBOX1, CYP4F15, CYP4A12A, CYP26B1

GO:MF	fatty acid omega-hydroxylase activity	GO:0120250	0.004392126	2.357325242	15	68	3	25299	CYP1A1, CYP4F15, CYP4A12A
GO:MF	long-chain fatty acid omega-hydroxylase activity	GO:0102033	0.004392126	2.357325242	15	68	3	25299	CYP1A1, CYP4F15, CYP4A12A
GO:MF	arachidonic acid monooxygenase activity	GO:0008391	0.004910234	2.308897787	49	68	4	25299	CYP1A1, CYP2A4, CYP4F15, CYP4A12A
GO:MF	CXCR chemokine receptor binding	GO:0045236	0.00853222	2.068937978	13	12	2	25299	CXCL5, CXCL3
GO:MF	hydro-lyase activity	GO:0016836	0.009414736	2.02619187	62	3	2	25299	CYP1A1, CYP1B1
GO:MF	carbon-oxygen lyase activity	GO:0016835	0.015332514	1.814386621	79	3	2	25299	CYP1A1, CYP1B1
GO:MF	flavonoid 3'-monooxygenase activity	GO:0016711	0.021025135	1.677261204	1	1	1	25299	CYP1A1
GO:MF	signaling receptor binding	GO:0005102	0.034061505	1.467736166	1682	64	14	25299	CXCL5, CXCL3, PIGR, GRIA1, PTH, NPY, AGR2, ARNT2, CD177, APOA2, COMP, TIMP1, TG, SLC9A3R2
GO:BP	organic hydroxy compound metabolic process	GO:1901615	1.53E-06	5.814325468	571	78	14	27350	CYP1A1, CYP1B1, NQO1, SPTLC3, SULT1D1, SLC27A2, CYP4F15, CES1, PTH, NPY, APOA2, TG, CYP4A12A, CYP26B1
GO:BP	response to xenobiotic stimulus	GO:0009410	1.31E-05	4.881150381	493	10	6	27350	CYP1A1, CYP1B1, NQO1, GCLC, LOX, CYP2A4
GO:BP	response to arsenic-containing substance	GO:0046685	0.000122444	3.912060967	34	7	3	27350	CYP1A1, CYP1B1, GCLC
GO:BP	response to nutrient	GO:0007584	0.000125072	3.902841609	180	7	4	27350	CYP1A1, CYP1B1, NQO1, GCLC
GO:BP	long-chain fatty acid metabolic process	GO:0001676	0.00133756	2.873686583	124	68	6	27350	CYP1A1, CYP1B1, CYP2A4, SLC27A2, CYP4F15, CYP4A12A
GO:BP	response to organic cyclic compound	GO:0014070	0.00185324	2.732068345	1137	10	6	27350	CYP1A1, CYP1B1, NQO1, GCLC, LOX, CYP2A4
GO:BP	toxin metabolic process	GO:0009404	0.002181035	2.66133746	17	3	2	27350	CYP1A1, CYP1B1
GO:BP	xenobiotic metabolic process	GO:0006805	0.003024988	2.519276329	124	78	6	27350	CYP1A1, CYP1B1, CYP2A4, FMO3, AHRR, CYP26B1
GO:BP	secondary metabolic process	GO:0019748	0.0031825	2.497231629	66	10	3	27350	CYP1A1, CYP1B1, CYP2A4

GO:BP	reactive oxygen species metabolic process	GO:0072593	0.004915473	2.308434718	234	4	3	27350	CYP1A1, CYP1B1, NQO1
GO:BP	response to toxic substance	GO:0009636	0.005042864	2.297322732	236	4	3	27350	CYP1A1, CYP1B1, NQO1
GO:BP	response to lipid	GO:0033993	0.005777784	2.2382387	1129	4	4	27350	CYP1A1, CXCL5, CYP1B1, NQO1
GO:BP	estrogen metabolic process	GO:0008210	0.008993038	2.046093587	34	3	2	27350	CYP1A1, CYP1B1
GO:BP	aging	GO:0007568	0.009652096	2.015378348	198	60	6	27350	CYP1A1, NQO1, GCLC, NPY, COMP, TIMP1
GO:BP	response to nutrient levels	GO:0031667	0.010484343	1.979458799	546	7	4	27350	CYP1A1, CYP1B1, NQO1, GCLC
GO:BP	response to extracellular stimulus	GO:0009991	0.013409404	1.872590513	581	7	4	27350	CYP1A1, CYP1B1, NQO1, GCLC
GO:BP	fatty acid metabolic process	GO:0006631	0.015383769	1.812937259	456	44	7	27350	CYP1A1, CYP1B1, CYP2A4, ATP6V1B1, SLC27A2, CYP4F15, CES1
GO:BP	cellular response to chemokine	GO:1990869	0.01691474	1.77173468	94	12	3	27350	CXCL5, LOX, CXCL3
GO:BP	response to chemokine	GO:1990868	0.01691474	1.77173468	94	12	3	27350	CXCL5, LOX, CXCL3
GO:BP	antimicrobial humoral response	GO:0019730	0.018187423	1.740228831	157	49	5	27350	CXCL5, CXCL3, REG3G, BCL3, NPY
GO:BP	phenylpropanoid metabolic process	GO:0009698	0.018729948	1.727463423	13	10	2	27350	CYP1A1, CYP2A4
GO:BP	coumarin metabolic process	GO:0009804	0.018729948	1.727463423	13	10	2	27350	CYP1A1, CYP2A4
GO:BP	response to inorganic substance	GO:0010035	0.022200248	1.653642165	660	7	4	27350	CYP1A1, CYP1B1, NQO1, GCLC
GO:BP	regulation of biological quality	GO:0065008	0.028864927	1.53962954	4003	63	23	27350	CYP1A1, CXCL5, CYP1B1, NQO1, ELN, GCLC, LOX, CXCL3, MYB, ATP6V1B1, PER3, SLC26A4, GRIA1, KCNMB2, CES1, PTH, NPY, AGR2, APOA2, COMP, DIO1, TG, CHGA
GO:BP	retinol metabolic process	GO:0042572	0.029316254	1.532891521	61	3	2	27350	CYP1A1, CYP1B1
GO:BP	response to follicle-stimulating hormone	GO:0032354	0.030911837	1.509875189	24	7	2	27350	CYP1B1, GCLC
GO:BP	cellular response to gonadotropin stimulus	GO:0071371	0.033595726	1.473715968	25	7	2	27350	CYP1B1, GCLC

GO:BP	small molecule metabolic process	GO:0044281	0.033729908	1.471984846	1847	79	17	27350	CYP1A1, CYP1B1, NQO1, GCLC, CYP2A4, SPTLC3, ATP6V1B1, SLC27A2, CYP4F15, CES1, PTH, UCK2, APOA2, DIO1, CYP4A12A, CYP26B1, ENTPD3
GO:BP	transcytosis	GO:0045056	0.036219303	1.441059907	23	61	3	27350	PIGR, GP2, TG
GO:BP	positive regulation of receptor signaling pathway via JAK-STAT	GO:0046427	0.036486893	1.437863112	68	3	2	27350	CXCL5, CYP1B1
GO:BP	positive regulation of receptor signaling pathway via STAT	GO:1904894	0.039799512	1.400122252	71	3	2	27350	CXCL5, CYP1B1
GO:BP	cellular response to chemical stimulus	GO:0070887	0.044002245	1.356525167	3503	12	8	27350	CYP1A1, CXCL5, CYP1B1, NQO1, GCLC, LOX, CYP2A4, CXCL3
GO:BP	response to external stimulus	GO:0009605	0.04459756	1.3506889	3135	13	8	27350	CYP1A1, CXCL5, CYP1B1, NQO1, GCLC, LOX, CXCL3, AA467197
GO:BP	cell redox homeostasis	GO:0045454	0.048684113	1.312612734	30	7	2	27350	NQO1, GCLC
GO:BP	cellular response to xenobiotic stimulus	GO:0071466	0.049745219	1.303248652	202	78	6	27350	CYP1A1, CYP1B1, CYP2A4, FMO3, AHRR, CYP26B1
GO:CC	extracellular space	GO:0005615	2.36E-06	5.626507044	1984	68	21	27145	CXCL5, ELN, LOX, CXCL3, REG3G, VPREB3, SLC26A4, PIGR, GP2, PTH, NPY, COL10A1, AGR2, DCPP1, APOA2, COMP, TIMP1, TG, ALB, CHGA, CYP4A12A
GO:CC	extracellular region	GO:0005576	3.05E-06	5.515316107	2774	75	26	27145	CXCL5, ELN, LOX, CXCL3, REG3G, VPREB3, SLC26A4, PIGR, DMBT1, GRIA1, TFF2, GP2, PTH, NPY, COL10A1, AGR2, DCPP1, APOA2, COMP, TIMP1, TG, ALB, CHGA, CYP4A12A, CALCB, SERPINB11
GO:CC	apical plasma membrane	GO:0016324	0.002309001	2.636575961	407	68	8	27145	ATP6V1B1, SLCO2B1, SLC26A4, SLC6A20A,

									CYP4F15, GP2, SLC9A3R2, CYP4A12A
GO:CC	collagen-containing extracellular matrix	GO:0062023	0.006453492	2.190205218	373	62	7	27145	ELN, LOX, REG3G, COL10A1, COMP, TIMP1, ALB
GO:CC	apical part of cell	GO:0045177	0.010537439	1.977264914	503	68	8	27145	ATP6V1B1, SLCO2B1, SLC26A4, SLC6A20A, CYP4F15, GP2, SLC9A3R2, CYP4A12A
GO:CC	endomembrane system	GO:0012505	0.011471031	1.940397542	4239	71	25	27145	CYP1A1, CYP1B1, RRP12, CYP2A4, SPTLC3, REG3G, VPREB3, ATP6V1B1, PIGR, SLC27A2, GRIA1, CES1, GP2, FMO3, UBXN10, NPY, AGR2, CD177, TG, ALB, CHGA, SLC9A3R2, CYP4A12A, PAQR8, AGR3
GO:CC	endoplasmic reticulum	GO:0005783	0.020944959	1.678920475	1877	78	16	27145	CYP1A1, CYP1B1, CYP2A4, SPTLC3, VPREB3, SLC27A2, GRIA1, CES1, FMO3, UBXN10, AGR2, TG, ALB, CYP4A12A, AGR3, CYP26B1
GO:CC	cytoplasm	GO:0005737	0.025451825	1.594281079	11483	54	37	27145	CYP1A1, CYP1B1, NQO1, AB182283, ELN, GCLC, RRP12, CYP2A4, AA467197, SPTLC3, BBOX1, REG3G, MICALCL, MYB, VPREB3, ATP6V1B1, SRXN1, SULT1D1, PER3, TTC29, BCL3, 1700007K13RIK, 1700013F07RIK, PIGR, SLC27A2, GRIA1, CYP4F15, CES1, GP2, FMO3, UBXN10, NPY, COL10A1, AGR2, UCK2, ARNT2, CD177
GO:CC	extracellular matrix	GO:0031012	0.042160229	1.375097041	503	62	7	27145	ELN, LOX, REG3G, COL10A1, COMP, TIMP1, ALB
GO:CC	external encapsulating structure	GO:0030312	0.043201254	1.364503649	505	62	7	27145	ELN, LOX, REG3G, COL10A1, COMP, TIMP1, ALB

KEGG	Chemical carcinogenesis - reactive oxygen species	KEGG:05208	0.00453287	2.343626747	220	4	3	8875	CYP1A1, CYP1B1, NQO1
KEGG	Tryptophan metabolism	KEGG:00380	0.007432943	2.128839176	51	3	2	8875	CYP1A1, CYP1B1
KEGG	Ovarian steroidogenesis	KEGG:04913	0.01101493	1.958018242	62	3	2	8875	CYP1A1, CYP1B1
KEGG	Metabolism of xenobiotics by cytochrome P450	KEGG:00980	0.01529519	1.81544512	73	3	2	8875	CYP1A1, CYP1B1
KEGG	Chemical carcinogenesis - DNA adducts	KEGG:05204	0.020271967	1.693104108	84	3	2	8875	CYP1A1, CYP1B1
KEGG	Steroid hormone biosynthesis	KEGG:00140	0.023800853	1.623407474	91	3	2	8875	CYP1A1, CYP1B1
REAC	Synthesis of epoxy (EET) and dihydroxyeicosatrienoic acids (DHET)	REAC:R-MMU-2142670	0.000484334	3.314854658	7	3	2	8384	CYP1A1, CYP1B1
REAC	Cytochrome P450 - arranged by substrate type	REAC:R-MMU-211897	0.000622087	3.206148586	66	78	7	8384	CYP1A1, CYP1B1, CYP2A4, CYP4F15, ARNT2, CYP4A12A, CYP26B1
REAC	Synthesis of (16-20)-hydroxyeicosatetraenoic acids (HETE)	REAC:R-MMU-2142816	0.000645728	3.189950479	8	3	2	8384	CYP1A1, CYP1B1
REAC	Phase I - Functionalization of compounds	REAC:R-MMU-211945	0.001142211	2.942253563	102	78	8	8384	CYP1A1, CYP1B1, CYP2A4, CYP4F15, FMO3, ARNT2, CYP4A12A, CYP26B1
REAC	Biological oxidations	REAC:R-MMU-211859	0.015322106	1.814681537	198	10	4	8384	CYP1A1, CYP1B1, GCLC, CYP2A4
REAC	Arachidonic acid metabolism	REAC:R-MMU-2142753	0.035379381	1.451249765	56	3	2	8384	CYP1A1, CYP1B1
WP	Estrogen metabolism	WP:WP1264	2.66E-06	5.574467381	13	4	3	4471	CYP1A1, CYP1B1, NQO1
WP	Oxidative stress response	WP:WP412	0.000262993	3.580056144	28	7	3	4471	CYP1A1, NQO1, GCLC
WP	Transcriptional activation by Nfe2l2 in response to phytochemicals	WP:WP1245	0.006581059	2.181704214	14	7	2	4471	NQO1, GCLC
WP	Oxidation by cytochrome P450	WP:WP1274	0.008084856	2.092327709	40	3	2	4471	CYP1A1, CYP1B1
WP	Tryptophan metabolism	WP:WP79	0.009799595	2.008791875	44	3	2	4471	CYP1A1, CYP1B1

Table S2.6: List of enriched gene ontology (GO) terms of cluster 4 and cluster 5 identified by STRING analysis.

Cluster # 4

#term ID	term description	observed gene count	background gene count	strength	false discovery rate	matching proteins in your network (labels)
GO:0030001	Metal ion transport	18	666	0.97	5.17E-09	KCNA7, KCNJ2, KCNH3, ABCC9, KCNQ4, SCN2B, GRIN2C, KCNG3, KCNJ4, KCND3, TRPM8, GRIN3A, KCNB1, SCN2A, CAMK2A, ATP2C2, ATP1A3, KCNJ12
GO:0098662	Inorganic cation transmembrane transport	17	628	0.97	1.31E-08	KCNA7, KCNJ2, KCNH3, ABCC9, KCNQ4, SCN2B, GRIN2C, KCNG3, KCNJ4, KCND3, TRPM8, GRIN3A, KCNB1, SCN2A, ATP2C2, ATP1A3, KCNJ12
GO:0034765	Regulation of ion transmembrane transport	14	477	1	1.29E-07	LRRC7, KCNA7, KCNJ2, KCNH3, KCNQ4, SCN2B, SHANK1, KCNG3, KCNJ4, KCND3, KCNB1, SCN2A, DLG2, KCNJ12
GO:0043269	Regulation of ion transport	16	696	0.9	1.42E-07	LRRC7, KCNA7, KCNJ2, KCNH3, KCNQ4, SCN2B, SHANK1, KCNG3, KCNJ4, KCND3, KCNB1, SCN2A, DLG2, CAMK2A, ATP2C2, KCNJ12
GO:0035637	Multicellular organismal signaling	8	137	1.3	8.80E-06	KCNJ2, ABCC9, SCN2B, KCNJ4, KCND3, SCN2A, ATP1A3, KCNJ12
GO:0051049	Regulation of transport	21	1776	0.61	1.08E-05	LRRC7, NNAT, KCNA7, KCNJ2, KCNH3, KCNQ4, SCN2B, SHANK1, KCNG3, KCNJ4, KCND3, GRIN3A, NPR1, KCNB1, SCN2A, DLG2, CAMK2A, ATP2C2, CADPS2, UNC13A, KCNJ12
GO:0061337	Cardiac conduction	7	94	1.41	1.42E-05	KCNJ2, ABCC9, SCN2B, KCNJ4, KCND3, ATP1A3, KCNJ12
GO:1990573	Potassium ion import across plasma membrane	5	43	1.6	0.00019	KCNJ2, ABCC9, KCNJ4, ATP1A3, KCNJ12
GO:0042391	Regulation of membrane potential	10	440	0.89	0.00041	KCNJ2, KCNH3, SCN2B, GRIN2C, SHANK1, KCND3, GRIN3A, KCNB1, SCN2A, ATP1A3
GO:0008016	Regulation of heart contraction	8	245	1.05	0.00042	KCNJ2, ABCC9, SCN2B, KCNJ4, KCND3, NPR1, ATP1A3, KCNJ12
GO:0032879	Regulation of localization	22	2740	0.44	0.0024	LRRC7, NNAT, KCNA7, KCNJ2, KCNH3, KCNQ4, SCN2B, SHANK1, KCNG3, KCNJ4, KCND3, NHLRC1, GRIN3A, NPR1, KCNB1, SCN2A, DLG2, CAMK2A, ATP2C2, CADPS2, UNC13A, KCNJ12
GO:0001508	Action potential	5	97	1.25	0.006	KCNJ2, SCN2B, KCND3, KCNB1, SCN2A
GO:0051260	Protein homooligomerization	6	182	1.05	0.0085	KCNA7, KCNJ2, KCNG3, KCND3, KCNB1, KCNJ12
GO:0035418	Protein localization to synapse	4	63	1.34	0.0196	LRRC7, GRIN2C, SHANK1, DLG2

GO:1900449	Regulation of glutamate receptor signaling pathway	4	63	1.34	0.0196	LRRC7, SHANK1, DLG2, UNC13A
GO:0044057	Regulation of system process	9	592	0.72	0.0232	KCNJ2, ABCC9, SCN2B, SHANK1, KCNJ4, KCND3, NPR1, ATP1A3, KCNJ12
GO:0086010	Membrane depolarization during action potential	3	30	1.54	0.0464	KCNJ2, SCN2B, SCN2A
Cluster # 5						
GO:0007166	Cell surface receptor signaling pathway	30	2325	0.69	3.16E-11	CLDN18, FGF22, PIK3R2, HES1, EGR1, EGF, NKD1, FGF18, KIT, CXCL5, FOS, CXCL8, CXCR2, MMP1, WNT11, IL20RB, GP1BA, SOCS1, HEY1, C5AR1, ROS1, LCN2, KLF4, ID1, NPPB, CCL26, ITGB2, PBX1, PGF, NTF4
GO:2000026	Regulation of multicellular organismal development	29	2096	0.72	3.16E-11	CLDN18, HES1, ID2, EGR1, EGR2, EGF, NKD1, MEGF10, FGF18, KIT, KNDC1, FOS, CXCL8, CXCR2, WNT11, GP1BA, SOCS1, HEY1, HOXA9, C5AR1, MAP2, KLF4, ID1, NPPB, ITGB2, PBX1, PGF, NTF4, RAB7B
GO:0050793	Regulation of developmental process	31	2648	0.64	5.83E-11	CLDN18, HES1, ID2, EGR1, EGR2, EGF, NKD1, MEGF10, FGF18, KIT, KNDC1, FOS, CXCL8, HAS2, CXCR2, WNT11, GP1BA, SOCS1, HEY1, HOXA9, C5AR1, MAP2, KLF4, ID1, NPPB, POSTN, ITGB2, PBX1, PGF, NTF4, RAB7B
GO:0051239	Regulation of multicellular organismal process	33	3227	0.59	1.49E-10	CLDN18, HES1, ID2, EGR1, EGR2, EGF, NKD1, MEGF10, FGF18, KIT, KNDC1, FOS, CXCL8, HAS2, CXCR2, WNT11, IL20RB, GP1BA, SOCS1, HEY1, HOXA9, C5AR1, MAP2, LCN2, KLF4, ID1, NPPB, POSTN, ITGB2, PBX1, PGF, NTF4, RAB7B
GO:0048518	Positive regulation of biological process	41	6112	0.4	5.90E-09	FGF22, PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, MEGF10, FGF18, KIT, CXCL5, KNDC1, FOS, CXCL8, HAS2, CXCR2, MMP1, WNT11, IL20RB, SOCS1, HEY1, HOXA9, C5AR1, MAP2, ROS1, LCN2, KLF4, ID1, NPPB, POSTN, CCL26, GPT, ITGB2, MASP2, ETV1, PBX1, PGF, NTF4, RAB7B
GO:0071310	Cellular response to organic substance	27	2369	0.63	9.00E-09	CLDN18, FGF22, PIK3R2, EGR1, EGR2, FGF18, KIT, CXCL5, FOS, CXCL8, HAS2, CXCR2, MMP1, WNT11, IL20RB, SOCS1, MAP2, LCN2, KLF4, ID1, POSTN, CCL26, GPT, ITGB2, PGF, NTF4, RAB7B
GO:0070887	Cellular response to chemical stimulus	29	2919	0.57	2.18E-08	CLDN18, FGF22, PIK3R2, ID2, EGR1, EGR2, FGF18, KIT, CXCL5, FOS, CXCL8, HAS2, CXCR2, MMP1, WNT11, IL20RB, SOCS1, C5AR1, MAP2, LCN2, KLF4, ID1, POSTN, CCL26, GPT, ITGB2, PGF, NTF4, RAB7B

GO:0048522	Positive regulation of cellular process	38	5579	0.41	6.32E-08	FGF22, PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, MEGF10, FGF18, KIT, CXCL5, KNDC1, FOS, CXCL8, HAS2, CXCR2, MMP1, WNT11, SOCS1, HEY1, HOXA9, C5AR1, MAP2, ROS1, LCN2, KLF4, ID1, POSTN, CCL26, GPT, ITGB2, ETV1, PBX1, PGF, NTF4, RAB7B
GO:1901342	Regulation of vasculature development	12	336	1.13	1.25E-07	EGR1, FGF18, KIT, CXCL8, CXCR2, HEY1, C5AR1, KLF4, ID1, NPPB, ITGB2, PGF
GO:0010033	Response to organic substance	28	3011	0.54	2.21E-07	CLDN18, FGF22, PIK3R2, EGR1, EGR2, FGF18, KIT, CXCL5, FOS, CXCL8, HAS2, CXCR2, MMP1, WNT11, IL20RB, SOCS1, C5AR1, MAP2, LCN2, KLF4, ID1, POSTN, CCL26, GPT, ITGB2, PGF, NTF4, RAB7B
GO:0009653	Anatomical structure morphogenesis	24	2165	0.62	2.56E-07	HES1, ID2, EGR2, HOXC13, EGF, NKD1, FGF18, KNDC1, CXCL8, HAS2, CXCR2, WNT11, GP1BA, HEY1, HOXA9, MAP2, KLF4, ID1, POSTN, ITGB2, ETV1, PBX1, PGF, NTF4
GO:0022603	Regulation of anatomical structure morphogenesis	18	1095	0.79	2.56E-07	HES1, NKD1, FGF18, KIT, KNDC1, CXCL8, HAS2, CXCR2, WNT11, HEY1, C5AR1, MAP2, KLF4, ID1, NPPB, POSTN, ITGB2, PGF
GO:0051240	Positive regulation of multicellular organismal process	22	1770	0.67	2.56E-07	HES1, ID2, EGR1, EGR2, EGF, FGF18, KIT, FOS, CXCL8, HAS2, CXCR2, WNT11, IL20RB, SOCS1, C5AR1, LCN2, KLF4, NPPB, POSTN, ITGB2, PGF, RAB7B
GO:0009893	Positive regulation of metabolic process	31	3893	0.48	3.91E-07	PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, FGF18, KIT, KNDC1, FOS, HAS2, WNT11, IL20RB, HEY1, HOXA9, C5AR1, ROS1, LCN2, KLF4, ID1, POSTN, CCL26, GPT, ITGB2, ETV1, PBX1, PGF, NTF4, RAB7B
GO:0071345	Cellular response to cytokine stimulus	17	1013	0.8	5.05E-07	CLDN18, EGR1, KIT, CXCL5, FOS, CXCL8, HAS2, CXCR2, MMP1, IL20RB, SOCS1, LCN2, KLF4, POSTN, CCL26, ITGB2, RAB7B
GO:0045595	Regulation of cell differentiation	22	1874	0.65	5.20E-07	CLDN18, HES1, ID2, EGR2, MEGF10, FGF18, KIT, KNDC1, FOS, HAS2, WNT11, GP1BA, SOCS1, HEY1, HOXA9, MAP2, KLF4, ID1, POSTN, PBX1, NTF4, RAB7B
GO:0009888	Tissue development	21	1760	0.65	1.07E-06	HES1, ID2, EGR1, EGR2, HOXC13, EGF, MEGF10, FGF18, FOS, HAS2, CXCR2, WNT11, HEY1, ROS1, KLF4, ID1, POSTN, ITGB2, PBX1, PGF, NTF4
GO:1904018	Positive regulation of vasculature development	9	189	1.25	1.55E-06	EGR1, FGF18, KIT, CXCL8, CXCR2, C5AR1, KLF4, ITGB2, PGF
GO:0051716	Cellular response to stimulus	38	6489	0.34	3.14E-06	CLDN18, FGF22, PIK3R2, HES1, ID2, EGR1, EGR2, EGF, NKD1, FGF18, KIT, CXCL5, KNDC1, FOS, CXCL8, HAS2, CXCR2, MMP1, WNT11, IL20RB, GP1BA, SOCS1, HEY1, C5AR1, MAP2,

						ROS1, LCN2, KLF4, ID1, NPPB, POSTN, CCL26, GPT, ITGB2, PBX1, PGF, NTF4, RAB7B
GO:0050896	Response to stimulus	42	8046	0.29	3.73E-06	CLDN18, FGF22, PIK3R2, HES1, ID2, EGR1, EGR2, EGF, NKD1, FGF18, KIT, PKD1L1, CXCL5, KNDC1, FOS, CXCL8, HAS2, CXCR2, MMP1, WNT11, IL20RB, GP1BA, SOCS1, HEY1, C5AR1, MAP2, ROS1, LCN2, KLF4, ID1, NPPB, POSTN, CCL26, GPT, ITGB2, MASP2, BEST1, ETV1, PBX1, PGF, NTF4, RAB7B
GO:0051094	Positive regulation of developmental process	18	1389	0.69	5.38E-06	HES1, ID2, EGR1, EGR2, EGF, FGF18, KIT, FOS, CXCL8, HAS2, CXCR2, WNT11, SOCS1, C5AR1, KLF4, ITGB2, PGF, RAB7B
GO:0008284	Positive regulation of cell population proliferation	15	919	0.79	6.12E-06	HES1, ID2, EGR1, EGF, MEGF10, FGF18, KIT, CXCL5, HAS2, CXCR2, C5AR1, ID1, CCL26, PBX1, PGF
GO:0010604	Positive regulation of macromolecule metabolic process	28	3600	0.47	6.12E-06	PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, FGF18, KIT, KNDC1, FOS, HAS2, WNT11, IL20RB, HEY1, HOXA9, C5AR1, LCN2, KLF4, ID1, POSTN, CCL26, ETV1, PBX1, PGF, NTF4, RAB7B
GO:0040011	Locomotion	17	1251	0.71	6.98E-06	PIK3R2, HES1, ID2, EGR2, MEGF10, KIT, CXCL5, CXCL8, CXCR2, MMP1, WNT11, C5AR1, ID1, CCL26, ITGB2, ETV1, PGF
GO:0009719	Response to endogenous stimulus	18	1447	0.67	8.55E-06	FGF22, PIK3R2, EGR1, EGR2, FGF18, KIT, PKD1L1, FOS, CXCL8, HAS2, SOCS1, KLF4, ID1, POSTN, GPT, ITGB2, PGF, NTF4
GO:0042127	Regulation of cell population proliferation	19	1642	0.64	9.14E-06	HES1, ID2, EGR1, EGF, MEGF10, FGF18, KIT, CXCL5, CXCL8, HAS2, CXCR2, WNT11, IL20RB, C5AR1, KLF4, ID1, CCL26, PBX1, PGF
GO:0019221	Cytokine-mediated signaling pathway	13	678	0.86	9.42E-06	CLDN18, EGR1, KIT, CXCL5, FOS, CXCL8, CXCR2, MMP1, IL20RB, SOCS1, LCN2, CCL26, ITGB2
GO:0051272	Positive regulation of cellular component movement	12	561	0.91	1.08E-05	EGF, FGF18, KIT, CXCL8, HAS2, CXCR2, WNT11, C5AR1, MAP2, POSTN, CCL26, PGF
GO:0007165	Signal transduction	32	4876	0.39	1.12E-05	CLDN18, FGF22, PIK3R2, HES1, EGR1, EGF, NKD1, FGF18, KIT, CXCL5, KNDC1, FOS, CXCL8, CXCR2, MMP1, WNT11, IL20RB, GP1BA, SOCS1, HEY1, C5AR1, ROS1, LCN2, KLF4, ID1, NPPB, CCL26, ITGB2, PBX1, PGF, NTF4, RAB7B
GO:0009605	Response to external stimulus	22	2310	0.55	1.29E-05	ID2, EGR1, EGR2, KIT, PKD1L1, CXCL5, FOS, CXCL8, CXCR2, WNT11, C5AR1, LCN2, NPPB, POSTN, CCL26, GPT, ITGB2, MASP2, BEST1, ETV1, PGF, RAB7B
GO:0042221	Response to chemical	30	4333	0.42	1.32E-05	CLDN18, FGF22, PIK3R2, ID2, EGR1, EGR2, FGF18, KIT, CXCL5, FOS, CXCL8, HAS2, CXCR2, MMP1, WNT11, IL20RB, SOCS1, C5AR1, MAP2, LCN2, KLF4, ID1, POSTN, CCL26, GPT, ITGB2, ETV1, PGF, NTF4, RAB7B

GO:0035295	Tube development	14	851	0.79	1.36E-05	CLDN18, HES1, ID2, EGF, FGF18, KIT, CXCL8, HAS2, CXCR2, WNT11, HEY1, ID1, PBX1, PGF
GO:0097529	Myeloid leukocyte migration	7	123	1.33	1.88E-05	KIT, CXCL5, CXCL8, CXCR2, C5AR1, CCL26, ITGB2
GO:0007154	Cell communication	33	5320	0.37	1.93E-05	CLDN18, FGF22, PIK3R2, HES1, EGR1, EGF, NKD1, FGF18, KIT, CXCL5, KNDC1, FOS, CXCL8, CXCR2, MMP1, WNT11, IL20RB, GP1BA, SOCS1, HEY1, C5AR1, ROS1, LCN2, KLF4, ID1, NPPB, POSTN, CCL26, ITGB2, PBX1, PGF, NTF4, RAB7B
GO:0030593	Neutrophil chemotaxis	6	74	1.48	2.38E-05	CXCL5, CXCL8, CXCR2, C5AR1, CCL26, ITGB2
GO:0071363	Cellular response to growth factor stimulus	11	494	0.92	2.38E-05	FGF22, EGR1, FGF18, FOS, CXCL8, HAS2, KLF4, ID1, POSTN, PGF, NTF4
GO:0031325	Positive regulation of cellular metabolic process	26	3413	0.46	2.90E-05	PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, FGF18, KIT, KNDC1, FOS, HAS2, WNT11, HEY1, HOXA9, C5AR1, ROS1, KLF4, ID1, CCL26, GPT, ITGB2, ETV1, PBX1, PGF, NTF4
GO:0048513	Animal organ development	25	3197	0.47	3.60E-05	HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, MEGF10, FGF18, KIT, KNDC1, FOS, CXCL8, HAS2, CXCR2, WNT11, HEY1, HOXA9, C5AR1, KLF4, ID1, ETV1, PBX1, PGF, NTF4
GO:0001932	Regulation of protein phosphorylation	17	1459	0.64	3.67E-05	PIK3R2, HES1, EGR1, EGF, FGF18, KIT, KNDC1, WNT11, SOCS1, C5AR1, ROS1, KLF4, ID1, CCL26, ITGB2, PGF, NTF4
GO:0030335	Positive regulation of cell migration	11	522	0.9	3.67E-05	EGF, FGF18, KIT, CXCL8, HAS2, CXCR2, WNT11, C5AR1, POSTN, CCL26, PGF
GO:0030595	Leukocyte chemotaxis	7	142	1.27	3.67E-05	KIT, CXCL5, CXCL8, CXCR2, C5AR1, CCL26, ITGB2
GO:0032502	Developmental process	34	5841	0.34	3.67E-05	CLDN18, FGF22, HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, MEGF10, FGF18, KIT, PKD1L1, KNDC1, FOS, CXCL8, HAS2, CXCR2, WNT11, GP1BA, SOCS1, HEY1, HOXA9, C5AR1, MAP2, ROS1, KLF4, ID1, POSTN, ITGB2, ETV1, PBX1, PGF, NTF4
GO:0035239	Tube morphogenesis	12	656	0.84	3.67E-05	HES1, ID2, EGF, FGF18, CXCL8, HAS2, CXCR2, WNT11, HEY1, ID1, PBX1, PGF
GO:0045765	Regulation of angiogenesis	9	303	1.05	3.67E-05	FGF18, CXCL8, CXCR2, C5AR1, KLF4, ID1, NPPB, ITGB2, PGF
GO:0051173	Positive regulation of nitrogen compound metabolic process	25	3239	0.46	3.71E-05	PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, FGF18, KIT, KNDC1, FOS, HAS2, WNT11, HEY1, HOXA9, C5AR1, KLF4, ID1, CCL26, ITGB2, ETV1, PBX1, PGF, NTF4
GO:0060284	Regulation of cell development	14	956	0.74	3.72E-05	CLDN18, HES1, ID2, EGR2, KIT, KNDC1, HAS2, HEY1, MAP2, KLF4, ID1, POSTN, PBX1, NTF4
GO:0065009	Regulation of molecular function	31	4913	0.38	3.72E-05	FGF22, PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, FGF18, KIT, CXCL5, KNDC1, FOS, CXCL8, WNT11, SOCS1, HEY1, HOXA9, C5AR1, MAP2, ROS1, KLF4, ID1, NPPB, CCL26, ITGB2, ETV1, PBX1, PGF, NTF4, RAB7B

GO:0050900	Leukocyte migration	9	316	1.03	4.01E-05	PIK3R2, KIT, CXCL5, CXCL8, CXCR2, MMP1, C5AR1, CCL26, ITGB2
GO:0042330	Taxis	11	547	0.88	4.46E-05	ID2, EGR2, KIT, CXCL5, CXCL8, CXCR2, C5AR1, CCL26, ITGB2, ETV1, PGF
GO:0010628	Positive regulation of gene expression	21	2337	0.53	4.64E-05	PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, KIT, FOS, WNT11, IL20RB, HEY1, HOXA9, C5AR1, LCN2, KLF4, ID1, POSTN, ETV1, PBX1, RAB7B
GO:0051093	Negative regulation of developmental process	14	983	0.73	4.64E-05	CLDN18, HES1, ID2, NKD1, WNT11, SOCS1, HEY1, HOXA9, MAP2, KLF4, ID1, NPPB, POSTN, PBX1
GO:0002682	Regulation of immune system process	17	1514	0.63	4.72E-05	CLDN18, PIK3R2, HES1, ID2, KIT, FOS, CXCL8, CXCR2, IL20RB, GP1BA, SOCS1, HOXA9, C5AR1, ITGB2, MASP2, PGF, RAB7B
GO:0003002	Regionalization	9	332	1.01	5.23E-05	HES1, EGR2, HOXC13, PKD1L1, WNT11, HEY1, HOXA9, PBX1, NTF4
GO:0007275	Multicellular organism development	31	5023	0.37	5.23E-05	CLDN18, HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, MEGF10, FGF18, KIT, PKD1L1, KNDC1, FOS, CXCL8, HAS2, CXCR2, WNT11, HEY1, HOXA9, C5AR1, MAP2, ROS1, KLF4, ID1, POSTN, ITGB2, ETV1, PBX1, PGF, NTF4
GO:0051270	Regulation of cellular component movement	14	1009	0.72	5.65E-05	EGF, NKD1, FGF18, KIT, CXCL8, HAS2, CXCR2, WNT11, C5AR1, MAP2, KLF4, POSTN, CCL26, PGF
GO:0071495	Cellular response to endogenous stimulus	15	1181	0.68	5.80E-05	FGF22, PIK3R2, EGR1, FGF18, KIT, FOS, CXCL8, HAS2, SOCS1, KLF4, ID1, POSTN, GPT, PGF, NTF4
GO:0007167	Enzyme linked receptor protein signaling pathway	12	720	0.8	6.81E-05	FGF22, PIK3R2, EGR1, EGF, FGF18, KIT, FOS, ROS1, ID1, NPPB, PGF, NTF4
GO:0048856	Anatomical structure development	32	5402	0.35	6.81E-05	CLDN18, HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, MEGF10, FGF18, KIT, PKD1L1, KNDC1, FOS, CXCL8, HAS2, CXCR2, WNT11, GP1BA, HEY1, HOXA9, C5AR1, MAP2, ROS1, KLF4, ID1, POSTN, ITGB2, ETV1, PBX1, PGF, NTF4
GO:0045766	Positive regulation of angiogenesis	7	169	1.19	7.58E-05	FGF18, CXCL8, CXCR2, C5AR1, KLF4, ITGB2, PGF
GO:0030154	Cell differentiation	26	3702	0.42	8.55E-05	FGF22, HES1, ID2, EGR1, EGR2, NKD1, MEGF10, FGF18, KIT, KNDC1, FOS, HAS2, WNT11, SOCS1, HEY1, C5AR1, MAP2, ROS1, KLF4, ID1, POSTN, ITGB2, ETV1, PBX1, PGF, NTF4
GO:0016477	Cell migration	13	896	0.74	9.17E-05	PIK3R2, HES1, MEGF10, KIT, CXCL5, CXCL8, CXCR2, MMP1, WNT11, C5AR1, ID1, CCL26, ITGB2
GO:0001822	Kidney development	8	271	1.05	0.00012	HES1, ID2, EGR1, HAS2, CXCR2, WNT11, PBX1, PGF
GO:0001568	Blood vessel development	10	500	0.88	0.00013	HES1, EGR1, EGF, FGF18, CXCL8, HAS2, WNT11, HEY1, ID1, PGF

GO:2000145	Regulation of cell motility	13	929	0.72	0.00013	EGF, NKD1, FGF18, KIT, CXCL8, HAS2, CXCR2, WNT11, C5AR1, KLF4, POSTN, CCL26, PGF
GO:0048731	System development	28	4426	0.38	0.00018	CLDN18, HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, MEGF10, FGF18, KIT, KNDC1, FOS, CXCL8, HAS2, CXCR2, WNT11, HEY1, HOXA9, C5AR1, MAP2, KLF4, ID1, POSTN, ETV1, PBX1, PGF, NTF4
GO:0009887	Animal organ morphogenesis	13	967	0.7	0.00019	HES1, ID2, HOXC13, NKD1, FGF18, HAS2, WNT11, HEY1, HOXA9, ID1, PBX1, PGF, NTF4
GO:0048514	Blood vessel morphogenesis	9	410	0.92	0.00022	HES1, EGF, FGF18, CXCL8, HAS2, WNT11, HEY1, ID1, PGF
GO:0051101	Regulation of dna binding	6	128	1.25	0.00022	HES1, ID2, EGF, HEY1, KLF4, ID1
GO:0006935	Chemotaxis	10	545	0.84	0.00025	EGR2, KIT, CXCL5, CXCL8, CXCR2, C5AR1, CCL26, ITGB2, ETV1, PGF
GO:0031328	Positive regulation of cellular biosynthetic process	18	2005	0.53	0.00032	PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, FOS, HAS2, WNT11, HEY1, HOXA9, KLF4, ID1, GPT, ITGB2, ETV1, PBX1
GO:0030334	Regulation of cell migration	12	865	0.72	0.00034	EGF, FGF18, KIT, CXCL8, HAS2, CXCR2, WNT11, C5AR1, KLF4, POSTN, CCL26, PGF
GO:0002376	Immune system process	20	2481	0.48	0.00036	PIK3R2, HES1, ID2, EGR1, KIT, CXCL5, CXCL8, CXCR2, MMP1, IL20RB, HOXA9, C5AR1, LCN2, KLF4, NPPB, CCL26, ITGB2, MASP2, PBX1, RAB7B
GO:0010721	Negative regulation of cell development	8	330	0.96	0.00038	CLDN18, HES1, ID2, HEY1, MAP2, ID1, POSTN, PBX1
GO:0032879	Regulation of localization	21	2740	0.46	0.00039	CLDN18, PIK3R2, EGF, NKD1, FGF18, KIT, CXCL8, HAS2, CXCR2, WNT11, SOCS1, C5AR1, MAP2, ROS1, KLF4, NPPB, POSTN, CCL26, ITGB2, BEST1, PGF
GO:0048646	Anatomical structure formation involved in morphogenesis	12	883	0.71	0.00039	EGR2, EGF, FGF18, KNDC1, CXCL8, WNT11, HEY1, KLF4, ID1, ITGB2, PGF, NTF4
GO:0051241	Negative regulation of multicellular organismal process	14	1231	0.63	0.00039	CLDN18, HES1, ID2, WNT11, IL20RB, GP1BA, SOCS1, HEY1, HOXA9, MAP2, KLF4, ID1, NPPB, PBX1
GO:0032501	Multicellular organismal process	35	6933	0.28	0.0004	CLDN18, HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, MEGF10, FGF18, KIT, PKD1L1, KNDC1, FOS, CXCL8, HAS2, CXCR2, WNT11, IL20RB, GP1BA, HEY1, HOXA9, C5AR1, MAP2, ROS1, KLF4, ID1, NPPB, POSTN, ITGB2, BEST1, ETV1, PBX1, PGF, NTF4
GO:0050921	Positive regulation of chemotaxis	6	147	1.19	0.0004	FGF18, CXCL8, CXCR2, C5AR1, CCL26, PGF
GO:0048583	Regulation of response to stimulus	26	4114	0.38	0.00047	FGF22, PIK3R2, HES1, EGR1, EGF, NKD1, FGF18, KIT, FOS, CXCL8, CXCR2, WNT11, IL20RB, GP1BA, SOCS1, HEY1,

						C5AR1, ROS1, KLF4, ID1, POSTN, CCL26, ITGB2, MASP2, PGF, RAB7B
GO:0022008	Neurogenesis	16	1657	0.56	0.0005	HES1, ID2, EGR2, NKD1, KIT, KNDC1, WNT11, HEY1, C5AR1, MAP2, KLF4, ID1, POSTN, ETV1, PBX1, NTF4
GO:0051960	Regulation of nervous system development	12	942	0.68	0.00069	HES1, ID2, EGR2, EGF, KIT, KNDC1, HEY1, MAP2, KLF4, ID1, PBX1, NTF4
GO:0006928	Movement of cell or subcellular component	15	1501	0.58	0.0007	PIK3R2, HES1, EGR2, MEGF10, KIT, CXCL5, CXCL8, CXCR2, MMP1, WNT11, C5AR1, ID1, CCL26, ITGB2, ETV1
GO:0002684	Positive regulation of immune system process	12	949	0.68	0.00073	PIK3R2, HES1, ID2, FOS, CXCL8, CXCR2, SOCS1, C5AR1, ITGB2, MASP2, PGF, RAB7B
GO:0060412	Ventricular septum morphogenesis	4	42	1.55	0.0008	HES1, ID2, WNT11, HEY1
GO:0006954	Inflammatory response	9	515	0.82	0.001	KIT, CXCL5, FOS, CXCL8, CXCR2, IL20RB, C5AR1, CCL26, ITGB2
GO:0048699	Generation of neurons	15	1551	0.56	0.001	HES1, ID2, EGR2, NKD1, KIT, KNDC1, WNT11, HEY1, MAP2, KLF4, ID1, POSTN, ETV1, PBX1, NTF4
GO:0008593	Regulation of notch signaling pathway	5	103	1.26	0.0011	HES1, EGF, KIT, HEY1, POSTN
GO:0050767	Regulation of neurogenesis	11	828	0.7	0.0011	HES1, ID2, EGR2, KIT, KNDC1, HEY1, MAP2, KLF4, ID1, PBX1, NTF4
GO:1901700	Response to oxygen-containing compound	15	1567	0.56	0.0011	CLDN18, PIK3R2, EGR1, EGR2, CXCL5, FOS, CXCL8, WNT11, SOCS1, C5AR1, LCN2, KLF4, ID1, POSTN, GPT
GO:0030182	Neuron differentiation	12	1019	0.65	0.0014	HES1, ID2, EGR2, NKD1, KNDC1, WNT11, MAP2, ID1, POSTN, ETV1, PBX1, NTF4
GO:0035914	Skeletal muscle cell differentiation	4	51	1.47	0.0015	EGR1, EGR2, MEGF10, FOS
GO:0050789	Regulation of biological process	45	11475	0.17	0.0015	CLDN18, FGF22, PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, MEGF10, FGF18, KIT, CXCL5, KNDC1, FOS, CXCL8, HAS2, CXCR2, MMP1, WNT11, IL20RB, GP1BA, SOCS1, HIST2H2AC, HEY1, HOXA9, C5AR1, MAP2, ROS1, LCN2, KLF4, ID1, NPPB, POSTN, CCL26, GPT, ITGB2, MASP2, BEST1, ETV1, PBX1, PGF, NTF4, RAB7B
GO:0072359	Circulatory system development	11	872	0.68	0.0018	HES1, ID2, EGR1, EGF, FGF18, CXCL8, HAS2, WNT11, HEY1, ID1, PGF
GO:0060675	Ureteric bud morphogenesis	4	54	1.45	0.0019	HES1, WNT11, PBX1, PGF
GO:0043551	Regulation of phosphatidylinositol 3-kinase activity	4	57	1.42	0.0022	PIK3R2, KIT, SOCS1, KLF4
GO:0045596	Negative regulation of cell differentiation	10	728	0.71	0.0022	CLDN18, HES1, ID2, SOCS1, HEY1, HOXA9, MAP2, ID1, POSTN, PBX1

GO:0051090	Regulation of dna-binding transcription factor activity	8	437	0.84	0.0022	ID2, KIT, FOS, KLF4, ID1, ITGB2, PBX1, RAB7B
GO:0060255	Regulation of macromolecule metabolic process	32	6407	0.27	0.0022	PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, FGF18, KIT, KNDC1, FOS, HAS2, WNT11, IL20RB, SOCS1, HIST2H2AC, HEY1, HOXA9, C5AR1, ROS1, LCN2, KLF4, ID1, POSTN, CCL26, ITGB2, ETV1, PBX1, PGF, NTF4, RAB7B
GO:1903508	Positive regulation of nucleic acid-templated transcription	15	1670	0.53	0.0022	PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, FOS, WNT11, HEY1, HOXA9, KLF4, ID1, ETV1, PBX1
GO:0007399	Nervous system development	18	2371	0.46	0.0024	HES1, ID2, EGR2, NKD1, KIT, KNDC1, FOS, CXCR2, WNT11, HEY1, C5AR1, MAP2, KLF4, ID1, POSTN, ETV1, PBX1, NTF4
GO:0009612	Response to mechanical stimulus	6	212	1.03	0.0024	KIT, PKD1L1, FOS, WNT11, POSTN, ETV1
GO:0010557	Positive regulation of macromolecule biosynthetic process	16	1906	0.5	0.0024	PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, FOS, HAS2, WNT11, HEY1, HOXA9, KLF4, ID1, ETV1, PBX1
GO:0042327	Positive regulation of phosphorylation	12	1093	0.62	0.0025	HES1, EGR1, EGF, FGF18, KIT, KNDC1, WNT11, C5AR1, ROS1, CCL26, PGF, NTF4
GO:0048568	Embryonic organ development	8	448	0.83	0.0025	HES1, ID2, KIT, CXCL8, WNT11, HEY1, HOXA9, PBX1
GO:0072006	Nephron development	5	126	1.17	0.0025	HES1, EGR1, WNT11, PBX1, PGF
GO:0006952	Defense response	13	1296	0.58	0.0026	EGR1, KIT, CXCL5, FOS, CXCL8, CXCR2, IL20RB, C5AR1, LCN2, CCL26, ITGB2, MASP2, RAB7B
GO:0048565	Digestive tract development	5	128	1.17	0.0026	CLDN18, ID2, KIT, CXCL8, WNT11
GO:0060429	Epithelium development	12	1109	0.61	0.0027	HES1, ID2, HOXC13, EGF, CXCR2, WNT11, HEY1, ROS1, KLF4, ID1, PBX1, PGF
GO:0072073	Kidney epithelium development	5	130	1.16	0.0028	HES1, CXCR2, WNT11, PBX1, PGF
GO:0045665	Negative regulation of neuron differentiation	6	222	1.01	0.0029	HES1, ID2, HEY1, MAP2, ID1, PBX1
GO:0035019	Somatic stem cell population maintenance	4	65	1.36	0.0031	HES1, KIT, KLF4, PBX1
GO:0051896	Regulation of protein kinase b signaling	6	226	1	0.0031	FGF22, PIK3R2, EGF, FGF18, KIT, KLF4
GO:0019222	Regulation of metabolic process	33	6948	0.25	0.0035	PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, FGF18, KIT, KNDC1, FOS, HAS2, WNT11, IL20RB, SOCS1, HIST2H2AC, HEY1, HOXA9, C5AR1, ROS1, LCN2, KLF4, ID1, POSTN, CCL26, GPT, ITGB2, ETV1, PBX1, PGF, NTF4, RAB7B
GO:0061061	Muscle structure development	8	479	0.8	0.0035	HES1, EGR1, EGR2, NKD1, MEGF10, FOS, HEY1, ETV1
GO:0050794	Regulation of cellular process	43	10932	0.17	0.0037	CLDN18, FGF22, PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, MEGF10, FGF18, KIT, CXCL5, KNDC1, FOS,

						CXCL8, HAS2, CXCR2, MMP1, WNT11, IL20RB, GP1BA, SOCS1, HIST2H2AC, HEY1, HOXA9, C5AR1, MAP2, ROS1, LCN2, KLF4, ID1, NPPB, POSTN, CCL26, GPT, ITGB2, ETV1, PBX1, PGF, NTF4, RAB7B
GO:0043586	Tongue development	3	23	1.69	0.0038	HOXC13, KIT, NTF4
GO:1903706	Regulation of hemopoiesis	8	493	0.79	0.0041	CLDN18, HES1, ID2, FOS, GP1BA, SOCS1, HOXA9, RAB7B
GO:0045893	Positive regulation of transcription, dna-templated	14	1587	0.52	0.0043	PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, FOS, WNT11, HEY1, HOXA9, KLF4, ETV1, PBX1
GO:0071356	Cellular response to tumor necrosis factor	6	245	0.96	0.0043	CLDN18, CXCL8, HAS2, LCN2, POSTN, CCL26
GO:0080090	Regulation of primary metabolic process	30	6032	0.27	0.0046	PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, FGF18, KIT, KNDC1, FOS, HAS2, WNT11, SOCS1, HIST2H2AC, HEY1, HOXA9, C5AR1, ROS1, KLF4, ID1, CCL26, GPT, ITGB2, ETV1, PBX1, PGF, NTF4, RAB7B
GO:0048519	Negative regulation of biological process	28	5389	0.29	0.0047	CLDN18, PIK3R2, HES1, ID2, EGR1, EGF, NKD1, KIT, CXCL8, CXCR2, WNT11, IL20RB, GP1BA, SOCS1, HIST2H2AC, HEY1, HOXA9, C5AR1, MAP2, ROS1, KLF4, ID1, NPPB, POSTN, ITGB2, PBX1, NTF4, RAB7B
GO:0007622	Rhythmic behavior	3	26	1.64	0.005	ID2, EGR1, EGR2
GO:0045664	Regulation of neuron differentiation	9	665	0.71	0.005	HES1, ID2, KNDC1, HEY1, MAP2, KLF4, ID1, PBX1, NTF4
GO:0051098	Regulation of binding	7	373	0.85	0.005	HES1, ID2, EGF, HEY1, MAP2, KLF4, ID1
GO:0090023	Positive regulation of neutrophil chemotaxis	3	26	1.64	0.005	CXCL8, CXCR2, C5AR1
GO:0001934	Positive regulation of protein phosphorylation	11	1019	0.61	0.0052	HES1, EGR1, EGF, FGF18, KIT, KNDC1, WNT11, C5AR1, CCL26, PGF, NTF4
GO:0007169	Transmembrane receptor protein tyrosine kinase signaling pathway	8	518	0.76	0.0052	FGF22, PIK3R2, EGF, FGF18, KIT, ROS1, PGF, NTF4
GO:0048468	Cell development	14	1629	0.51	0.0052	HES1, ID2, EGR2, MEGF10, FGF18, KIT, C5AR1, MAP2, ROS1, ID1, POSTN, ETV1, PBX1, NTF4
GO:1903707	Negative regulation of hemopoiesis	5	156	1.08	0.0052	CLDN18, HES1, ID2, SOCS1, HOXA9
GO:0098759	Cellular response to interleukin-8	2	3	2.4	0.0053	EGR1, CXCR2
GO:0045637	Regulation of myeloid cell differentiation	6	260	0.94	0.0054	CLDN18, ID2, FOS, GP1BA, HOXA9, RAB7B
GO:0070098	Chemokine-mediated signaling pathway	4	80	1.27	0.0055	CXCL5, CXCL8, CXCR2, CCL26
GO:0001656	Metanephros development	4	84	1.25	0.0064	HES1, ID2, EGR1, CXCR2

GO:0048523	Negative regulation of cellular process	26	4874	0.3	0.0064	CLDN18, PIK3R2, HES1, ID2, EGR1, EGF, NKD1, KIT, CXCL8, CXCR2, WNT11, IL20RB, SOCS1, HIST2H2AC, HEY1, HOXA9, C5AR1, MAP2, KLF4, ID1, NPPB, POSTN, ITGB2, PBX1, NTF4, RAB7B
GO:0051171	Regulation of nitrogen compound metabolic process	29	5836	0.27	0.0064	PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, NKD1, FGF18, KIT, KNDC1, FOS, HAS2, WNT11, SOCS1, HIST2H2AC, HEY1, HOXA9, C5AR1, ROS1, KLF4, ID1, CCL26, ITGB2, ETV1, PBX1, PGF, NTF4, RAB7B
GO:1901701	Cellular response to oxygen-containing compound	11	1055	0.59	0.0065	PIK3R2, EGR1, CXCL5, FOS, CXCL8, WNT11, SOCS1, LCN2, KLF4, ID1, GPT
GO:0051897	Positive regulation of protein kinase b signaling	5	170	1.04	0.007	FGF22, PIK3R2, EGF, FGF18, KIT
GO:0030155	Regulation of cell adhesion	9	712	0.68	0.0072	HES1, MEGF10, CXCL8, HAS2, IL20RB, SOCS1, KLF4, POSTN, ITGB2
GO:1900127	Positive regulation of hyaluronan biosynthetic process	2	4	2.27	0.0072	EGF, HAS2
GO:0001775	Cell activation	11	1075	0.59	0.0073	ID2, EGR1, MEGF10, KIT, CXCL8, CXCR2, GP1BA, HOXA9, C5AR1, LCN2, ITGB2
GO:0071347	Cellular response to interleukin-1	5	174	1.03	0.0074	EGR1, CXCL8, HAS2, LCN2, CCL26
GO:0000902	Cell morphogenesis	9	726	0.67	0.0078	HES1, ID2, EGR2, GP1BA, MAP2, ID1, POSTN, ETV1, NTF4
GO:0006950	Response to stress	21	3485	0.36	0.0078	PIK3R2, ID2, EGR1, KIT, CXCL5, FOS, CXCL8, HAS2, CXCR2, IL20RB, GP1BA, C5AR1, LCN2, KLF4, POSTN, CCL26, GPT, ITGB2, MASP2, PGF, RAB7B
GO:0043408	Regulation of mapk cascade	9	725	0.67	0.0078	PIK3R2, EGF, FGF18, KIT, C5AR1, ROS1, KLF4, ID1, CCL26
GO:0048729	Tissue morphogenesis	8	561	0.73	0.0078	HES1, EGF, CXCR2, WNT11, HEY1, KLF4, PBX1, PGF
GO:0010646	Regulation of cell communication	21	3514	0.35	0.0086	FGF22, PIK3R2, HES1, EGR1, EGR2, EGF, NKD1, FGF18, KIT, CXCL8, WNT11, SOCS1, HEY1, C5AR1, ROS1, KLF4, ID1, POSTN, CCL26, NTF4, RAB7B
GO:0022407	Regulation of cell-cell adhesion	7	424	0.79	0.0086	HES1, MEGF10, HAS2, IL20RB, SOCS1, KLF4, ITGB2
GO:0043549	Regulation of kinase activity	10	918	0.61	0.0088	PIK3R2, EGR1, EGF, FGF18, KIT, WNT11, SOCS1, C5AR1, ROS1, KLF4
GO:0009987	Cellular process	50	15024	0.1	0.0096	CLDN18, FGF22, PIK3R2, HES1, ID2, EGR1, EGR2, EGF, NKD1, MEGF10, FGF18, MMP10, KIT, PKD1L1, U2AFBP, CXCL5, KNDC1, FOS, CXCL8, HAS2, CXCR2, MMP1, WNT11, IL20RB, GP1BA, SOCS1, HIST2H2AC, HEY1, HOXA9, C5AR1, MAP2, ROS1, LCN2, KLF4, ID1, NPPB, POSTN, AMACR, ALPP, CCL26,

						GPT, ITGB2, BEST1, ETV1, PBX1, PGF, NTF4, HIST1H2BO, HIST1H2BH, RAB7B
GO:0002690	Positive regulation of leukocyte chemotaxis	4	98	1.19	0.0097	CXCL8, CXCR2, C5AR1, PGF
GO:0002009	Morphogenesis of an epithelium	7	435	0.78	0.0098	HES1, EGF, CXCR2, WNT11, KLF4, PBX1, PGF
GO:0023051	Regulation of signaling	21	3553	0.35	0.0098	FGF22, PIK3R2, HES1, EGR1, EGR2, EGF, NKD1, FGF18, KIT, CXCL8, WNT11, SOCS1, HEY1, C5AR1, ROS1, KLF4, ID1, POSTN, CCL26, NTF4, RAB7B
GO:0050679	Positive regulation of epithelial cell proliferation	5	192	0.99	0.0106	HAS2, C5AR1, ID1, CCL26, PGF
GO:0060562	Epithelial tube morphogenesis	6	307	0.87	0.0106	HES1, EGF, CXCR2, WNT11, PBX1, PGF
GO:0002683	Negative regulation of immune system process	7	450	0.77	0.0117	CLDN18, HES1, ID2, CXCR2, IL20RB, SOCS1, HOXA9
GO:0001525	Angiogenesis	6	315	0.86	0.012	EGF, FGF18, CXCL8, HEY1, ID1, PGF
GO:1903037	Regulation of leukocyte cell-cell adhesion	6	315	0.86	0.012	HES1, HAS2, IL20RB, SOCS1, KLF4, ITGB2
GO:1902895	Positive regulation of pri-mirna transcription by rna polymerase ii	3	40	1.45	0.0121	EGR1, FOS, KLF4
GO:0006955	Immune response	13	1588	0.49	0.0125	EGR1, KIT, CXCL5, CXCL8, CXCR2, IL20RB, C5AR1, LCN2, NPPB, CCL26, ITGB2, MASP2, RAB7B
GO:0010468	Regulation of gene expression	25	4813	0.29	0.0125	PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, KIT, FOS, WNT11, IL20RB, SOCS1, HIST2H2AC, HEY1, HOXA9, C5AR1, ROS1, LCN2, KLF4, ID1, POSTN, ITGB2, ETV1, PBX1, RAB7B
GO:0044344	Cellular response to fibroblast growth factor stimulus	4	108	1.14	0.0129	FGF22, FGF18, CXCL8, POSTN
GO:0051246	Regulation of protein metabolic process	18	2828	0.38	0.0136	PIK3R2, HES1, EGR1, EGF, NKD1, FGF18, KIT, KNDC1, WNT11, SOCS1, C5AR1, ROS1, KLF4, ID1, CCL26, ITGB2, PGF, NTF4
GO:0009966	Regulation of signal transduction	19	3107	0.36	0.0142	FGF22, PIK3R2, HES1, EGR1, EGF, NKD1, FGF18, KIT, CXCL8, WNT11, SOCS1, HEY1, C5AR1, ROS1, KLF4, ID1, POSTN, CCL26, RAB7B
GO:0010243	Response to organonitrogen compound	10	987	0.58	0.0142	PIK3R2, EGR1, EGR2, FOS, SOCS1, C5AR1, KLF4, ID1, GPT, ITGB2
GO:0045687	Positive regulation of glial cell differentiation	3	43	1.42	0.0142	HES1, ID2, EGR2
GO:0009790	Embryo development	10	1002	0.57	0.0159	HES1, ID2, KIT, CXCL8, WNT11, HEY1, HOXA9, KLF4, ITGB2, PBX1

GO:0009952	Anterior/posterior pattern specification	5	214	0.94	0.0159	HES1, HOXC13, HEY1, HOXA9, PBX1
GO:0050678	Regulation of epithelial cell proliferation	6	339	0.82	0.0164	HES1, HAS2, C5AR1, ID1, CCL26, PGF
GO:0010863	Positive regulation of phospholipase c activity	3	46	1.39	0.0167	KIT, C5AR1, NTF4
GO:0032101	Regulation of response to external stimulus	10	1013	0.57	0.017	FGF18, CXCL8, CXCR2, IL20RB, GP1BA, SOCS1, C5AR1, KLF4, CCL26, PGF
GO:0031323	Regulation of cellular metabolic process	29	6239	0.24	0.0179	PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, FGF18, KIT, KNDC1, FOS, HAS2, WNT11, SOCS1, HIST2H2AC, HEY1, HOXA9, C5AR1, ROS1, KLF4, ID1, CCL26, GPT, ITGB2, ETV1, PBX1, PGF, NTF4, RAB7B
GO:0022604	Regulation of cell morphogenesis	7	498	0.72	0.0194	KIT, KNDC1, HAS2, MAP2, ID1, POSTN, ITGB2
GO:1904796	Regulation of core promoter binding	2	9	1.92	0.0202	ID2, KLF4
GO:0045944	Positive regulation of transcription by rna polymerase ii	11	1253	0.52	0.0212	PIK3R2, HES1, EGR1, EGR2, HOXC13, FOS, HEY1, HOXA9, KLF4, ETV1, PBX1
GO:0051247	Positive regulation of protein metabolic process	13	1715	0.46	0.0232	HES1, EGR1, EGF, NKD1, FGF18, KIT, KNDC1, WNT11, C5AR1, KLF4, CCL26, PGF, NTF4
GO:0046425	Regulation of receptor signaling pathway via jak-stat	4	133	1.05	0.0246	HES1, EGF, KIT, SOCS1
GO:0043392	Negative regulation of dna binding	3	56	1.3	0.0263	ID2, HEY1, ID1
GO:2000677	Regulation of transcription regulatory region dna binding	3	56	1.3	0.0263	ID2, HEY1, KLF4
GO:0072173	Metanephric tubule morphogenesis	2	11	1.84	0.0268	HES1, CXCR2
GO:0051347	Positive regulation of transferase activity	8	705	0.63	0.0269	EGR1, EGF, FGF18, KIT, WNT11, C5AR1, ROS1, KLF4
GO:0045165	Cell fate commitment	5	248	0.88	0.0272	HES1, ID2, WNT11, KLF4, NTF4
GO:0003007	Heart morphogenesis	5	251	0.87	0.0286	HES1, ID2, HAS2, WNT11, HEY1
GO:0043388	Positive regulation of dna binding	3	59	1.28	0.0294	HES1, EGF, KLF4
GO:0048584	Positive regulation of response to stimulus	15	2257	0.4	0.0294	FGF22, PIK3R2, HES1, EGF, NKD1, FGF18, KIT, CXCL8, CXCR2, WNT11, C5AR1, CCL26, ITGB2, MASP2, PGF
GO:0045475	Locomotor rhythm	2	12	1.8	0.0302	ID2, EGR1
GO:0048711	Positive regulation of astrocyte differentiation	2	12	1.8	0.0302	HES1, ID2
GO:0050730	Regulation of peptidyl-tyrosine phosphorylation	5	258	0.86	0.0315	HES1, EGF, KIT, SOCS1, ITGB2

GO:0048844	Artery morphogenesis	3	61	1.27	0.0317	HES1, WNT11, HEY1
GO:0007155	Cell adhesion	9	925	0.56	0.0338	CLDN18, HES1, MEGF10, KIT, PKD1L1, GP1BA, ID1, POSTN, ITGB2
GO:0048715	Negative regulation of oligodendrocyte differentiation	2	13	1.76	0.0339	HES1, ID2
GO:1902947	Regulation of tau-protein kinase activity	2	13	1.76	0.0339	EGR1, C5AR1
GO:0032963	Collagen metabolic process	3	63	1.25	0.034	MMP10, MMP1, ID1
GO:0007423	Sensory organ development	7	563	0.67	0.0344	HES1, HOXC13, NKD1, KIT, KLF4, PBX1, NTF4
GO:0001817	Regulation of cytokine production	8	742	0.61	0.0351	EGR1, WNT11, IL20RB, SOCS1, C5AR1, KLF4, POSTN, RAB7B
GO:0006355	Regulation of transcription, dna-templated	19	3388	0.32	0.0351	PIK3R2, HES1, ID2, EGR1, EGR2, HOXC13, EGF, KIT, FOS, WNT11, HIST2H2AC, HEY1, HOXA9, KLF4, ID1, ITGB2, ETV1, PBX1, RAB7B
GO:0031623	Receptor internalization	3	64	1.25	0.0351	CXCL8, CXCR2, ITGB2
GO:0007420	Brain development	8	745	0.61	0.0356	HES1, ID2, EGR2, KNDC1, CXCR2, C5AR1, ID1, PBX1
GO:0048762	Mesenchymal cell differentiation	4	152	1	0.0359	HES1, HAS2, WNT11, HEY1
GO:0032922	Circadian regulation of gene expression	3	65	1.24	0.036	ID2, EGR1, ID1
GO:0007267	Cell-cell signaling	10	1145	0.52	0.0361	HES1, EGF, NKD1, FGF18, CXCL5, WNT11, KLF4, CCL26, ITGB2, PGF
GO:0048598	Embryonic morphogenesis	7	571	0.66	0.0361	HES1, ID2, WNT11, HOXA9, KLF4, ITGB2, PBX1
GO:0048732	Gland development	6	410	0.74	0.0361	HES1, ID2, EGF, WNT11, HOXA9, PBX1
GO:0007610	Behavior	7	572	0.66	0.0362	ID2, EGR1, EGR2, KIT, FOS, ETV1, NTF4
GO:0009628	Response to abiotic stimulus	10	1147	0.52	0.0363	ID2, EGR1, KIT, PKD1L1, FOS, WNT11, POSTN, BEST1, ETV1, PGF
GO:0048511	Rhythmic process	5	271	0.84	0.0364	ID2, EGR1, EGR2, HAS2, ID1
GO:0097242	Amyloid-beta clearance	2	14	1.73	0.0364	C5AR1, ITGB2
GO:0030902	Hindbrain development	4	154	0.99	0.0365	HES1, EGR2, KNDC1, C5AR1
GO:0022409	Positive regulation of cell-cell adhesion	5	272	0.84	0.0366	HES1, MEGF10, HAS2, SOCS1, ITGB2
GO:0006959	Humoral immune response	5	275	0.84	0.0383	CXCL5, CXCL8, LCN2, NPPB, MASP2
GO:0044093	Positive regulation of molecular function	13	1842	0.42	0.0387	HES1, EGR1, EGF, FGF18, KIT, WNT11, C5AR1, ROS1, KLF4, CCL26, ITGB2, NTF4, RAB7B
GO:0035810	Positive regulation of urine volume	2	15	1.7	0.0405	HAS2, NPPB

GO:0044849	Estrous cycle	2	15	1.7	0.0405	EGR1, HAS2
GO:0050930	Induction of positive chemotaxis	2	15	1.7	0.0405	CXCL8, PGF
GO:0014706	Striated muscle tissue development	5	280	0.83	0.0407	ID2, EGR1, EGR2, MEGF10, FOS
GO:0019216	Regulation of lipid metabolic process	6	424	0.73	0.0407	PIK3R2, ID2, EGR1, KIT, SOCS1, KLF4
GO:0019730	Antimicrobial humoral response	4	160	0.97	0.0407	CXCL5, CXCL8, LCN2, NPPB
GO:0001837	Epithelial to mesenchymal transition	3	70	1.21	0.0413	HAS2, WNT11, HEY1
GO:0032270	Positive regulation of cellular protein metabolic process	12	1635	0.44	0.0439	HES1, EGR1, EGF, FGF18, KIT, KNDC1, WNT11, C5AR1, KLF4, CCL26, PGF, NTF4
GO:0048585	Negative regulation of response to stimulus	12	1636	0.44	0.0439	PIK3R2, EGR1, EGF, NKD1, CXCL8, WNT11, IL20RB, GP1BA, SOCS1, HEY1, KLF4, RAB7B
GO:2000341	Regulation of chemokine (c-x-c motif) ligand 2 production	2	16	1.67	0.0439	KLF4, POSTN
GO:0043433	Negative regulation of dna-binding transcription factor activity	4	166	0.96	0.045	ID2, KLF4, ID1, PBX1
GO:0007517	Muscle organ development	5	291	0.81	0.046	EGR1, EGR2, MEGF10, FOS, ETV1
GO:0007417	Central nervous system development	9	988	0.53	0.0465	HES1, ID2, EGR2, KNDC1, CXCR2, C5AR1, MAP2, ID1, PBX1
GO:0070372	Regulation of erk1 and erk2 cascade	5	292	0.81	0.0465	FGF18, C5AR1, ROS1, KLF4, CCL26
GO:0021783	Preganglionic parasympathetic fiber development	2	17	1.65	0.0475	HES1, EGR2
GO:0045597	Positive regulation of cell differentiation	9	993	0.53	0.0476	HES1, ID2, EGR2, FGF18, KIT, FOS, HAS2, SOCS1, RAB7B
GO:0007422	Peripheral nervous system development	3	75	1.18	0.0477	EGR2, ETV1, NTF4
GO:0030324	Lung development	4	170	0.95	0.0477	HES1, FGF18, WNT11, ID1

Table S2.7: List of functionally enriched top dysregulated genes identified by using g:GOSt module of g:Profiler.

source	term name	term id	adjusted p.value	-log10 (adj.p.value)	term size	query size	intersection size	effective domain size	intersections
GO:MF	transcription regulator inhibitor activity	GO:0140416	0.027928792	1.553947848	21	14	2	20166	ID1, ID2
GO:MF	heparin binding	GO:0008201	0.038609657	1.413304061	173	12	3	20166	PGF, CXCL8, SMOC1

GO:MF	hemi-methylated DNA-binding	GO:0044729	0.044540494	1.351244975	3	1	1	20166	EGR1
GO:BP	regulation of multicellular organismal process	GO:0051239	0.000172675	3.762770177	2753	23	14	21092	EGR1, ID1, LAPTM5, PGF, KLF4, HES1, CXCL8, RGS2, ID2, ZFHX2, KCNJ12, NPPB, TGM2, OAS2
GO:BP	circulatory system development	GO:0072359	0.0002147	3.668167752	1109	14	8	21092	EGR1, ID1, PGF, KLF4, HES1, CXCL8, RGS2, ID2
GO:BP	positive regulation of multicellular organismal process	GO:0051240	0.000241222	3.6175827	1507	23	11	21092	EGR1, LAPTM5, PGF, KLF4, HES1, CXCL8, RGS2, ID2, NPPB, TGM2, OAS2
GO:BP	animal organ development	GO:0048513	0.000586177	3.231970953	3591	23	15	21092	EGR1, ID1, CYP1A1, PGF, KLF4, HES1, CXCL8, SMOC1, RGS2, ID2, KRT81, ZFHX2, NPPB, TGM2, OAS2
GO:BP	tube development	GO:0035295	0.000635593	3.196821073	1076	21	9	21092	ID1, CYP1A1, PGF, KLF4, HES1, CXCL8, ID2, NPPB, TGM2
GO:BP	cellular response to growth factor stimulus	GO:0071363	0.000693193	3.159146089	673	11	6	21092	EGR1, ID1, PGF, KLF4, HES1, CXCL8
GO:BP	response to growth factor	GO:0070848	0.000888304	3.051438291	702	11	6	21092	EGR1, ID1, PGF, KLF4, HES1, CXCL8
GO:BP	blood vessel development	GO:0001568	0.000895764	3.047806365	703	11	6	21092	EGR1, ID1, PGF, KLF4, HES1, CXCL8
GO:BP	regulation of developmental process	GO:0050793	0.001028692	2.987714837	2499	24	13	21092	ID1, PGF, KLF4, HES1, CXCL8, SMOC1, RGS2, ID2, ZFHX2, NPPB, TGM2, OAS2, UNC13A
GO:BP	vasculature development	GO:0001944	0.001154039	2.937779531	734	11	6	21092	EGR1, ID1, PGF, KLF4, HES1, CXCL8
GO:BP	regulation of multicellular organismal development	GO:2000026	0.001325223	2.877710929	1394	23	10	21092	ID1, PGF, KLF4, HES1, CXCL8, RGS2, ID2, NPPB, TGM2, OAS2
GO:BP	tube morphogenesis	GO:0035239	0.001604217	2.794736841	863	21	8	21092	ID1, PGF, KLF4, HES1, CXCL8, ID2, NPPB, TGM2
GO:BP	digestive tract development	GO:0048565	0.002400413	2.619714073	133	14	4	21092	CYP1A1, HES1, CXCL8, ID2
GO:BP	digestive system development	GO:0055123	0.00329635	2.481966693	144	14	4	21092	CYP1A1, HES1, CXCL8, ID2
GO:BP	cellular response to organic substance	GO:0071310	0.005652297	2.247775034	2405	11	8	21092	EGR1, ID1, CYP1A1, LAPTM5, PGF, KLF4, HES1, CXCL8
GO:BP	positive regulation of gene expression	GO:0010628	0.006006658	2.221367121	1158	14	7	21092	EGR1, ID1, LAPTM5, KLF4, HES1, CXCL8, ID2
GO:BP	negative regulation of DNA-binding transcription factor activity	GO:0043433	0.006839915	2.164949275	173	14	4	21092	ID1, KLF4, HES1, ID2

GO:BP	regulation of vasculature development	GO:1901342	0.008122927	2.090287448	292	19	5	21092	ID1, PGF, KLF4, CXCL8, NPPB
GO:BP	response to oxygen-containing compound	GO:1901700	0.009181164	2.037102248	1655	15	8	21092	EGR1, ID1, CYP1A1, KLF4, HES1, CXCL8, RGS2, KCNB1
GO:BP	response to organic substance	GO:0010033	0.009338161	2.029738651	3034	15	10	21092	EGR1, ID1, CYP1A1, LAPTM5, PGF, KLF4, HES1, CXCL8, RGS2, KCNB1
GO:BP	multicellular organism development	GO:0007275	0.009792458	2.00910826	4823	24	16	21092	EGR1, ID1, CYP1A1, PGF, KLF4, HES1, CXCL8, SMOC1, RGS2, ID2, KCNB1, ZFHX2, NPPB, TGM2, OAS2, UNC13A
GO:BP	tissue development	GO:0009888	0.011207728	1.950482412	1973	21	10	21092	EGR1, ID1, CYP1A1, KLF4, HES1, RGS2, ID2, KRT81, NPPB, TGM2
GO:BP	system development	GO:0048731	0.011453118	1.941076264	4369	21	14	21092	EGR1, ID1, CYP1A1, PGF, KLF4, HES1, CXCL8, SMOC1, RGS2, ID2, KCNB1, ZFHX2, NPPB, TGM2
GO:BP	defense response to tumor cell	GO:0002357	0.013821079	1.859458042	12	8	2	21092	LAPTM5, KLF4
GO:BP	blood vessel morphogenesis	GO:0048514	0.013911092	1.856638777	616	11	5	21092	ID1, PGF, KLF4, HES1, CXCL8
GO:BP	epithelial cell differentiation	GO:0030855	0.014356999	1.842936337	711	16	6	21092	ID1, CYP1A1, KLF4, HES1, ID2, KRT81
GO:BP	cell differentiation	GO:0030154	0.01665366	1.778490318	4256	17	12	21092	EGR1, ID1, CYP1A1, PGF, KLF4, HES1, SMOC1, RGS2, ID2, KCNB1, KRT81, ZFHX2
GO:BP	cellular developmental process	GO:0048869	0.017703852	1.751932225	4280	17	12	21092	EGR1, ID1, CYP1A1, PGF, KLF4, HES1, SMOC1, RGS2, ID2, KCNB1, KRT81, ZFHX2
GO:BP	cell population proliferation	GO:0008283	0.01827144	1.738227221	1988	14	8	21092	EGR1, ID1, CYP1A1, PGF, KLF4, HES1, CXCL8, ID2
GO:BP	circadian regulation of gene expression	GO:0032922	0.018620451	1.730009796	71	2	2	21092	EGR1, ID1
GO:BP	anatomical structure development	GO:0048856	0.022527398	1.647288967	5836	24	17	21092	EGR1, ID1, CYP1A1, PGF, KLF4, HES1, CXCL8, SMOC1, RGS2, ID2, KCNB1, KRT81, ZFHX2, NPPB, TGM2, OAS2, UNC13A
GO:BP	fat cell differentiation	GO:0045444	0.022955881	1.639106033	235	14	4	21092	KLF4, HES1, RGS2, ID2
GO:BP	multicellular organismal process	GO:0032501	0.024319384	1.614047432	7463	24	19	21092	EGR1, ID1, CYP1A1, LAPTM5, PGF, KLF4, HES1, CXCL8, SMOC1, RGS2, ID2, KCNB1, KRT81, ZFHX2,

									KCNJ12, NPPB, TGM2, OAS2, UNC13A
GO:BP	cellular response to interleukin-1	GO:0071347	0.032947986	1.482171126	106	11	3	21092	EGR1, HES1, CXCL8
GO:BP	negative regulation of multicellular organismal process	GO:0051241	0.033512036	1.474799187	1035	19	7	21092	ID1, LAPTM5, KLF4, HES1, RGS2, ID2, NPPB
GO:BP	cellular response to chemical stimulus	GO:0070887	0.03450026	1.462177628	3049	11	8	21092	EGR1, ID1, CYP1A1, LAPTM5, PGF, KLF4, HES1, CXCL8
GO:BP	regulation of DNA binding	GO:0051101	0.035766542	1.446523042	121	10	3	21092	ID1, KLF4, HES1
GO:BP	induction of positive chemotaxis	GO:0050930	0.043113111	1.365390639	15	11	2	21092	PGF, CXCL8
GO:BP	positive regulation of cellular process	GO:0048522	0.043160301	1.364915536	5641	15	12	21092	EGR1, ID1, MMP1, CYP1A1, LAPTM5, PGF, KLF4, HES1, CXCL8, RGS2, ID2, KCNB1
REAC	NGF-stimulated transcription	REAC:R-HSA-9031628	0.003960201	2.402282789	39	2	2	10461	EGR1, ID1
REAC	Nuclear Events (kinase and transcription factor activation)	REAC:R-HSA-198725	0.009780253	2.009649907	61	2	2	10461	EGR1, ID1
REAC	Signaling by NTRK1 (TRKA)	REAC:R-HSA-187037	0.033220794	1.478589991	112	2	2	10461	EGR1, ID1
REAC	Signaling by NTRKs	REAC:R-HSA-166520	0.045507571	1.341916345	131	2	2	10461	EGR1, ID1
WP	miR-517 relationship with ARCN1 and USP1	WP:WP3596	0.003226717	2.491239131	5	14	2	7827	ID1, ID2
WP	let-7 inhibition of ES cell reprogramming	WP:WP3299	0.010387526	1.983487869	15	8	2	7827	EGR1, KLF4
WP	Neurogenesis regulation in the olfactory epithelium	WP:WP5265	0.013718365	1.862697648	57	14	3	7827	ID1, HES1, ID2
WP	Oncostatin M signaling pathway	WP:WP2374	0.022074934	1.656100589	65	3	2	7827	EGR1, MMP1
WP	ID signaling pathway	WP:WP53	0.038287482	1.416943195	16	14	2	7827	ID1, ID2
WP	Burn wound healing	WP:WP5055	0.048405062	1.31510922	113	11	3	7827	MMP1, KLF4, CXCL8

WP	Gastrin signaling pathway	WP:WP4659	0.049674972	1.303862369	114	11	3	7827	EGR1, KLF4, CXCL8

Table S3.1: Primers used for quantification of miRNAs through qPCR.

Targets	Species	Primer	Sequence	Annealing Temperature	Supplier	Purification
snRNA U6	Mus musculus	Forward	CGGCAGCACATATACTAAAATTGG	50	IDT	Standard Desalting
		Reverse	GCCATGCTAATCTTCTCTGTATC		IDT	Standard Desalting
snoRNA U49	Mus musculus	Forward	ATCACTAATAGGAAGTGCCGTC	50	IDT	Standard Desalting
		Reverse	ACAGGAGTAGTCTTCGTCAGT		IDT	Standard Desalting
mmu-miR-181a-5p	Mus musculus	Forward	AACATTCAACGCTGTC	50	IDT	Standard Desalting
		Reverse	GTAAAACGACGGCCAGTACTCACC		IDT	Standard Desalting
mmu-miR-181b-5p	Mus musculus	Forward	AACATTCATTGCTGTCG	50	IDT	Standard Desalting
		Reverse	GTAAAACGACGGCCAGTAACCCAC		IDT	Standard Desalting
mmu-miR-181c-5p	Mus musculus	Forward	AA+CATTCAACCTGTC	50	Exiqon	Standard Desalting
		Reverse	GTAAAACGACGGCCAGTACTCACC		IDT	Standard Desalting
mmu-miR-30b-5p	Mus musculus	Forward	T+GT+AAACATCCTACA	50	Exiqon	Standard Desalting
		Reverse	GTAAAACGACGGCCAGTAGCTGAG		IDT	Standard Desalting
mmu-let-7b-5p	Mus musculus	Forward	TG+AGGTAGTAGGTTG	50	Exiqon	Standard Desalting
		Reverse	GTAAAACGACGGCCAGTAACCACA		IDT	Standard Desalting
mmu-miR-150-5p	Mus musculus	Forward	TCTCCCAACCCTTGT	50	IDT	Standard Desalting

		Reverse	GTAAAACGACGGCCAGTCACTGGT		IDT	Standard Desalting
mmu-miR-151-5p	Mus musculus	Forward	TCGAGGAGCTCACA	50	IDT	Standard Desalting
		Reverse	GTAAAACGACGGCCAGTACTAGAC		IDT	Standard Desalting

Table S3.2: Primers used for quantification of genes through qPCR.

Targets	Species	Forward Sequence 5' à 3' (For)	Reverse Sequence (Rev)	Annealing Temperature	Design
Hprt	*Mus musculus*	AGGCCAGACTTTGTTGGATTTGAA	CAACTTGCGCTCATCTTAGGCTTT	55-65	Custom
Arnt2	*Mus musculus*	GCAGTAAATATTGCCTCGTG	AATATCCTTTCCCAGAAGGTC	56	Custom
Bbox1	*Mus musculus*	GACATTGGAGTGGATTACTG	CTTGGCCTTTATCATCCAAC	56	Predesign-KiCqStart
Kcnmb2	*Mus musculus*	AGATCAATCAAAAGTGCTCC	TTTCCTTCTGGGTCAGAATAG	56	Predesign-KiCqStart
Mfsd4	*Mus musculus*	AGTTACAGTTTCCTGGTCTG	GTAAGAACTGATGAGAGTTCAG	56	Predesign-KiCqStart
Myb	*Mus musculus*	CACAAAACATCTCCAGTCAC	TCTTCGTCGTTATAGTGTCTC	56	Predesign-KiCqStart
Slco2b1	*Mus musculus*	ACATGAGTTTCATTTCAGGC	AAGACTTTGACAAACTGGAC	56	Predesign-KiCqStart
Ttc25	*Mus musculus*	AGAGTTATGGGGAAAAGTCTC	GGCCTTCTCAAAGTTATTCAC	56	Predesign-KiCqStart

Table S3.3: Differentially expressed miRNAs identified during longitudinal cigarette smoke (CS) exposure in mice from 4 to 12 weeks.

miRNA	logFC	AveExpr	t	P.Value	adj.P.Val	B	Legend
miR-195-5p	-0.46066667	11.96480645	-5.70088658	6.26E-06	2.58E-03	3.96201538	downregulated
miR-449a-5p	1.64633333	7.615258065	5.60727424	7.94E-06	0.00258138	3.73991621	upregulated
miR-10a-3p	-0.59983333	4.367645161	-5.35414588	1.52E-05	3.29E-03	3.13574934	downregulated

miR-676-3p	-0.81125	5.959741935	-4.37609176	1.89E-04	3.08E-02	0.77504099	downregulated
miR-135b-5p	1.19708333	5.897451613	4.23799896	2.71E-04	0.035167832	0.44222568	upregulated
miR-23a-3p	-0.36141667	14.35093548	-4.12481537	3.62E-04	3.75E-02	0.17032389	downregulated
miR-150-5p	-1.23966667	8.103032258	-4.01391229	4.81E-04	3.75E-02	-0.09508876	downregulated
miR-30d-5p	-0.24291667	11.72932258	-3.98031142	5.25E-04	3.75E-02	-0.17526948	downregulated
miR-30b-5p	-0.47075	11.89032258	-3.95263453	5.63E-04	3.75E-02	-0.24122419	downregulated
miR-10a-5p	-0.68533333	11.56322581	-3.94347937	5.77E-04	0.037479784	-0.2630226	downregulated
miR-1904	-0.7775	5.756645161	-3.8742644	6.88E-04	0.040665446	-0.42750719	downregulated
miR-99a-5p	-0.50216667	10.37996774	-3.80027814	8.31E-04	0.041898302	-0.60266319	downregulated
miR-100-5p	-0.59116667	8.465967742	-3.79067946	8.52E-04	0.041898302	-0.62533283	downregulated
miR-361-5p	-0.80208333	8.309935484	-3.7678447	9.02E-04	0.041898302	-0.67920999	downregulated
miR-151-5p	-0.5815	11.21641935	-3.71438566	1.03E-03	0.043483307	-0.80504125	downregulated
miR-28-5p	-0.51658333	7.756645161	-3.6901619	1.10E-03	0.043483307	-0.8619133	downregulated
miR-365-3p	-0.79541667	10.04370968	-3.67662345	1.14E-03	0.043483307	-0.89365747	downregulated
miR-539-5p	0.26691667	2.47616129	3.57034696	1.49E-03	0.053693933	-1.14176164	upregulated
miR-135a-1-3p	-0.42358333	3.906258065	-3.54624282	1.58E-03	0.054044568	-1.19774912	downregulated
miR-181c-5p	-0.26458333	7.922548387	-3.49928679	1.78E-03	0.057758113	-1.3064929	downregulated
miR-30c-5p	-0.37641667	12.31709677	-3.43987855	2.06E-03	0.061789303	-1.44343334	downregulated
miR-467b-5p	-0.50808333	5.264129032	-3.41138083	2.21E-03	0.061789303	-1.50885615	downregulated
miR-690	-0.92408333	7.523612903	-3.39919249	2.28E-03	0.061789303	-1.53678254	downregulated
miR-760-3p	0.15383333	2.242032258	3.40228342	2.26E-03	0.061789303	-1.52970363	upregulated
miR-320-3p	-0.64425	5.977806452	-3.3608469	2.51E-03	0.06524399	-1.62442208	downregulated
miR-1224-5p	-0.68483333	9.373709677	-3.32368431	2.75E-03	0.067957038	-1.70903082	downregulated
miR-145-5p	-0.735	6.939419355	-3.29856797	2.93E-03	0.067957038	-1.76602522	downregulated
miR-34c-5p	0.881	10.06609677	3.30854941	2.86E-03	0.067957038	-1.74339369	upregulated
miR-674-3p	-0.6175	6.522870968	-3.26771919	3.16E-03	0.070797091	-1.83581349	downregulated
miR-215-5p	0.227	2.624032258	3.19852652	3.74E-03	0.081104588	-1.99145226	upregulated
miR-744-5p	-0.39916667	4.788548387	-3.13956771	4.32E-03	0.087050696	-2.12304726	downregulated
miR-30c-2-3p	-0.51908333	6.837	-3.13684693	4.35E-03	0.087050696	-2.12909634	downregulated
miR-181a-5p	-0.45425	9.889645161	-3.13045104	4.42E-03	0.087050696	-2.14330789	downregulated

miR-706	0.20708333	2.468645161	3.11463289	4.59E-03	0.087800958	-2.17840461	upregulated
miR-22-5p	-0.31933333	7.597677419	-3.07425167	5.06E-03	0.092457266	-2.26766637	downregulated
miR-21-5p	0.80041667	14.39367742	3.06970867	5.12E-03	0.092457266	-2.27767796	upregulated
miR-139-5p	-0.80416667	6.292064516	-3.03004565	5.63E-03	0.097123716	-2.36481558	downregulated
miR-181b-5p	-0.47966667	8.042870968	-3.02172674	5.75E-03	0.097123716	-2.3830296	downregulated
miR-202-3p	-0.716	5.344967742	-2.97790473	6.39E-03	0.097123716	-2.47861143	downregulated
miR-652-3p	-0.333	10.2226129	-2.97230789	6.47E-03	0.097123716	-2.49077401	downregulated
miR-342-5p	-0.40933333	3.336870968	-2.96801547	6.54E-03	0.097123716	-2.50009494	downregulated
miR-92a-3p	-0.589	10.30964516	-2.96763491	6.55E-03	0.097123716	-2.50092102	downregulated
miR-30e-3p	-0.58833333	8.365387097	-2.96575634	6.57E-03	0.097123716	-2.50499816	downregulated
miR-205-5p	2.00116667	4.419225806	3.0038458	6.00E-03	0.097123716	-2.42210505	upregulated
let-7b-5p	-0.3725	13.79919355	-2.94605178	6.89E-03	0.099536054	-2.54769303	downregulated
miR-23b-3p	-0.36575	13.03654839	-2.9310179	7.14E-03	0.099536054	-2.58017995	downregulated
miR-10b-5p	-0.55983333	7.915032258	-2.92451588	7.25E-03	0.099536054	-2.59420647	downregulated
miR-342-3p	-0.701	9.290258065	-2.91897333	7.35E-03	0.099536054	-2.60615177	downregulated

Table S3.4: Functional enrichment analysis of negatively correlated direct targets of disease related miRNAs identified by using g:GOSt module of g:Profiler.

source	term_name	term_id	Adj.p.	-log10 (adj. p)	Term size	Query size	Intersection size	Effective domain size	intersection
GO:MF	gamma-butyrobetaine dioxygenase activity	GO:0008336	0.014961219	1.825033021	1	5	1	25399	BBOX1
GO:MF	WD40-repeat domain binding	GO:0071987	0.037391267	1.427229815	5	5	1	25399	MYB

GO:MF	aryl hydrocarbon receptor binding	GO:0017162	0.042939049	1.367147582	9	5	1	25399	ARNT2
GO:MF	sodium-independent organic anion transmembrane transporter activity	GO:0015347	0.042939049	1.367147582	22	5	1	25399	SLCO2B1
GO:MF	calcium-activated potassium channel activity	GO:0015269	0.042939049	1.367147582	12	5	1	25399	KCNMB2
GO:MF	bile acid transmembrane transporter activity	GO:0015125	0.042939049	1.367147582	23	5	1	25399	SLCO2B1
GO:MF	DNA-binding transcription activator activity, RNA polymerase II-specific	GO:0001228	0.042939049	1.367147582	503	5	2	25399	ARNT2, MYB
GO:MF	DNA-binding transcription activator activity	GO:0001216	0.042939049	1.367147582	510	5	2	25399	ARNT2, MYB
GO:MF	calcium activated cation channel activity	GO:0005227	0.043136292	1.365157194	26	5	1	25399	KCNMB2
REAC	Ca2+ activated K+ channels	REAC:R-MMU-1296052	0.017360506	1.760437613	6	4	1	8748	KCNMB2
REAC	Carnitine synthesis	REAC:R-MMU-71262	0.017360506	1.760437613	2	4	1	8748	BBOX1
REAC	Aryl hydrocarbon receptor signalling	REAC:R-MMU-8937144	0.017360506	1.760437613	5	4	1	8748	ARNT2
REAC	Transport of organic anions	REAC:R-MMU-879518	0.022541576	1.647015716	11	4	1	8748	SLCO2B1

REAC	Heme degradation	REAC:R-MMU-189483	0.022541576	1.647015716	13	4	1	8748	SLCO2B1
REAC	Metabolism of porphyrins	REAC:R-MMU-189445	0.028437779	1.546104326	23	4	1	8748	SLCO2B1
REAC	Endogenous sterols	REAC:R-MMU-211976	0.028437779	1.546104326	23	4	1	8748	ARNT2
REAC	Xenobiotics	REAC:R-MMU-211981	0.037787632	1.422650319	35	4	1	8748	ARNT2
REAC	Transport of vitamins, nucleosides, and related molecules	REAC:R-MMU-425397	0.03835451	1.41618356	40	4	1	8748	SLCO2B1
REAC	Metabolism	REAC:R-MMU-1430728	0.04617583	1.33558529	1682	4	3	8748	ARNT2, SLCO2B1, BBOX1
WP	Wnt signaling pathway (NetPath)	WP:WP539	0.035960044	1.444179781	108	1	1	4505	MYB
WP	Neural crest differentiation	WP:WP2074	0.035960044	1.444179781	99	1	1	4505	MYB

Table S4.1: Hits identified during first round of CSE exposure (Time point T1) in A549 KO cells using MAGeCK.

id	num	pos|score	pos|p-value	pos|fdr	pos|rank	pos|goodsgrna	pos|lfc
CDK1	4	3.26E-07	2.74E-07	0.002475	1	4	4.3714
RPL34	1	0.00077304	2.74E-07	0.002475	75	1	5.2999
OR6K6	4	6.34E-06	3.57E-06	0.015677	2	4	3.7266
UPF2	4	1.17E-05	5.21E-06	0.015677	3	4	3.0192

CHEK1	4	1.42E-05	5.21E-06	0.015677	4	3	2.6409
NOL6	4	1.44E-05	5.21E-06	0.015677	5	4	1.8715
EIF2S1	4	1.68E-05	7.95E-06	0.020509	6	3	3.5185
EIF3G	4	2.91E-05	1.56E-05	0.029208	8	3	3.1296
PRPF8	4	4.14E-05	1.62E-05	0.029208	9	3	3.3416
DDX39B	4	4.46E-05	2.50E-05	0.040954	10	2	2.2262
UBE2I	3	0.0001049	3.81E-05	0.057343	19	3	2.2657
SDAD1	4	5.20E-05	5.07E-05	0.070449	11	3	2.6809
ANK2	4	6.99E-05	6.44E-05	0.083098	12	4	1.1562
DHPS	4	8.03E-05	8.97E-05	0.087252	13	2	2.0546
POLE2	4	8.53E-05	9.02E-05	0.087252	15	3	2.9277
PRMT5	4	9.82E-05	0.00011545	0.087252	16	3	2.059
CCT3	4	9.91E-05	0.00011545	0.087252	17	3	2.9265
SNX31	4	9.97E-05	0.00011545	0.087252	18	4	1.1107
RTFDC1	4	0.00010523	0.000116	0.087252	20	4	1.6997
CKAP5	4	0.00010695	0.000116	0.087252	21	2	1.9559
KHSRP	4	0.00010803	0.000116	0.087252	22	3	2.5622
MMP15	4	0.00011238	0.000116	0.087252	23	3	2.5812
ALG2	4	0.00012033	0.00014891	0.09676	24	4	1.7896
NMI	4	0.00012172	0.00014891	0.09676	25	2	2.1482
NOL9	4	0.0001256	0.00014891	0.09676	26	4	1.8871
RPS21	4	0.00014648	0.00017633	0.09676	27	3	2.282
BPIFB4	4	0.00015168	0.00017633	0.09676	28	3	1.7466
PSMB7	4	0.0001582	0.00017633	0.09676	29	3	2.4915
CCDC84	4	0.00017403	0.00017688	0.09676	30	4	1.8262
RBP5	4	0.00017642	0.00017688	0.09676	31	3	1.8742
MRPL11	4	0.00017741	0.00017688	0.09676	32	4	1.6138

Table S4.2: Hits identified during second round of CSE exposure (Time point T2) in A549 KO cells using MAGeCK.

id	num	pos|score	pos|p-value	pos|fdr	pos|rank	pos|goodsgrna	pos|lfc
COX20	1	0.0054702	2.74E-07	0.00495	401	1	2.4154
YEATS4	4	6.53E-06	6.31E-06	0.023927	1	4	1.9814
MMS22L	4	8.37E-06	6.86E-06	0.023927	2	3	2.0618
ZBTB17	4	8.93E-06	6.86E-06	0.023927	3	3	2.1503
HEATR1	4	1.44E-05	7.95E-06	0.023927	4	3	1.757
RRM1	4	1.44E-05	7.95E-06	0.023927	5	3	2.402
ATP5I	4	2.83E-05	3.70E-05	0.095474	6	3	0.36217

Table S4.3: Hits identified during third round of CSE exposure (Time point T3) in A549 KO cells using MAGeCK.

id	num	pos|score	pos|p-value	pos|fdr	pos|rank	pos|goodsgrna	pos|lfc
CELA3B	1	0.013512	2.74E-07	0.00045	876	1	3.0008
NF2	4	3.29E-08	2.74E-07	0.00045	1	4	3.4537
DDX49	4	3.89E-06	2.74E-07	0.00045	4	4	4.0997
OST4	1	0.018917	2.74E-07	0.00045	1145	1	4.2324
MSTO1	4	4.60E-07	2.74E-07	0.00045	2	4	4.4049
KRAS	4	7.29E-06	2.74E-07	0.00045	6	4	4.5944
PTGES3L-AARSD1	1	0.01491	2.74E-07	0.00045	961	1	4.7539
CNTNAP3	1	0.011881	2.74E-07	0.00045	788	1	5.0122
RPL39	1	0.0075329	2.74E-07	0.00045	544	1	5.4796
ELL	4	5.33E-06	2.74E-07	0.00045	5	3	5.6293
ADSL	3	1.33E-06	2.74E-07	0.00045	3	3	5.7294
PDCD2	4	1.78E-05	1.37E-06	0.002063	7	4	4.6538

VTN	4	2.84E-05	1.92E-06	0.002666	8	3	0.50436
CRTC3	4	3.87E-05	4.11E-06	0.00495	10	4	2.544
TNPO1	4	3.84E-05	4.11E-06	0.00495	9	4	4.3379
PSMB5	4	6.28E-05	1.01E-05	0.010176	13	2	3.0845
RPL11	4	5.45E-05	9.60E-06	0.010176	12	3	3.9794
RPS3	4	5.45E-05	9.60E-06	0.010176	11	4	4.2876
SLURP1	4	8.51E-05	2.28E-05	0.018295	16	4	2.8237
TUBB	4	8.50E-05	2.28E-05	0.018295	15	4	3.8021
FAM103A1	4	7.77E-05	2.28E-05	0.018295	14	3	3.8813
MCM5	4	9.79E-05	2.33E-05	0.018295	17	4	4.068
SF3A3	3	0.00018766	2.22E-05	0.018295	30	3	5.1039
HIST1H2AK	4	0.00014185	3.26E-05	0.01841	22	2	0.91923
RALGAPB	4	0.00014259	3.26E-05	0.01841	23	4	1.8735
HSPA14	4	0.00011849	3.21E-05	0.01841	18	4	2.2011
UBE4B	4	0.00013842	3.26E-05	0.01841	21	4	2.7078
CDK6	4	0.00014828	3.26E-05	0.01841	25	3	3.4771
NAA15	4	0.00016011	3.26E-05	0.01841	26	4	3.6702
FASN	4	0.00012777	3.21E-05	0.01841	19	4	3.7008
CXXC1	4	0.00013336	3.26E-05	0.01841	20	3	4.0883
CTCF	4	0.00014652	3.26E-05	0.01841	24	4	4.2539
CDK7	4	0.00017311	4.25E-05	0.022568	27	3	2.9166
EZH1	4	0.00017462	4.25E-05	0.022568	28	3	2.918
LCLAT1	4	0.00031206	7.98E-05	0.022867	45	2	0.26332
OR52N5	4	0.00036879	7.98E-05	0.022867	48	4	0.52311
ZNF787	4	0.00025533	7.93E-05	0.022867	40	3	0.88967
MBNL1	4	0.00029785	7.93E-05	0.022867	44	4	1.1243
NFIC	4	0.00034839	7.98E-05	0.022867	46	4	1.2544

LRRC47	4	0.00038296	7.98E-05	0.022867	55	4	1.4646
GPR26	4	0.00024649	7.93E-05	0.022867	38	4	1.5296
TP53	4	0.00024129	7.93E-05	0.022867	37	4	1.6496
RPS25	4	0.00037334	7.98E-05	0.022867	49	2	2.0146
AHR	4	0.00036769	7.98E-05	0.022867	47	4	2.1512
PAXIP1	4	0.0002092	7.93E-05	0.022867	34	4	2.339
NIP7	4	0.00021137	7.93E-05	0.022867	35	4	2.3473
OTOP1	4	0.00037334	7.98E-05	0.022867	53	2	2.4869
RTTN	4	0.00037334	7.98E-05	0.022867	50	2	2.5883
TOMM22	4	0.00023183	7.93E-05	0.022867	36	2	2.5925
TLN1	4	0.00037334	7.98E-05	0.022867	54	2	2.6018
TBL3	4	0.00037334	7.98E-05	0.022867	51	2	2.6261
AP2M1	4	0.00018844	5.73E-05	0.022867	31	4	2.7864
VHL	4	0.0004013	7.98E-05	0.022867	56	4	2.7904
ADM5	4	0.00028028	7.93E-05	0.022867	43	3	2.8438
LSM2	4	0.00028028	7.93E-05	0.022867	42	3	2.8781
PPAT	4	0.00040548	7.98E-05	0.022867	57	3	3.1098
PSMA2	4	0.00037334	7.98E-05	0.022867	52	4	3.2437
HNRNPK	4	0.00026416	7.93E-05	0.022867	41	4	3.3733
ZNF239	4	0.00019413	5.79E-05	0.022867	32	3	3.6897
DDX10	4	0.00018123	5.73E-05	0.022867	29	3	3.7916
WRB	4	0.0002511	7.93E-05	0.022867	39	4	3.8374
RPL30	2	0.00099362	5.29E-05	0.022867	110	2	4.0308
DNAJA4	4	0.00042552	0.00011381	0.031607	59	2	0.50036
PTF1A	4	0.00041001	0.00011326	0.031607	58	3	2.6357
PCGF1	3	0.00048934	0.00012807	0.035029	66	2	0.46397
PEX11G	4	0.00045108	0.00017524	0.044073	64	4	1.1838

DPF2	4	0.00048224	0.00017578	0.044073	65	3	1.2545
NUP85	4	0.00044805	0.00017524	0.044073	61	3	3.4339
CDIPT	4	0.00044805	0.00017524	0.044073	60	4	3.5636
RAD51	4	0.00044805	0.00017524	0.044073	63	3	3.9134
KAT8	4	0.00044805	0.00017524	0.044073	62	3	3.9803
LAMB4	4	0.00055685	0.00021582	0.045418	72	4	1.4014
TAOK1	4	0.00051598	0.00021582	0.045418	67	4	1.5358
DNAI2	4	0.00063729	0.00021637	0.045418	78	4	1.9194
MAP4K4	4	0.00066044	0.00021637	0.045418	80	4	1.9665
IRF2BP2	4	0.00056394	0.00021582	0.045418	73	3	2.1049
C12orf75	4	0.0005318	0.00021582	0.045418	70	4	2.1976
ATP2A2	4	0.00061237	0.00021637	0.045418	76	2	2.3307
PGLYRP1	4	0.00065814	0.00021637	0.045418	79	2	2.435
USP10	4	0.00052887	0.00021582	0.045418	69	2	2.7524
LUC7L3	4	0.00055568	0.00021582	0.045418	71	3	2.9267
MDN1	4	0.00057341	0.00021582	0.045418	74	3	3.1629
SLITRK5	4	0.00057643	0.00021582	0.045418	75	4	3.4231
ZWINT	4	0.00052887	0.00021582	0.045418	68	4	3.6495
PRMT5	4	0.00061237	0.00021637	0.045418	77	4	4.0781
TVP23A	4	0.00076584	0.00026299	0.045744	88	2	0.015329
MIIP	4	0.00082255	0.00026354	0.045744	95	2	0.077921
ANKS3	4	0.00071964	0.00026299	0.045744	86	4	1.1881
CRYBA2	4	0.00070912	0.00026299	0.045744	84	3	1.2699
MEF2B	4	0.00075441	0.00026299	0.045744	87	4	1.3288
DPH6	4	0.00079836	0.00026354	0.045744	93	3	1.9076
CCBL1	4	0.00078937	0.00026354	0.045744	92	3	1.9369
MYOM2	4	0.00068567	0.00026299	0.045744	82	2	2.1811

FAM117B	4	0.00071251	0.00026299	0.045744	85	3	2.3089
SYMPK	4	0.00077538	0.00026354	0.045744	89	2	2.4406
GSDMC	4	0.00067045	0.00026299	0.045744	81	4	2.7033
FOSL1	4	0.00068567	0.00026299	0.045744	83	3	2.7901
AIFM1	4	0.00085245	0.00026354	0.045744	97	3	2.8003
DOLK	4	0.0008695	0.00026354	0.045744	98	3	2.879
ZZZ3	4	0.00077538	0.00026354	0.045744	90	3	2.9031
XRCC3	4	0.00081539	0.00026354	0.045744	94	3	3.1342
ATP5O	4	0.00077538	0.00026354	0.045744	91	3	3.197
CDK5R1	4	0.00084576	0.00026354	0.045744	96	3	3.4692
TMEM165	4	0.00087925	0.00031016	0.051458	99	1	0.068578
OR6B1	4	0.00088978	0.00031016	0.051458	100	2	1.6179
DAPK1	4	0.00090963	0.00031071	0.051458	103	3	1.9162
TRNT1	4	0.00089107	0.00031016	0.051458	101	4	2.087
HSCB	4	0.00090259	0.00031016	0.051458	102	3	2.9056
CCRL2	4	0.00099266	0.00037269	0.057998	109	4	0.82815
ASB17	4	0.00093596	0.00037214	0.057998	104	4	1.4645
NUP50	4	0.0009828	0.00037269	0.057998	108	2	1.9043
EXT1	4	0.0009753	0.00037269	0.057998	106	3	2.1254
PGS1	4	0.00096792	0.00037214	0.057998	105	4	2.4599
KIF14	4	0.0009828	0.00037269	0.057998	107	3	2.7775
UROD	4	0.00099387	0.00037269	0.057998	111	4	2.814
KRTAP10-8	2	0.0025519	0.00045605	0.070365	226	1	3.2709
KCNC1	4	0.0011061	0.00052132	0.071938	117	2	0.25883
ST14	4	0.0013895	0.00059591	0.071938	133	2	0.57469
PPP4R4	4	0.0012195	0.00059536	0.071938	123	4	0.59772
TMEM71	4	0.0016162	0.00069683	0.071938	156	2	0.62891

MRC2	4	0.0016162	0.00069683	0.071938	154	4	0.64294
PARP10	4	0.0012195	0.00059536	0.071938	122	4	0.82732
GALR2	4	0.0017863	0.00069738	0.071938	169	3	0.83606
SIGLEC14	4	0.0015789	0.00069683	0.071938	152	4	1.0845
SARDH	4	0.001516	0.00069683	0.071938	146	4	1.0956
KIRREL	4	0.0014776	0.00059591	0.071938	143	4	1.1874
PAPPA2	4	0.0012195	0.00059536	0.071938	124	3	1.2101
ATF4	4	0.0010901	0.00052132	0.071938	114	4	1.2378
LRIG3	4	0.001272	0.00059536	0.071938	129	4	1.2857
CDHR1	4	0.0014124	0.00059591	0.071938	137	4	1.3109
KDM8	4	0.0015659	0.00069683	0.071938	150	3	1.4476
KIF1C	4	0.0013064	0.00059536	0.071938	130	4	1.4572
ATP6V0E1	4	0.0010494	0.00052077	0.071938	113	2	1.5994
COPS2	4	0.0013895	0.00059591	0.071938	135	4	1.6433
ARNT	4	0.0017264	0.00069738	0.071938	163	3	1.6477
ALPK1	4	0.001774	0.00069738	0.071938	166	3	1.7046
POLR2A	4	0.0014378	0.00059591	0.071938	140	4	1.7138
SHOC2	4	0.0013094	0.00059591	0.071938	131	3	1.8203
SLC17A8	4	0.0015637	0.00069683	0.071938	149	2	1.8269
BPI	4	0.0015363	0.00069683	0.071938	148	3	1.8877
RTKN	4	0.0012639	0.00059536	0.071938	128	3	1.9564
ILF3	4	0.0014321	0.00059591	0.071938	139	2	1.9819
TMEM200B	4	0.0016469	0.00069683	0.071938	159	4	1.9999
SGSM3	4	0.0016415	0.00069683	0.071938	158	3	2.0177
HSF1	4	0.0012393	0.00059536	0.071938	125	4	2.0732
RPS2	4	0.0014671	0.00059591	0.071938	141	2	2.1023
TGFBRAP1	4	0.0014314	0.00059591	0.071938	138	3	2.1552

CCDC96	4	0.0015694	0.00069683	0.071938	151	3	2.1894
CD8A	4	0.0011168	0.00052132	0.071938	119	3	2.1932
ELP5	4	0.0016162	0.00069683	0.071938	157	3	2.1948
NSUN3	4	0.0017816	0.00069738	0.071938	167	4	2.2221
PIGA	4	0.0017826	0.00069738	0.071938	168	3	2.2424
ADC	4	0.0015105	0.00069683	0.071938	145	3	2.2687
C1orf27	4	0.0013459	0.00059591	0.071938	132	2	2.2897
FTSJ3	4	0.0011097	0.00052132	0.071938	118	2	2.2999
ARMC7	4	0.0015207	0.00069683	0.071938	147	4	2.4791
ANXA11	4	0.0017381	0.00069738	0.071938	164	3	2.6434
SF3B2	4	0.0017447	0.00069738	0.071938	165	3	2.7366
RNF40	4	0.001042	0.00052077	0.071938	112	3	2.7574
RNGTT	4	0.0011008	0.00052132	0.071938	116	3	2.791
SDHA	4	0.001249	0.00059536	0.071938	126	3	2.8141
KRT77	4	0.0011556	0.00059536	0.071938	120	3	2.9363
ABHD11	4	0.0017235	0.00069738	0.071938	162	4	2.9636
PSMC3	4	0.0016578	0.00069683	0.071938	160	4	2.9743
RRN3	4	0.001249	0.00059536	0.071938	127	3	3.1965
FBL	4	0.0014092	0.00059591	0.071938	136	3	3.372
BRF2	4	0.0011649	0.00059536	0.071938	121	3	3.3876
RCL1	4	0.0016578	0.00069683	0.071938	161	4	3.4042
SF3B3	4	0.0014671	0.00059591	0.071938	142	4	3.453
PELP1	4	0.001095	0.00052132	0.071938	115	3	3.4594
SSRP1	4	0.001611	0.00069683	0.071938	153	3	3.5053
SAP30BP	3	0.0014893	0.00062389	0.071938	144	2	4.1828
HSF2	4	0.0018996	0.00086137	0.078188	175	2	0.29697
OR4K2	4	0.0018996	0.00086137	0.078188	174	2	0.31194

CSRNP1	4	0.0020979	0.00086192	0.078188	187	4	0.78405
OVCA2	4	0.0020129	0.00086137	0.078188	183	3	0.86614
ALG3	4	0.0020979	0.00086192	0.078188	186	3	1.0339
DFNB59	4	0.0018996	0.00086137	0.078188	173	2	1.1258
DLGAP4	4	0.0021793	0.00086192	0.078188	192	4	1.1442
ITPR1	4	0.0020583	0.00086192	0.078188	184	4	1.2872
CACNG6	4	0.002139	0.00086192	0.078188	190	4	1.398
PAX7	4	0.0019095	0.00086137	0.078188	177	4	1.7138
ZKSCAN5	4	0.0019772	0.00086137	0.078188	180	2	1.8445
RETSAT	4	0.0019963	0.00086137	0.078188	181	3	1.9351
CMIP	4	0.0021204	0.00086192	0.078188	188	4	1.9559
DNAJC9	4	0.0020078	0.00086137	0.078188	182	4	2.0498
RNASE2	4	0.0019472	0.00086137	0.078188	179	3	2.1101
FICD	4	0.0019055	0.00086137	0.078188	176	4	2.1931
TMOD2	4	0.0021496	0.00086192	0.078188	191	3	2.4731
KHDRBS1	4	0.0018497	0.00086137	0.078188	170	3	2.6062
CARS	4	0.0022033	0.00086192	0.078188	193	3	2.7637
C11orf48	4	0.0020641	0.00086192	0.078188	185	3	2.8385
UPF3A	4	0.0019419	0.00086137	0.078188	178	3	2.8572
SFRP5	4	0.0018937	0.00086137	0.078188	172	3	2.9751
ATP5C1	4	0.0021336	0.00086192	0.078188	189	4	2.9768
SCAP	4	0.0018497	0.00086137	0.078188	171	4	3.199
CACNB3	4	0.0022395	0.00093871	0.08074	196	2	0.00060472
ING5	3	0.0027425	0.00092554	0.08074	237	2	1.2097
ZNF407	4	0.0022802	0.00093926	0.08074	199	2	1.4577
THOC3	4	0.0022802	0.00093926	0.08074	198	2	1.8148
GPN3	4	0.0022802	0.00093926	0.08074	200	2	2.1042

LILRB5	4	0.0022465	0.00093871	0.08074	197	4	2.1391
LELP1	4	0.0022339	0.00093871	0.08074	194	3	2.4883
IMP4	3	0.0024664	0.00092554	0.08074	211	3	2.5703
CSNK1A1	4	0.0022802	0.00093926	0.08074	201	3	2.5897
EIF2S3	3	0.0025301	0.00092554	0.08074	217	3	2.9219
SERGEF	4	0.0023528	0.0012332	0.099603	204	2	0.33433
SNF8	4	0.0025511	0.0013347	0.099603	225	3	0.81596
ARHGEF38	4	0.0024305	0.0013347	0.099603	207	4	0.81911
MAP1A	4	0.0026659	0.0013353	0.099603	234	4	1.0771
STAT4	4	0.0025269	0.0013347	0.099603	215	4	1.1295
MRM1	4	0.0025511	0.0013347	0.099603	221	2	1.1538
ZNF569	4	0.0025511	0.0013347	0.099603	220	4	1.2556
SMARCE1	4	0.0026729	0.0013353	0.099603	235	4	1.2954
OR8B4	4	0.0023336	0.0012332	0.099603	203	3	1.4681
C3orf18	4	0.0025746	0.0013353	0.099603	227	4	1.5413
PCOLCE	4	0.002646	0.0013353	0.099603	233	4	1.5419
TSC22D1	4	0.0025511	0.0013347	0.099603	222	4	1.5565
PSMD2	4	0.0025378	0.0013347	0.099603	219	4	1.5685
RPL7A	4	0.0025511	0.0013347	0.099603	223	3	1.5745
MTCH1	4	0.0024535	0.0013347	0.099603	209	4	1.5793
FAM122A	4	0.0024054	0.0012338	0.099603	206	4	1.5983
AQP3	4	0.0025033	0.0013347	0.099603	213	3	1.6635
GABARAP	4	0.0025288	0.0013347	0.099603	216	2	1.7245
OR1F1	4	0.0025307	0.0013347	0.099603	218	3	1.8158
FAM92A1	4	0.0024345	0.0013347	0.099603	208	3	1.9
G6PD	4	0.0026456	0.0013353	0.099603	232	3	1.9188
PARN	4	0.00258	0.0013353	0.099603	228	3	2.039

DHPS	4	0.0025254	0.0013347	0.099603	214	4	2.2799
C16orf80	4	0.0026363	0.0013353	0.099603	231	3	2.2938
OR5A2	4	0.0023233	0.0012332	0.099603	202	4	2.3512
DHX33	4	0.0025017	0.0013347	0.099603	212	4	2.452
PCNA	4	0.0026345	0.0013353	0.099603	230	3	2.4742
PPP3R1	2	0.0041239	0.0012579	0.099603	330	1	2.8632
RAD51D	4	0.0024663	0.0013347	0.099603	210	4	2.998
SLC2A6	4	0.0023949	0.0012338	0.099603	205	3	3.0324
TUT1	4	0.002602	0.0013353	0.099603	229	3	3.42

Table S4.4: Functional enrichment analysis of hits identified during first round of CSE exposure (Time point T1) in A549 KO cells using enrichr GO:BP and REACTOME database.

Term	GeneOntology	Intersection	TotalSize	P-value	Adjusted P-value	Genes
cellular macromolecule biosynthetic process	GO:0034645	6	314	6.17E-06	0.002554405	POLE2, CHEK1, DHPS, CDK1, RPS21, MRPL11
gene expression	GO:0010467	6	356	1.26E-05	0.002611868	UPF2, DDX39B, DHPS, NOL9, RPS21, MRPL11
RNA processing	GO:0006396	4	179	0.000141743	0.017379634	DDX39B, KHSRP, NOL9, PRPF8
endonucleolytic cleavage of tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA	GO:0000479	2	13	0.000167919	0.017379634	NOL9, RPS21
cellular protein metabolic process	GO:0044267	5	417	0.000356562	0.028932097	UBE2I, MMP15, DHPS, RPS21, MRPL11
RNA splicing	GO:0008380	3	98	0.000420738	0.028932097	DDX39B, KHSRP, PRPF8

maturation of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA	GO:0000466	2	22	0.000493146	0.028932097	NOL9, RPS21
DNA replication	GO:0006260	3	108	0.000559074	0.028932097	POLE2, CHEK1, CDK1
RNA splicing, via transesterification reactions	GO:0000375	2	25	0.000638662	0.029378435	KHSRP, PRPF8
mitotic G2 DNA damage checkpoint signaling	GO:0007095	2	33	0.0011157	0.046189974	CHEK1, CDK1

Pathway name	#Entities found	#Entities total	Entities ratio	Entities pValue	Entities FDR	Submitted entities found
Metabolism of RNA	11	829	0.054396325	2.08E-06	7.87E-04	UPF2, PSMB7, PRMT5, DDX39B, KHSRP, NOL9, RPS21, PRPF8, NOL6
Formation of the ternary complex, and subsequently, the 43S complex	4	54	0.003543307	8.48E-06	0.001390998	EIF3G, EIF2S1, RPS21
Translation initiation complex formation	4	62	0.004068241	1.45E-05	0.001390998	EIF3G, EIF2S1, RPS21
Ribosomal scanning and start codon recognition	4	64	0.004199475	1.65E-05	0.001390998	EIF3G, EIF2S1, RPS21
Activation of the mRNA upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S	4	66	0.004330709	1.85E-05	0.001390998	EIF3G, EIF2S1, RPS21
G1/S Transition	5	150	0.00984252	2.70E-05	0.001701076	PSMB7, POLE2, CDK1
Mitotic G1 phase and G1/S transition	5	174	0.011417323	5.44E-05	0.002940237	PSMB7, POLE2, CDK1
Regulation of expression of SLITs and ROBOs	5	183	0.012007874	6.90E-05	0.003243124	UPF2, PSMB7, RPS21

SARS-CoV-1-host interactions	4	118	0.007742782	1.74E-04	0.005274793	UBE2I, NMI, RPS21
GTP hydrolysis and joining of the 60S ribosomal subunit	4	120	0.007874016	1.85E-04	0.005274793	EIF3G, EIF2S1, RPS21
L13a-mediated translational silencing of Ceruloplasmin expression	4	120	0.007874016	1.85E-04	0.005274793	EIF3G, EIF2S1, RPS21
Signaling by ROBO receptors	5	235	0.015419948	2.21E-04	0.005274793	UPF2, PSMB7, RPS21
Cap-dependent Translation Initiation	4	130	0.008530184	2.51E-04	0.005274793	EIF3G, EIF2S1, RPS21
Eukaryotic Translation Initiation	4	130	0.008530184	2.51E-04	0.005274793	EIF3G, EIF2S1, RPS21
Cell Cycle	8	734	0.04816273	2.60E-04	0.005274793	PSMB7, UBE2I, POLE2, CHEK1, CDK1, CKAP5
Mitotic Anaphase	5	249	0.016338583	2.88E-04	0.005274793	PSMB7, UBE2I, CDK1, CKAP5
p53-Independent G1/S DNA damage checkpoint	3	54	0.003543307	2.91E-04	0.005274793	PSMB7, CHEK1
p53-Independent DNA Damage Response	3	54	0.003543307	2.91E-04	0.005274793	PSMB7, CHEK1
Ubiquitin Mediated Degradation of Phosphorylated Cdc25A	3	54	0.003543307	2.91E-04	0.005274793	PSMB7, CHEK1
Mitotic Metaphase and Anaphase	5	250	0.016404199	2.93E-04	0.005274793	PSMB7, UBE2I, CDK1, CKAP5
Cell Cycle, Mitotic	7	596	0.039107612	4.29E-04	0.007724279	PSMB7, UBE2I, POLE2, CDK1, CKAP5
G2/M Checkpoints	4	153	0.01003937	4.63E-04	0.007726806	PSMB7, CHEK1, CDK1
Cell Cycle Checkpoints	5	279	0.018307087	4.83E-04	0.007726806	PSMB7, CHEK1, CDK1, CKAP5

Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex	2	15	9.84E-04	5.97E-04	0.008952934	CHEK1, CDK1
G1/S DNA Damage Checkpoints	3	72	0.004724409	6.70E-04	0.009760242	PSMB7, CHEK1
Cdc20:Phospho-APC/C mediated degradation of Cyclin A	3	73	0.004790026	6.97E-04	0.009760242	PSMB7, CDK1
APC:Cdc20 mediated degradation of cell cycle proteins prior to satisfation of the cell cycle checkpoint	3	74	0.004855643	7.25E-04	0.010150441	PSMB7, CDK1
APC/C:Cdc20 mediated degradation of mitotic proteins	3	76	0.004986877	7.83E-04	0.010177506	PSMB7, CDK1
Activation of APC/C and APC/C:Cdc20 mediated degradation of mitotic proteins	3	77	0.005052493	8.13E-04	0.01056746	PSMB7, CDK1
Regulation of APC/C activators between G1/S and early anaphase	3	83	0.005446194	0.001008271	0.011161766	PSMB7, CDK1
The role of GTSE1 in G2/M progression after G2 checkpoint	3	83	0.005446194	0.001008271	0.011161766	PSMB7, CDK1
Major pathway of rRNA processing in the nucleolus and cytosol	4	189	0.012401575	0.001014873	0.011161766	NOL9, RPS21, NOL6
Transcription of E2F targets under negative control by p107 (RBL1) and p130	2	20	0.001312336	0.001053246	0.011161766	CDK1

(RBL2) in complex with HDAC1						
Metabolism of proteins	13	2215	0.145341207	0.001102194	0.011161766	CCT3, PSMB7, UBE2I, DHPS, ALG2, EIF3G, CDK1, ANK2, EIF2S1, RPS21, MRPL11
TP53 Regulates Transcription of DNA Repair Genes	3	86	0.005643045	0.001116177	0.011161766	CHEK1
SARS-CoV-1 Infection	4	194	0.012729659	0.001117216	0.011172161	UBE2I, NMI, RPS21
Translation	5	339	0.022244094	0.001155142	0.011551424	EIF3G, EIF2S1, RPS21, MRPL11
Regulation of mitotic cell cycle	3	92	0.006036745	0.001353291	0.012179621	PSMB7, CDK1
APC/C-mediated degradation of cell cycle proteins	3	92	0.006036745	0.001353291	0.012179621	PSMB7, CDK1
Regulation of mRNA stability by proteins that bind AU-rich elements	3	93	0.006102362	0.001395646	0.012560811	PSMB7, KHSRP
rRNA processing in the nucleus and cytosol	4	208	0.013648294	0.001442149	0.012798016	NOL9, RPS21, NOL6
G2/M Transition	4	212	0.013910761	0.001545945	0.012798016	PSMB7, CDK1, CKAP5
Mitotic G2-G2/M phases	4	214	0.014041995	0.001599752	0.012798016	PSMB7, CDK1, CKAP5
Formation of a pool of free 40S subunits	3	106	0.006955381	0.002023619	0.016188951	EIF3G, RPS21
MAPK6/MAPK4 signaling	3	106	0.006955381	0.002023619	0.016188951	PSMB7, CDK1
Axon guidance	6	585	0.038385827	0.002309467	0.018475732	UPF2, PSMB7, ANK2, RPS21
Response of EIF2AK4 (GCN2) to amino acid deficiency	3	115	0.007545932	0.002546821	0.018535253	RPS21, EIF2S1
rRNA processing	4	246	0.016141732	0.002647893	0.018535253	NOL9, RPS21, NOL6

M Phase	5	416	0.027296588	0.002820709	0.019744962	PSMB7, UBE2I, CDK1, CKAP5
Nervous system development	6	621	0.040748031	0.003107965	0.021546862	UPF2, PSMB7, ANK2, RPS21
Nonsense-Mediated Decay (NMD)	3	124	0.008136483	0.003146788	0.021546862	UPF2, RPS21
Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC)	3	124	0.008136483	0.003146788	0.021546862	UPF2, RPS21
HDR through Homologous Recombination (HRR) or Single Strand Annealing (SSA)	3	124	0.008136483	0.003146788	0.021546862	UBE2I, POLE2, CHEK1
ABC-family proteins mediated transport	3	124	0.008136483	0.003146788	0.021546862	PSMB7, EIF2S1
Homology Directed Repair	3	130	0.008530184	0.003591144	0.021546862	UBE2I, POLE2, CHEK1
G0 and Early G1	2	38	0.002493438	0.003702173	0.022213037	CDK1
Synthesis of DNA	3	132	0.008661417	0.003747356	0.022484137	PSMB7, POLE2
DNA Replication Pre-Initiation	3	132	0.008661417	0.003747356	0.022484137	PSMB7, POLE2
SARS-CoV-1 modulates host translation machinery	2	41	0.002690289	0.004290718	0.025744311	RPS21
PERK regulates gene expression	2	42	0.002755906	0.004495921	0.026791001	KHSRP, EIF2S1
G1/S-Specific Transcription	2	43	0.002821522	0.004705599	0.026791001	CDK1
Transcriptional regulation by RUNX2	3	147	0.009645669	0.005052013	0.026791001	PSMB7, CDK1
Transcriptional Regulation by TP53	5	484	0.03175853	0.0053582	0.026791001	PRMT5, CHEK1, CDK1

Regulation of activated PAK-2p34 by proteasome mediated degradation	2	50	0.00328084	0.006296883	0.028536425	PSMB7
Regulation of ornithine decarboxylase (ODC)	2	51	0.003346457	0.006541609	0.028536425	PSMB7
Vpu mediated degradation of CD4	2	53	0.00347769	0.007043905	0.028536425	PSMB7
Cross-presentation of soluble exogenous antigens (endosomes)	2	53	0.00347769	0.007043905	0.028536425	PSMB7
SUMOylation of nuclear envelope proteins	1	3	1.97E-04	0.007062256	0.028536425	UBE2I
Defective ALG2 causes CDG-1i	1	3	1.97E-04	0.007062256	0.028536425	ALG2
GSK3B and BTRC:CUL1-mediated-degradation of NFE2L2	2	54	0.003543307	0.007301433	0.028536425	PSMB7
Ubiquitin-dependent degradation of Cyclin D	2	54	0.003543307	0.007301433	0.028536425	PSMB7
Regulation of Apoptosis	2	54	0.003543307	0.007301433	0.028536425	PSMB7
Autodegradation of the E3 ubiquitin ligase COP1	2	54	0.003543307	0.007301433	0.028536425	PSMB7
DNA Replication	3	169	0.011089239	0.007409773	0.028536425	PSMB7, POLE2
DNA Double-Strand Break Repair	3	169	0.011089239	0.007409773	0.028536425	UBE2I, POLE2, CHEK1
FBXL7 down-regulates AURKA during mitotic entry and in early mitosis	2	55	0.003608924	0.007563189	0.028536425	PSMB7
SCF-beta-TrCP mediated degradation of Emi1	2	55	0.003608924	0.007563189	0.028536425	PSMB7
AUF1 (hnRNP D0) binds and destabilizes mRNA	2	56	0.003674541	0.007829151	0.028536425	PSMB7

Vif-mediated degradation of APOBEC3G	2	56	0.003674541	0.007829151	0.028536425	PSMB7
Degradation of DVL	2	57	0.003740157	0.0080993	0.028536425	PSMB7
Degradation of AXIN	2	57	0.003740157	0.0080993	0.028536425	PSMB7
Negative regulation of NOTCH4 signaling	2	57	0.003740157	0.0080993	0.028536425	PSMB7
Regulation of RUNX3 expression and activity	2	57	0.003740157	0.0080993	0.028536425	PSMB7
Cellular response to starvation	3	176	0.011548556	0.008275568	0.028536425	RPS21, EIF2S1
Regulation of TP53 Activity	3	178	0.01167979	0.008533448	0.028536425	PRMT5, CHEK1, CDK1
Stabilization of p53	2	59	0.003871391	0.008652077	0.028536425	PSMB7
S Phase	3	179	0.011745407	0.008664152	0.028536425	PSMB7, POLE2
Hh mutants are degraded by ERAD	2	61	0.004002625	0.00922136	0.028536425	PSMB7
NIK-->noncanonical NF-kB signaling	2	61	0.004002625	0.00922136	0.028536425	PSMB7
Degradation of GLI1 by the proteasome	2	62	0.004068241	0.009512142	0.028536425	PSMB7
Degradation of GLI2 by the proteasome	2	62	0.004068241	0.009512142	0.028536425	PSMB7
GLI3 is processed to GLI3R by the proteasome	2	62	0.004068241	0.009512142	0.028536425	PSMB7
SARS-CoV-2 modulates host translation machinery	2	62	0.004068241	0.009512142	0.028536425	RPS21
SCF(Skp2)-mediated degradation of p27/p21	2	62	0.004068241	0.009512142	0.028536425	PSMB7
Autodegradation of Cdh1 by Cdh1:APC/C	2	64	0.004199475	0.010105886	0.030317657	PSMB7
Hh mutants abrogate ligand secretion	2	64	0.004199475	0.010105886	0.030317657	PSMB7

Asymmetric localization of PCP proteins	2	66	0.004330709	0.01071574	0.032147221	PSMB7
Defective CFTR causes cystic fibrosis	2	66	0.004330709	0.01071574	0.032147221	PSMB7
Dectin-1 mediated noncanonical NF-kB signaling	2	66	0.004330709	0.01071574	0.032147221	PSMB7
Separation of Sister Chromatids	3	195	0.012795276	0.010917323	0.03275197	PSMB7, CKAP5
APC/C:Cdc20 mediated degradation of Securin	2	68	0.004461942	0.011341549	0.033056308	PSMB7
Phosphorylation of proteins involved in the G2/M transition by Cyclin A:Cdc2 complexes	1	5	3.28E-04	0.011743462	0.033056308	CDK1
p53-Dependent G1 DNA Damage Response	2	70	0.004593176	0.011983157	0.033056308	PSMB7
p53-Dependent G1/S DNA damage checkpoint	2	70	0.004593176	0.011983157	0.033056308	PSMB7
Loss of proteins required for interphase microtubule organization from the centrosome	2	71	0.004658793	0.012309837	0.033056308	CDK1, CKAP5
Loss of Nlp from mitotic centrosomes	2	71	0.004658793	0.012309837	0.033056308	CDK1, CKAP5
Regulation of RAS by GAPs	2	71	0.004658793	0.012309837	0.033056308	PSMB7
Activation of NF-kappaB in B cells	2	72	0.004724409	0.012640409	0.033056308	PSMB7
Oxygen-dependent proline hydroxylation of Hypoxia-inducible Factor Alpha	2	72	0.004724409	0.012640409	0.033056308	PSMB7
Hedgehog ligand biogenesis	2	72	0.004724409	0.012640409	0.033056308	PSMB7

Orc1 removal from chromatin	2	73	0.004790026	0.012974854	0.033056308	PSMB7
HDR through Homologous Recombination (HRR)	2	73	0.004790026	0.012974854	0.033056308	POLE2, CHEK1
Cilium Assembly	3	210	0.013779528	0.013311406	0.033056308	CCT3, CDK1, CKAP5
AURKA Activation by TPX2	2	74	0.004855643	0.013313153	0.033056308	CDK1, CKAP5
APC/C:Cdh1 mediated degradation of Cdc20 and other APC/C:Cdh1 targeted proteins in late mitosis/early G1	2	74	0.004855643	0.013313153	0.033056308	PSMB7
Regulation of PTEN stability and activity	2	74	0.004855643	0.013313153	0.033056308	PSMB7
CDK-mediated phosphorylation and removal of Cdc6	2	75	0.00492126	0.013655287	0.033056308	PSMB7
G2/M DNA damage checkpoint	2	80	0.005249344	0.015422818	0.033056308	CHEK1, CDK1
Metabolism of polyamines	2	80	0.005249344	0.015422818	0.033056308	PSMB7
Recruitment of mitotic centrosome proteins and complexes	2	81	0.005314961	0.015787566	0.033056308	CDK1, CKAP5
SUMO is transferred from E1 to E2 (UBE2I, UBC9)	1	7	4.59E-04	0.01640321	0.033056308	UBE2I
G2/M DNA replication checkpoint	1	7	4.59E-04	0.01640321	0.033056308	CDK1
Centrosome maturation	2	83	0.005446194	0.016528154	0.033056308	CDK1, CKAP5
Regulation of RUNX2 expression and activity	2	83	0.005446194	0.016528154	0.033056308	PSMB7
RNA Polymerase II Transcription	9	1730	0.11351706	0.016929203	0.033858406	PSMB7, PRMT5, UBE2I, DDX39B, CHEK1, CDK1

Cellular response to hypoxia	2	86	0.005643045	0.017666491	0.035332983	PSMB7
Nuclear Envelope (NE) Reassembly	2	88	0.005774278	0.018443473	0.036886946	UBE2I, CDK1
Cyclin E associated events during G1/S transition	2	88	0.005774278	0.018443473	0.036886946	PSMB7
Phosphorylation of Emi1	1	8	5.25E-04	0.018725066	0.037450132	CDK1
Processing of DNA double-strand break ends	2	89	0.005839895	0.018837336	0.037674673	UBE2I, CHEK1
Cyclin A:Cdk2-associated events at S phase entry	2	90	0.005905512	0.019234757	0.038469515	PSMB7
Downstream signaling events of B Cell Receptor (BCR)	2	91	0.005971129	0.019635718	0.039271436	PSMB7
Degradation of beta-catenin by the destruction complex	2	91	0.005971129	0.019635718	0.039271436	PSMB7
Regulation of PLK1 Activity at G2/M Transition	2	92	0.006036745	0.0200402	0.040080401	CDK1, CKAP5
Global Genome Nucleotide Excision Repair (GG-NER)	2	92	0.006036745	0.0200402	0.040080401	UBE2I, POLE2
Hedgehog 'on' state	2	92	0.006036745	0.0200402	0.040080401	PSMB7
Signaling by NOTCH4	2	92	0.006036745	0.0200402	0.040080401	PSMB7
Switching of origins to a post-replicative state	2	94	0.006167979	0.020859658	0.041719317	PSMB7
E2F-enabled inhibition of pre-replication complex formation	1	9	5.91E-04	0.021041593	0.042083186	CDK1
Activation of NIMA Kinases NEK9, NEK6, NEK7	1	9	5.91E-04	0.021041593	0.042083186	CDK1
Recruitment of NuMA to mitotic centrosomes	2	97	0.006364829	0.022114814	0.044229627	CDK1, CKAP5
Peptide chain elongation	2	97	0.006364829	0.022114814	0.044229627	RPS21
UCH proteinases	2	98	0.006430446	0.022540053	0.045080106	PSMB7

Anchoring of the basal body to the plasma membrane	2	99	0.006496063	0.02296869	0.045252011	CDK1, CKAP5
ABC transporter disorders	2	99	0.006496063	0.02296869	0.045252011	PSMB7
PCP/CE pathway	2	99	0.006496063	0.02296869	0.045252011	PSMB7
Recycling of eIF2:GDP	1	10	6.56E-04	0.023352803	0.045252011	EIF2S1
Nonsense Mediated Decay (NMD) independent of the Exon Junction Complex (EJC)	2	101	0.006627297	0.023836089	0.045252011	RPS21
Eukaryotic Translation Elongation	2	102	0.006692913	0.024274815	0.045252011	RPS21
TNFR2 non-canonical NF-kB pathway	2	104	0.006824147	0.025162236	0.045252011	PSMB7
SUMOylation of immune response proteins	1	11	7.22E-04	0.025658708	0.045252011	UBE2I
Negative regulation of activity of TFAP2 (AP-2) family transcription factors	1	11	7.22E-04	0.025658708	0.045252011	UBE2I
Eukaryotic Translation Termination	2	106	0.006955381	0.026062834	0.045252011	RPS21
RUNX1 regulates transcription of genes involved in differentiation of HSCs	2	106	0.006955381	0.026062834	0.045252011	PSMB7
Nuclear events mediated by NFE2L2	2	110	0.007217848	0.027903013	0.045252011	PSMB7
Hypusine synthesis from eIF5A-lysine	1	12	7.87E-04	0.027959319	0.045252011	DHPS
MASTL Facilitates Mitotic Progression	1	12	7.87E-04	0.027959319	0.045252011	CDK1
Gene expression (Transcription)	9	1893	0.124212598	0.028823614	0.045252011	PSMB7, PRMT5, UBE2I, DDX39B, CHEK1, CDK1

Selenocysteine synthesis	2	112	0.007349081	0.028842325	0.045252011	RPS21
Generic Transcription Pathway	8	1588	0.104199475	0.02927634	0.045252011	PSMB7, PRMT5, UBE2I, CHEK1, CDK1
Cellular responses to stress	6	1009	0.066207349	0.029484406	0.045252011	PSMB7, KHSRP, EIF2S1, RPS21
Viral mRNA Translation	2	114	0.007480315	0.029794272	0.045252011	RPS21
Assembly of the pre-replicative complex	2	114	0.007480315	0.029794272	0.045252011	PSMB7
Folding of actin by CCT/TriC	1	13	8.53E-04	0.030254649	0.045252011	CCT3
DNA replication initiation	1	13	8.53E-04	0.030254649	0.045252011	POLE2
Processing and activation of SUMO	1	13	8.53E-04	0.030254649	0.045252011	UBE2I
MAPK3 (ERK1) activation	1	13	8.53E-04	0.030254649	0.045252011	CDK1
Deubiquitination	3	289	0.018963255	0.030579304	0.045252011	PSMB7, CDK1
Transcriptional regulation by RUNX3	2	118	0.007742782	0.031735545	0.045252011	PSMB7
Cellular responses to stimuli	6	1027	0.067388451	0.031791597	0.045252011	PSMB7, KHSRP, EIF2S1, RPS21
SRP-dependent cotranslational protein targeting to membrane	2	119	0.007808399	0.032228555	0.045252011	RPS21
Nucleotide Excision Repair	2	119	0.007808399	0.032228555	0.045252011	UBE2I, POLE2
CLEC7A (Dectin-1) signaling	2	120	0.007874016	0.032724608	0.045252011	PSMB7
Hedgehog 'off' state	2	124	0.008136483	0.034738935	0.045252011	PSMB7
Downstream TCR signaling	2	124	0.008136483	0.034738935	0.045252011	PSMB7
Condensation of Prometaphase Chromosomes	1	15	9.84E-04	0.03482951	0.045252011	CDK1
Interleukin-1 signaling	2	125	0.0082021	0.035249964	0.045252011	PSMB7

SUMOylation of DNA methylation proteins	1	16	0.001049869	0.037109065	0.045252011	UBE2I
Golgi Cisternae Pericentriolar Stack Reorganization	1	17	0.001115486	0.039383386	0.045252011	CDK1
Maturation of nucleoprotein	1	17	0.001115486	0.039383386	0.045252011	UBE2I
Resolution of Sister Chromatid Cohesion	2	134	0.008792651	0.039979959	0.045252011	CDK1, CKAP5
MAPK1/MAPK3 signaling	3	329	0.021587927	0.042333329	0.045252011	PSMB7, CDK1
SARS-CoV-1 targets host intracellular signalling and regulatory pathways	1	19	0.001246719	0.04391637	0.045252011	UBE2I
Organelle biogenesis and maintenance	3	338	0.022178478	0.045252011	0.045252011	CCT3, CDK1, CKAP5
SUMOylation of transcription factors	1	20	0.001312336	0.046175056	0.046175056	UBE2I
KSRP (KHSRP) binds and destabilizes mRNA	1	20	0.001312336	0.046175056	0.046175056	KHSRP
Phosphorylation of the APC/C	1	20	0.001312336	0.046175056	0.046175056	CDK1
KEAP1-NFE2L2 pathway	2	146	0.009580052	0.046638303	0.046638303	PSMB7
TCR signaling	2	147	0.009645669	0.047210582	0.047210582	PSMB7
TP53 Regulates Transcription of Genes Involved in G2 Cell Cycle Arrest	1	21	0.001377953	0.048428555	0.048428555	CDK1
Host Interactions of HIV factors	2	151	0.009908136	0.049525647	0.049525647	PSMB7

Table S4.5: Functional enrichment analysis of hits identified during second round of CSE exposure (Time point T2) in A549 KO cells using enrichr GO:BP and REACTOME database.

Term	Gene Ontology	Intersection	Total Size	P-value	Adjusted P-value	Genes
deoxyribonucleotide metabolic process	GO:0009262	1	5	0.00149922	0.026172549	RRM1
positive regulation of rRNA processing	GO:2000234	1	10	0.002996578	0.026172549	HEATR1
positive regulation of transcription, DNA-templated	GO:0045893	3	1183	0.003606272	0.026172549	ZBTB17, HEATR1, YEATS4
histone H2A acetylation	GO:0043968	1	16	0.004790942	0.026172549	YEATS4
regulation of rRNA processing	GO:2000232	1	16	0.004790942	0.026172549	HEATR1
nucleotide biosynthetic process	GO:0009165	1	18	0.005388466	0.026172549	RRM1
regulation of transcription by RNA polymerase I	GO:0006356	1	18	0.005388466	0.026172549	HEATR1
positive regulation of transcription by RNA polymerase I	GO:0045943	1	23	0.006880968	0.027717226	HEATR1
nucleobase-containing small molecule interconversion	GO:0015949	1	27	0.008073627	0.027717226	RRM1
carbohydrate derivative biosynthetic process	GO:1901137	1	28	0.008371605	0.027717226	RRM1
maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	GO:0000462	1	30	0.008967338	0.027717226	HEATR1
maturation of SSU-rRNA	GO:0030490	1	35	0.010455365	0.029623535	HEATR1

replication fork processing	GO:0031297	1	39	0.011644447	0.030420498	MMS22L
DNA-dependent DNA replication maintenance of fidelity	GO:0045005	1	43	0.012832338	0.030420498	MMS22L
positive regulation of RNA metabolic process	GO:0051254	1	45	0.013425837	0.030420498	HEATR1
histone H4 acetylation	GO:0043967	1	48	0.014315529	0.030420498	YEATS4
IRE1-mediated unfolded protein response	GO:0036498	1	53	0.015796862	0.031593725	ZBTB17
histone acetylation	GO:0016573	1	71	0.021114315	0.039882595	YEATS4
recombinational repair	GO:0000725	1	79	0.023469932	0.040398888	MMS22L
negative regulation of cell cycle	GO:0045786	1	80	0.023764052	0.040398888	ZBTB17
positive regulation of transcription by RNA polymerase II	GO:0045944	2	908	0.027338863	0.044262921	ZBTB17, YEATS4
double-strand break repair via homologous recombination	GO:0000724	1	97	0.028752804	0.044436152	MMS22L

Pathway name	#Entities found	#Entities total	Entities ratio	Entities pValue	Entities FDR	Submitted entities found
XBP1(S) activates chaperone genes	2	95	0.006233596	0.001061202	0.015568686	ZBTB17
IRE1alpha activates chaperones	2	101	0.006627297	0.001197591	0.015568686	ZBTB17
Unfolded Protein Response (UPR)	2	155	0.010170604	0.002780745	0.025026707	ZBTB17
Activation of the TFAP2 (AP-2) family of transcription factors	1	12	7.87E-04	0.00628188	0.037691279	YEATS4

Table S4.6: Functional enrichment analysis of hits identified during second third of CSE exposure (Time point T3) in A549 KO cells using enrichr GO:BP and REACTOME database.

Term	Gene Ontology	Intersection	Total Size	P-value	Adjusted P-value	Genes
ribosome biogenesis	GO:0042254	8	192	5.46E-06	0.00490652	RPS25, RPL30, TBL3, DDX49, NIP7, RPL11, DDX10, MDN1
rRNA metabolic process	GO:0016072	7	162	1.71E-05	0.007687144	RPS25, RPL30, TBL3, DDX49, RPL11, DDX10, MDN1
rRNA processing	GO:0006364	7	173	2.61E-05	0.007828223	RPS25, RPL30, TBL3, DDX49, RPL11, DDX10, MDN1
ncRNA processing	GO:0034470	7	201	6.78E-05	0.01523817	RPS25, RPL30, TBL3, DDX49, RPL11, DDX10, MDN1
double-strand break repair via synthesis-dependent strand annealing	GO:0045003	2	5	0.000245087	0.044066673	RAD51, XRCC3
regulation of gene expression	GO:0010468	15	1079	0.000294666	0.04415086	PRMT5, MBNL1, RPL11, CXXC1, AHR, CTCF, CDK6, HNRNPK, DNAJA4, RPS3, PCGF1, KRAS, VHL, PTF1A, TP53
Pathway name	**#Entities found**	**#Entities total**	**Entities ratio**	**Entities pValue**	**Entities FDR**	**Submitted entities found**
Metabolism of RNA	19	829	0.054396325	2.13E-05	0.016343191	SF3A3, PRMT5, RPL30, DDX49, NIP7, RPL11, LSM2, RPS25, TBL3, CDK7, HNRNPK, PSMB5, NUP85,

						PSMA2, RPS3, SYMPK, LUC7L3, TNPO1
Cellular responses to stress	20	1009	0.066207349	9.60E-05	0.031059323	RPL30, CRTC3, RPL11, CXXC1, HSPA14, RPS25, PSMB5, CDK6, NUP85, DNAJA4, PSMA2, RPS3, VHL, TLN1, TP53, MAP4K4
Cellular responses to stimuli	20	1027	0.067388451	1.22E-04	0.031059323	RPL30, CRTC3, RPL11, CXXC1, HSPA14, RPS25, PSMB5, CDK6, NUP85, DNAJA4, PSMA2, RPS3, VHL, TLN1, TP53, MAP4K4
Regulation of RAS by GAPs	5	71	0.004658793	2.48E-04	0.046845815	PSMB5, PSMA2, KRAS
Regulation of TP53 Expression	2	4	2.62E-04	4.66E-04	0.046845815	TP53
Regulation of expression of SLITs and ROBOs	7	183	0.012007874	5.94E-04	0.046845815	RPS25, RPL30, PSMB5, PSMA2, RPL11, RPS3
Major pathway of rRNA processing in the nucleolus and cytosol	7	189	0.012401575	7.17E-04	0.046845815	RPS25, TBL3, RPL30, DDX49, NIP7, RPL11, RPS3
Cell Cycle, Mitotic	13	596	0.039107612	7.50E-04	0.046845815	CDK7, PSMB5, CDK6, NUP85, PSMA2, TUBB, TAOK1, SLITRK5, MCM5, TNPO1, TP53, ZWINT
Downstream signaling events of B Cell Receptor (BCR)	5	91	0.005971129	7.59E-04	0.046845815	PSMB5, PSMA2, KRAS
Signaling by RAS mutants	4	54	0.003543307	8.77E-04	0.046845815	KRAS, TLN1
Paradoxical activation of RAF signaling by kinase inactive BRAF	4	54	0.003543307	8.77E-04	0.046845815	KRAS, TLN1

Signaling by moderate kinase activity BRAF mutants	4	54	0.003543307	8.77E-04	0.046845815	KRAS, TLN1
Signaling downstream of RAS mutants	4	54	0.003543307	8.77E-04	0.046845815	KRAS, TLN1
Autodegradation of the E3 ubiquitin ligase COP1	4	54	0.003543307	8.77E-04	0.046845815	PSMB5, PSMA2, TP53
Signaling by NTRK1 (TRKA)	6	143	0.009383202	9.19E-04	0.046845815	FOSL1, KRAS, AP2M1, CDK5R1

Table S4.7: Hits identified in BCi NS1.1 KO cells during CSE exposure using MAGeCK.

id	num	pos|score	pos|p-value	pos|fdr	pos|rank	pos|goodsgrna	pos|lfc
SEMA4F	3	2.05E-05	2.97E-07	8.00E-05	1	3	4.1074
SNAP91	1	9.22E-05	2.97E-07	8.00E-05	4	1	1.428
GYPB	1	0.00011849	2.97E-07	8.00E-05	6	1	5.1195
CABP2	1	0.00022381	2.97E-07	8.00E-05	9	1	5.0255
C1orf145	1	0.00036862	2.97E-07	8.00E-05	16	1	4.9077
EBPL	1	0.00046078	2.97E-07	8.00E-05	19	1	4.8357
LDHC	1	0.001277	2.97E-07	8.00E-05	49	1	4.4583
BTN3A2	1	0.001435	2.97E-07	8.00E-05	55	1	0.92309
SRSF11	1	0.001514	2.97E-07	8.00E-05	56	1	4.4096
PLEKHM1	1	0.001751	2.97E-07	8.00E-05	67	1	4.3846
COG6	1	0.001751	2.97E-07	8.00E-05	68	1	4.3846
GDAP2	1	0.0019748	2.97E-07	8.00E-05	73	1	4.3592
ZNF750	1	0.0024092	2.97E-07	8.00E-05	90	1	4.3069
NFE2L1	1	0.0036336	2.97E-07	8.00E-05	128	1	4.1963
PEX1	1	0.0036336	2.97E-07	8.00E-05	129	1	4.1963
SLCO2A1	1	0.0036336	2.97E-07	8.00E-05	130	1	4.1963
TSR3	1	0.0041075	2.97E-07	8.00E-05	146	1	4.1673

HAUS8	1	0.0046078	2.97E-07	8.00E-05	163	1	4.1377
DKK3	1	0.0046078	2.97E-07	8.00E-05	164	1	4.1377
CD207	1	0.0046078	2.97E-07	8.00E-05	165	1	4.1377
GRK7	1	0.0046078	2.97E-07	8.00E-05	166	1	4.1377
UHRF1BP1L	1	0.0048579	2.97E-07	8.00E-05	180	1	4.1198
ZNF638	1	0.005503	2.97E-07	8.00E-05	194	1	4.0765
RNF113A	1	0.0058585	2.97E-07	8.00E-05	205	1	4.045
OTX2	1	0.0064246	2.97E-07	8.00E-05	234	1	4.0127
FBLN1	1	0.0064246	2.97E-07	8.00E-05	235	1	4.0127
EXOSC8	1	0.0071092	2.97E-07	8.00E-05	251	1	3.9797
GPR68	1	0.0075831	2.97E-07	8.00E-05	268	1	3.9333
C11orf53	1	0.0078464	2.97E-07	8.00E-05	274	1	3.9459
GLYCTK	1	0.0078464	2.97E-07	8.00E-05	275	1	3.9459
C1orf109	1	0.0078464	2.97E-07	8.00E-05	276	1	3.9459
RPL3L	1	0.0080702	2.97E-07	8.00E-05	286	1	3.6591
NT5C	1	0.0081492	2.97E-07	8.00E-05	289	1	3.7591
WDR44	1	0.0083467	2.97E-07	8.00E-05	303	1	3.9113
KRT7	1	0.0088997	2.97E-07	8.00E-05	315	1	3.8759
SOX9	1	0.0088997	2.97E-07	8.00E-05	316	1	3.8759
TAF1A	1	0.0088997	2.97E-07	8.00E-05	317	1	3.8759
GALR1	1	0.0093209	2.97E-07	8.00E-05	331	1	3.842
COL15A1	1	0.0096369	2.97E-07	8.00E-05	345	1	3.8395
C5orf34	1	0.0096369	2.97E-07	8.00E-05	346	1	3.8395
APOC4	1	0.0096369	2.97E-07	8.00E-05	347	1	3.8395
HS3ST1	1	0.009887	2.97E-07	8.00E-05	351	1	3.7537
ZBTB11	1	0.010519	2.97E-07	8.00E-05	370	1	3.8023
NLRP4	1	0.010519	2.97E-07	8.00E-05	371	1	3.8023
C20orf194	1	0.010519	2.97E-07	8.00E-05	372	1	3.8023
PBRM1	1	0.010519	2.97E-07	8.00E-05	373	1	3.8023
BRD7	1	0.01119	2.97E-07	8.00E-05	397	1	3.764

LHX8	1	0.01119	2.97E-07	8.00E-05	398	1	3.764
STK17A	1	0.01119	2.97E-07	8.00E-05	399	1	3.764
NR2C2AP	1	0.012072	2.97E-07	8.00E-05	425	1	3.7247
CASP8	1	0.012757	2.97E-07	8.00E-05	445	1	3.6843
PIGP	1	0.012757	2.97E-07	8.00E-05	446	1	3.6843
PLCXD3	1	0.013428	2.97E-07	8.00E-05	465	1	3.6427
TEX33	1	0.013428	2.97E-07	8.00E-05	466	1	3.6427
SUMO4	1	0.013428	2.97E-07	8.00E-05	467	1	3.6427
SEC14L3	1	0.013428	2.97E-07	8.00E-05	468	1	3.6427
SNAPC4	1	0.014021	2.97E-07	8.00E-05	485	1	3.0411
SPARC	1	0.014074	2.97E-07	8.00E-05	486	1	3.4524
TRAPPC8	1	0.014337	2.97E-07	8.00E-05	499	1	3.5999
UNC13A	1	0.014784	2.97E-07	8.00E-05	510	1	2.2293
TLK2	1	0.014995	2.97E-07	8.00E-05	515	1	3.5558
DHPS	1	0.015206	2.97E-07	8.00E-05	519	1	3.4692
ETV3L	3	3.98E-05	8.91E-07	0.000232	3	3	2.7487
YTHDF1	3	9.69E-05	9.80E-06	0.002513	5	3	3.6386
FLYWCH2	3	0.00012412	1.04E-05	0.002625	7	3	2.9135
TMEM182	3	0.0002362	9.77E-05	0.024309	10	2	3.8759
ZNF536	2	0.00028961	0.00013691	0.027615	12	1	1.6424
TOE1	2	0.00034227	0.00013691	0.027615	15	1	0.5475
TMEM87B	2	0.00043707	0.00013691	0.027615	18	2	3.3744
ST3GAL3	2	0.00046787	0.00013691	0.027615	20	2	2.492
FRRS1	2	0.00050021	0.00013691	0.027615	22	2	2.7157
ALG13	2	0.00069467	0.00013691	0.027615	27	2	3.1693
DDR2	2	0.00073711	0.00013691	0.027615	28	1	2.434
FADS2	2	0.00090812	0.00013751	0.027615	32	2	3.5676
DECR2	2	0.00092566	0.00013751	0.027615	33	2	2.7967
CDK20	2	0.0010008	0.00013751	0.027615	34	2	3.2082
ARL4C	2	0.0010266	0.00013751	0.027615	35	1	1.1861

MAS1L	2	0.0011319	0.00013751	0.027615	41	1	0.73932
CHST13	2	0.0011319	0.00013751	0.027615	42	1	2.2936
IMMP2L	2	0.0012504	0.00013751	0.027615	46	2	3.4766
ZNF512	2	0.0012635	0.00013751	0.027615	47	1	1.0804
ACOT7	2	0.0012635	0.00013751	0.027615	48	1	1.9502
STRBP	3	0.00030304	0.00018502	0.035977	13	2	3.7941
SLC35F5	3	0.00032962	0.00018502	0.035977	14	3	3.8023
CTNND1	3	0.00037778	0.00018562	0.035977	17	3	0.87593
KIAA1522	4	0.00015797	0.00027887	0.053431	8	2	0.66727
SEMA4C	3	0.00053613	0.00052418	0.084111	23	2	3.6427
OR6K2	3	0.00053613	0.00052418	0.084111	24	2	3.6427
CBL	3	0.00054461	0.00052418	0.084111	25	3	0.98035
ISYNA1	3	0.00059232	0.00052478	0.084111	26	2	0.30049
EPHA5	2	0.0013174	0.00052003	0.084111	52	2	2.8831
CPD	2	0.0016592	0.00052003	0.084111	60	2	2.769
ADSS	2	0.0018686	0.00052003	0.084111	71	2	0.53147
WDR47	2	0.0020785	0.00052003	0.084111	75	2	2.9085
PPIA	2	0.0022189	0.00052062	0.084111	77	2	2.3825
ANKRD61	2	0.0022189	0.00052062	0.084111	78	2	2.5261
RGS4	2	0.0022276	0.00052062	0.084111	79	2	2.4684
TTC39C	2	0.0022876	0.00052062	0.084111	81	2	2.5204
C1orf123	2	0.0023689	0.00052062	0.084111	87	2	2.0652
ELK4	2	0.0025146	0.00052062	0.084111	92	2	2.7158
BBOX1	2	0.0025524	0.00052062	0.084111	93	2	2.5942
TIMM17A	2	0.0027111	0.00052062	0.084111	99	2	2.5551
JOSD2	2	0.0027891	0.00052062	0.084111	103	2	2.932