



FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

Friedrich-Schiller-University Jena
Department of Informatics and Mathematics
Institute for Computer Science

Analysis of Gene Regulation in Respiratory Syncytial Virus Infected Cells

for the attainment of the academic degree of
Master of Science (M.Sc.)

Presented by Fionn Daire Keogh
MatNr. 176219
Born March 18th 1997 in Lörrach

1st Supervisor: Prof. Dr. Rainer König
2nd Supervisor: Dr. Emanuel Barth

Jena, June 11th 2024

Abstract

Respiratory Syncytial Virus (RSV) is one of the most potent causes of hospitalization with acute respiratory infections. It is also a leading cause of death in infants under the age of one year, the elderly, and immunocompromised patients. Rising numbers in infections over the last decade call for new therapeutic interventions. Two vaccines are newly approved. However, both do not target children directly under the age of 6 months or at all above 6 months. They are also not suitable for immune-compromised patients. Finding host genes that regulate changes induced by RSV could present potent targets for new therapeutic interventions. Analyzing gene expression data and performing a gene set enrichment analysis confirmed the findings of other works found in the literature describing the upregulation of morphogenesis. Analysis of host dependency and restriction factors, in conjunction with the expression datasets, identified cell death and cytokine-related pathways as significant areas for further investigation. Using then a gene regulatory network, regression models were optimized with mixed integer linear programming that predicted gene expression values in dependency on the activity profile of transcription factors. These models revealed 7 regulators for cell death and cytokine related pathways – ATF2, IRF3, MAFF, POU5F1, SIX5, STAT2 and TFAP2C – that, when employed in the models, predict the expression of genes of these processes most accurately.

Zusammenfassung

Das Respiratorische Syncytial Virus (RSV) ist eine der häufigsten Ursachen für Hospitalisierungen mit akuten Atemwegsinfektionen. Zusätzlich ist es eine der häufigsten Todesursachen bei Kleinkindern unter dem Alter von einem Jahr, älteren Menschen und bei Patienten mit einem geschwächten Immunsystem. Die steigenden Zahlen der letzten zehn Jahre verlangt nach neuen Therapiemöglichkeiten. Vor kurzem wurden zwei neue Impfstoffe zugelassen. Diese sind jedoch nur indirekt für Kinder unter 6 Monaten zugelassen und gar nicht für Kinder älter als 6 Monate. Patienten mit einem geschwächten Immunsystem profitieren auch nicht von den Vakzinen. Die Identifizierung von Wirtsfaktoren, welche die durch RSV ausgelösten Veränderungen regulieren, könnte wirksame Ziele für neue medikamentöse Therapien darstellen. Die Analyse von Genexpressionsdaten und die Durchführung einer Genmengenanreicherung bestätigten die Ergebnisse anderer Publikationen, in denen die Hochregulierung der Morphogenese beschrieben wurde. Die Analyse von Wirtsabhängigkeits- und Restriktionsfaktoren in Verbindung mit Expressionsdatensätzen identifizierte Zell Tod und Zytokin-assoziierte Signalwege als bedeutende Bereiche für weitere Untersuchungen. Mittels eines genregulatorischen Netzwerks wurden dann Regressionsmodelle mit gemischt-ganzzahliger linearer Programmierung optimiert und die Genexpressionswerte in Abhängigkeit vom Aktivitätsprofil von Transkriptionsfaktoren vorhergesagt. Die Modelle ergaben 7 Regulatoren für Zelltod- und Zytokin-basierte Pfade – ATF2, IRF3, MAFF, POU5F1, SIX5, STAT2 and TFAP2C – die die Genexpression am genauesten vorhersagen konnten.

Contents

1	Introduction	12
1.1	Current Cases of Respiratory Syncytial Virus Infections	12
1.2	Respiratory Syncytial Virus	14
1.2.1	Nonstructural Proteins	14
1.2.2	Genomic Stability Proteins	16
1.2.3	Envelop Structure	17
1.2.4	Surface Proteins	17
1.2.5	Viral Life Cycle	19
1.3	Clinical Relevance and Medical Intervention	19
1.3.1	Antibodies	19
1.3.2	Antiviral Drugs - Ribavirin	20
1.3.3	Vaccines	20
1.3.4	Host Factors	21
1.4	Gene Regulatory Networks	22
1.5	Objectives	23
2	Materials & Methods	24
2.1	Datasets	24
2.1.1	Gene Expression Data	24
2.1.2	Host Factor Screening Data	25
2.2	Differential Gene Expression Analysis	26
2.2.1	Data Exploration	26
2.2.2	Normalization	26
2.2.3	Gene Expression Analysis	27
2.3	Gene Set Enrichment Analysis Pipeline	27
2.3.1	Gene Set Enrichment Analysis	27
2.3.2	Redundancy Reduction	28
2.3.3	Categorization of Enriched Pathways	29
2.4	Prediction of Transcription Factors using a Gene Regulatory Network Model	30
3	Results	33
3.1	Data Exploration Leads to the Removal of Duplicate Samples	33
3.2	Gene Expression Shows Upregulation of Host Restriction Factors	36

3.3	Gene Set Enrichment Analysis Shows Upregulation of Genes Involved in Morphogenesis and Immune Response	38
3.4	Comparing Gene Expression Profiles with Data from Host Factor Screens Shows High Overlap in Genes Involved in Cytokine Related Pathways and Cell Death	44
3.5	Gene Regulatory Models Predict MAFF, TFAP2C, IRF3 and POU5F1 Regulating Cytokine Related Pathways due to Virus Infection	50
3.6	Gene Regulatory Models Predict STAT2, ATF2, TFAP2C and SIX5 Regulating Cell Death Related Pathways	54
4	Discussion	61
4.1	Predicted Transcription Factors	61
4.1.1	Cytokine Pathways are regulated by MAF BZIP Transcription Factor F	61
4.1.2	Cytokine Pathways are regulated by Interferon Regulatory Factor 3	62
4.1.3	Cytokine Pathways are regulated by POU Class 5 Homeobox 1 .	62
4.1.4	Cytokine Pathways and Cell Death are predicted to be regulated by Transcription Factor AP-2 γ	63
4.1.5	Cell Death is regulated by Signal Transducer and Activator of Transcription 2	64
4.1.6	Cell Death is regulated by Activating Transcription Factor 2 .	64
4.1.7	Cell Death is regulated by SIX Homeobox 5	65
4.2	Limitations of the Study	65
4.2.1	Handling of the Data Sets	65
4.2.2	HDFs versus Expressed HDFs	66
4.2.3	Current Network Model and Prediction of TFs	66
5	Conclusion and Future Outlook	67
6	Supplemental Material	93

List of Figures

1	RSV infections in the EEA and USA from 2018 - 2023	13
2	RSV infections of children younger than one year in the USA and Germany from 2018 - 2023	14
3	RSV Genome Sketch	15
4	Hierarchical clustering of the expression data	34
5	PCA of the expression data	35
6	Volcano plot of hSAEC late infected vs control group	37
7	Upset plot of the redundancy reduced upregulated gene sets	39
8	Illustrative example of redundancy reduction	39
9	Upset plot of the redundancy reduced downregulated gene sets	41
10	Pie charts of the upregulated categories	42
11	Pie charts of the downregulated categories	43
12	Pie chart of the combined upregulated categories from hSAEC with HF as Bar plot	45
13	Correlation plot of the (predicted) gene expression	49
14	Frequency of the TFs selected for the prediction of <i>cytokine pathway</i> genes	50
15	Activity of the four TFs IRF3, Maff, TFAP2C and POU5F1 over time for the hSAEC data.	53
16	Cytokine network	54
17	Frequency of the TFs selected for the prediction of <i>cell death</i> genes	55
18	Activity of the four TFs SIX5, STAT2, TFAP2C and ATF2 for the hSAEC data.	59
19	Cell death network	60
S1	Hierarchical clustering of the A549 data set	93
S2	Hierarchical clustering of the Hep-2 data set	94
S3	Hierarchical clustering of the hSAEC data set	95
S4	Volcano plot of the A549 data set	96
S5	Volcano plot of the Hep-2 data set	97
S6	Volcano plot of the hSAEC data set to the early time point	98
S7	Auxiliary plots for the redundancy reduction of upregulated A549 gene sets	99
S8	Auxiliary plots for the redundancy reduction of downregulated A549 gene sets	99

LIST OF FIGURES

S9	Auxiliary plots for the redundancy reduction of upregulated HEp-2 gene sets	100
S10	Auxiliary plots for the redundancy reduction of downregulated HEp-2 gene sets	100
S11	Auxiliary plots for the redundancy reduction of upregulated hSAEC gene sets at the early time point	101
S12	Auxiliary plots for the redundancy reduction of downregulated hSAEC gene sets at the early time point	101
S13	Auxiliary plots for the redundancy reduction of upregulated hSAEC gene sets at the late time point	102
S14	Auxiliary plots for the redundancy reduction of downregulated hSAEC gene sets at the late time point	102

List of Tables

1	Up- and downregulated HDFs and HRFs in gene expression data	37
2	Remaining genes after reduction of upregulated gene sets.	38
3	Remaining genes after reduction of downregulated gene sets.	40
4	Ordered list of categories by the ratio of unique HDFs to genes	46
5	Gene sets of the sub-category cytokine pathways.	47
6	Gene sets of the category cell death.	48
7	Genes regulated by the 4 predicted transcription factors related to cytokine pathways.	52
8	PCC <i>cytokine pathways</i> TFs	53
9	Genes regulated by the 4 predicted transcription factors related to <i>cell death</i>	57
10	Regulatory interactions between TFs and selected TGs of apoptosis and autophagy.	58
11	PCC <i>cell death</i> TFs	59

Abbreviations

A3G Apolipoprotein B mRNA Editing Enzyme Catalytic Subunit 3G

ABCE1 ATP binding cassette E 1

ANOVA analysis of variance

ARP2 Actin-related protein 2

ATF2 Activating Transcription Factor 2

ATP1A1 ATPase NA⁺/K⁺ transporting subunit alpha 1

BA RSV B Buenos Aires

Cav2 Caveolin 2

CCD central conserved domain

CD14 Monocyte differentiation antigen CD 14

CD8+ cells cytotoxic T lymphocytes

CDK4 Cyclin-dependent kinase 4

Cof1 Cofilin 1

CPM count per million

CR Cox-Reid profile-adjusted likelihood

CRE Cyclic AMP-responsive element

CREBBP CREB Binding Protein

Crm1 Exportin 1

CX3C C-X3-C chemokine

CX3CR1 CX3C chemokine receptor

CXCL4 CXC Chemokine Ligand 4

DEDD2 Death Effector 2

LIST OF TABLES

- DEGs** differentially expressed genes
- DGE** Differential Gene Expression
- E3 ligase** ubiquitin ligase
- EGFR** Epidermal growth factor receptor
- eIF α** eukaryotic initiation factor alpha
- EMA** European Medicines Agency
- ER** endoplasmic reticulum
- EU** European Union
- FBS** fetal bovine serum
- FDA** Food & Drug Administration
- FDR** false discovery rate
- g:SCS** Set Counts and Sizes
- GA1** RSV A Tracy
- GB1** RSV B 18537
- GBP5** Guanylate Binding Protein 5
- GBS** Guillain Barré syndrome
- GEO** Gene Expression Omnibus
- GFP** green fluorescent protein
- gHRN** generic human regulatory network
- GLMs** generalized linear models
- GMT** Gene Matrix Transposed
- GO** Gene Ontology
- GO:BP** Gene Ontology Biological Process
- GO:CC** Gene Ontology Cellular Component
- GO:MF** Gene Ontology Molecular Function
- GRCh38** human genome hg38
- GRN** gene regulatory network
- GRNs** gene regulatory networks

LIST OF TABLES

- GSEA** gene set enrichment analysis
- HBD** heparin-binding domain
- HBV** Hepatitis B virus
- HDF** host dependency factors
- HEK293** human embryonic kidney cells
- HF** host factors
- HNF-4 α** Hepatocyte Nuclear Factor 4 α
- HRF** host restriction factors
- hSAEC** human small airway epithelial cells
- Hsp70** heat shock protein 70
- HSPGs** Heparan sulfate proteoglycans
- HTRIdb** Human Transcriptional Regulation Interactions database
- ICAM-1** Intercellular adhesion molecule 1
- IFI44** Interferon-induced protein 44
- IFI44L** Interferon induced protein 44-like
- IFITM** Interferon-induced transmembrane protein
- IFN** interferon
- IGF1R** Insuline-like growth factor 1 receptor
- IKK ϵ** nuclear factor kappa B kinase subunit epsilon
- IL** interleukins
- IL-10** interleukin-10
- IRF** interferon regulatory factor
- IRF3** interferon regulatory factor 3
- IRF7** interferon regulatory factor 7
- ISGF3** Interferon Stimulated Gene Factor 3
- ISMARA** Integrated System for Motif Activity Response Analysis
- JNKs** MAP kinases
- LFC** log2-fold change

LIST OF TABLES

- M-CSF** Macrophage colony-stimulating factor
- MAD** median absolute deviation
- MAFF** MAF BZIP Transcription Factor F gene
- MARA** motif activity response analysis
- MAVS** mitochondrial antiviral-signaling protein
- MDA-5** melanoma-differentiation-associated gene 5
- MEM** minimum essential medium
- MIF** Macrophage Migration Inhibition Factor
- MIPRIP** Mixed Integer linear Programming based Regulatory Interaction Predictor
- MOI** multiplicity of infection
- nBreg** neonatal regulatory B lymphocytes
- NF- κ B** nuclear factor kappa B
- NLP** natural language processing
- Npm** nucleophosmin
- ON** RSV A Ontario
- ORF** open reading frame
- PBS** phosphate buffered saline
- PCA**
- PCC** Pearson's correlation coefficient
- PCR** Polymerase chain reaction
- PKR** protein kinase R
- POU5F1** POU class 5 homeobox 1 gene
- Rab5a** Ras-related protein Rab-5A
- RACK1** receptor of activated C kinase 1
- RdRp** RNA-dependent RNA polymerase
- RF** random forest
- RIG-I** retinoic acid-inducible gene I

LIST OF TABLES

- RNPs** ribonucleoprotein complexes
- RSV** Respiratory Syncytial Virus
- sG** truncated G protein
- SIX5** Six Homeobox 5
- STAT2** Signal Transducer and Activator of Transcription 2
- TBA** Total Binding Affinity
- TCC** tag count comparison
- TERT** telomerase reverse transcriptase
- TF** transcription factor
- TFAP2C** Transcription Factor AP-2 γ
- TFBS** transcription factor binding site
- TG** target gene
- TLR** Toll-like receptor
- TLR4** Toll-like receptor 4
- Tm** tropomyosin
- TMM** trimmed mean of M-values
- TNF- α** tumor necrosis factor-alpha
- TNFSF10 or TRAIL** Tumor Necrosis Factor Superfamily member 10
- TRAF3** tumor necrosis factor receptor-associated factor 3
- TRM** median of ratios
- UK** United Kingdom
- UPGMA** unweighted pair group method with arithmetic mean
- USA** United States of America
- VEGFs** vascular endothelial growth factors
- XAF1** X-linked inhibitor of apoptosis (XIAP)-associated factor 1

1. Introduction

1.1 Current Cases of Respiratory Syncytial Virus Infections

The last few years have shown how viral infections can influence events around the world and what devastating effects they can have.

In October 2022, the European Center for Disease Prevention and Control as well as the World Health Organization highlighted "the risk of potential co-circulation of COVID-19 and influenza putting increased pressure on both hospitals and healthcare workers. Elevated levels of Respiratory Syncytial Virus (RSV) circulation coinciding with peaks of these viruses could therefore add additional pressure on the system" [1].

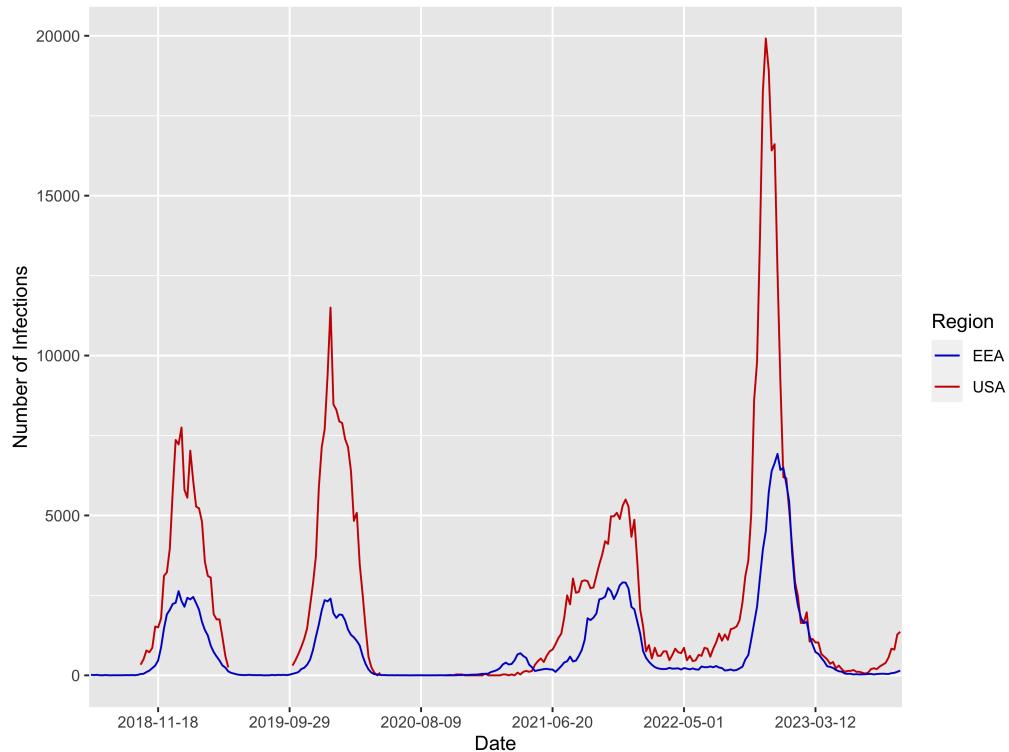


Figure 1: RSV infections in the European Economic Area (EEA, blue) and in the United States of America (USA, red) from July 2018 to August 2023. There is a visible dent in infections in the 2020/2021 season.

Additionally, data made available by the European Union [2] and the Center for Disease Control of the United States [3] shows the rise of cases of human Respiratory Syncytial Virus (RSV) infections across both areas in the last years. Figure ?? shows a reduction in cases only during the COVID-19 pandemic in 2020 and 2021. This can be explained by the strict hygiene measures during the years 2020 and 2021. RSV is one of the most potent causes of hospitalization with acute respiratory infections and a leading cause of death in infants under the age of one year, the elderly, and immunocompromised patients [4]. Figure ?? shows the number of infections per 1,000 children under one year of age in the United States [3] and in Germany [5]. The numbers show an upward trend and resemble a similar structure to the numbers for all ages.

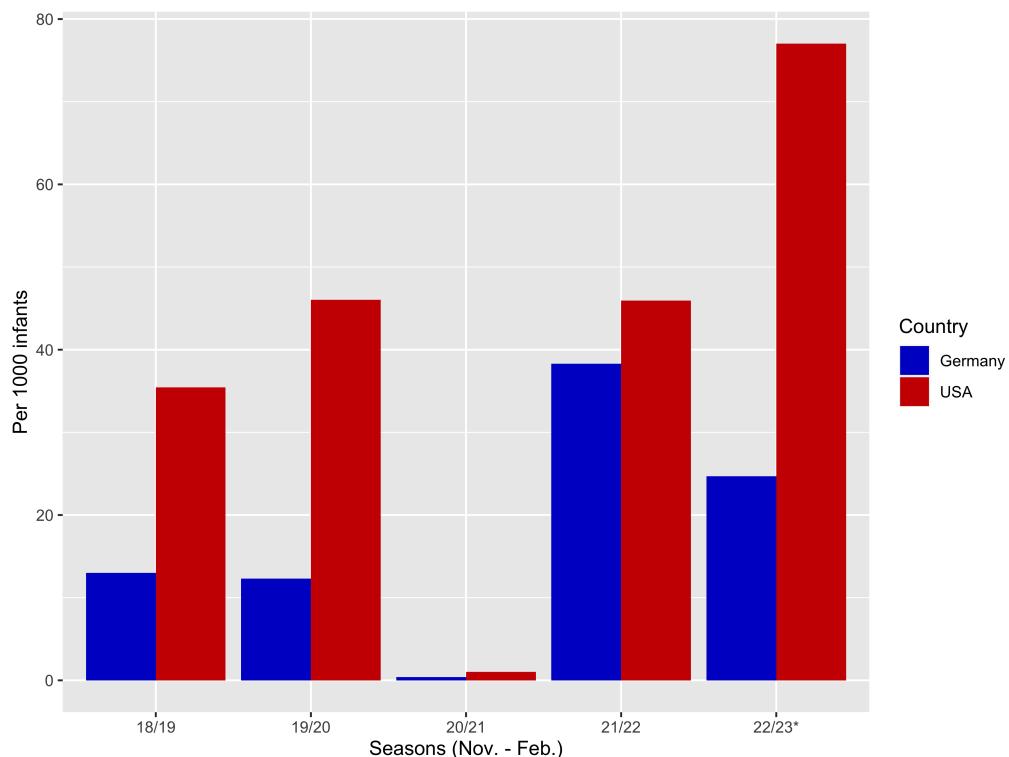


Figure 2: RSV infections of children younger than one year in the USA (red) and Germany (blue). The low infection numbers of the 20/21 season are explained by the measures against the COVID-19 pandemic. * The numbers for Germany are not complete. The season 22/23 has data only until the 31st of December 2022.

1.2 Respiratory Syncytial Virus

RSV is a member of the Orthopneumovirus genus in the family of Pneumoviridae within the order Mononegavirales [6]. Pneumoviridae are large enveloped negative-strand RNA viruses. Their genome contains 8-10 transcriptional elements. The RSV genome is not-segmented and about 15.2 kb in length [7]. Being a member of the Orthopneumovirus genus, RSV has 10 transcriptional elements in the order 3'-NS1-NS2-N-P-M-SH-G-F-M2-L-5' (see Figure ??), coding for 11 proteins. This is due to an overlap between the M2 and the L protein. A second open reading frame allows the M2 protein to encode two separate proteins, M2-1 and M2-2.

1.2.1 Nonstructural Proteins

The nonstructural proteins NS1 and NS2 are 139 and 124 amino acids long, respectively [8], [9], [10]. They are the first two proteins which are transcribed [11]. Although, both proteins improve virus growth, they are not essential [4], [12]. NS1 can not only be found in the cytoplasm but also in the nucleus [13]. Because NS1 is found in the same regions as other nuclear proteins it is strongly suggested that

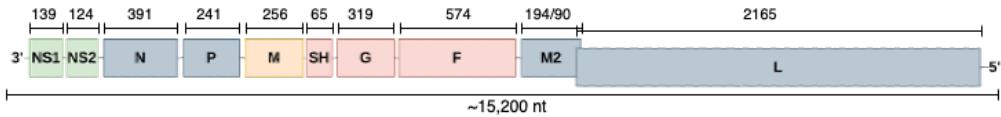


Figure 3: Sketch of the RSV Genome with proportional gene sizes. Green are the nonstructural proteins, blue the structural proteins, orange the envelop protein, and red the surface proteins.

active transportation into the nucleus is the primary way of localization into the nucleus. Even though it would fit through the nuclear pore complex, which allows for 50 kDa in size for passive transport [14], with its smaller protein size of about 15 kDa. Moreover, Pei *et al.* propose that NS1 has a regulatory effect on the gene expression of the host, due to interactions with the mediator complex and chromatin. NS1 inhibits virus RNA translation and transcription [10]. It does not hinder the packaging of RNA, suggesting that the effect of NS1 is prevalent after this step. NS1 affects the RNA polymerase in an early stage of the process [10]. Levels of inhibition of nuclear factor kappa B kinase subunit epsilon (IKK ϵ) were reduced by NS1, however they were not decreased by NS2 [15]. IKK ϵ is a kinase, usually responsible for interferon regulatory factor (IRF) activation. Both proteins, particularly NS1, can decrease the expression of tumor necrosis factor receptor-associated factor 3 (TRAF3), a critical intersection point between the retinoic acid-inducible gene I (RIG-I) and Toll-like receptor (TLR) pathways, thereby leading to the induction of interferon (IFN)[15].

NS2 seems to induce autophagy, through the increased expression of LC3II [16]. LC3II is part of the autophagosome. Higher levels of LC3II are also correlated with a block of autophagic flux. However, Chiok *et al.* [16] showed that NS2 does not block autophagy and thus the higher levels of LC3II must be due to the induction of autophagy. It is proposed that this occurs due to NS2 impairing the ISGylation of Beclin1, a key positive regulator of autophagy. The ISGylation of Beclin1 suppresses its pro-autophagic activity. ISGylation itself is a potent antiviral defense mechanism of the innate immune system [17]. In mice, the NS2 protein has been shown to suppress the production of cytotoxic T lymphocytes (CD8+ cells). This is due to a suppression of type I IFN response by NS2 [18]. In line with this, NS2 decreases STAT2 levels [15], which is an essential transcription factor for the expression of IFN genes. Spann *et al.* have shown that Δ NS1, Δ NS2 and Δ NS1/NS2 knockout mutants of human RSV lacked the ability to repress type I IFN response [19].

Atreya *et al.*[10] suggest that NS1 and NS2 do not depend on each other, but their effects are additive. They interact with the mitochondrial antiviral-signaling protein (MAVS) and RIG-I leading to an inhibited activation of interferon regulatory factor 3 (IRF3) which, in turn, reduced the nuclear factor kappa B (NF- κ B) signal[20]. However, unexpectedly reduced NF- κ B activation was reported in Vero cells infected with RSV with NS2 and NS1/NS2 knockdowns, even though IFN was not present[19]. NS1 and NS2 activity inhibits IFN synthesis and increases anti-apoptotic gene expression, which is an inflammatory response, ending in potential necroptosis, an inflammatory cell death. Due to an extended cellular lifespan, viral

growth is promoted[20].

1.2.2 Genomic Stability Proteins

The nucleoprotein (N) is 391 amino acids long[8]. It is part of the viral replication complex further consisting of the P and L proteins. It interacts non-specifically with RNA.

Oliveira *et al.*[21] promote the idea of interactions between heat shock protein 70 (Hsp70) and the N protein similar to the interactions between the measles N protein and the Hsp70. In addition, the N protein is attributed to have an immunomodulating role[20]. The proteins induce the formation of inclusion bodies. These trap MAVS and melanoma-differentiation-associated gene 5 (MDA-5) and inactivate them. MDA-5 is a member of the RIG-I-like receptor family and also activates IRF3 in downstream signaling[22]. The N protein also interacts with the protein kinase R (PKR) reducing the phosphorylation of eukaryotic initiation factor alpha (eIF α) and thus reducing the eIF α -induced stress response[23]. The interaction between the N protein and the phosphoprotein (P) allows for a more specific interaction with the viral RNA[21].

The P protein is 241 amino acids long[24]. It can have up to five phosphorylations in two serine residue clusters at positions 116, 117, 119, and 232, 237[25]. Oliveira *et al.*[21] showed that the P protein also interacts with the Hsp70 protein. They also showed that there is an interaction between the P protein and tropomyosin (Tm), which is an important protein for the organization of the cytoskeletal microfilaments, most notably for F-actin[21]. Elimination of all five phosphorylations of the P protein reduced P protein reporter gene expression by 20% [25]. Mutation of all five serine positions reduced the interaction between the N and the P protein by 60%. The interaction between N, P, and M2-1 proteins is essential for virus replication[7]. It also stabilizes the L protein[24].

The L protein is the major subgroup of the viral RNA polymerase[24]. It is 2,165 amino acids long. The L protein is also called the RNA-dependent RNA polymerase (RdRp). The enzymatic activity required for RNA replication is mainly associated with the L protein[26]. Six conserved regions have been identified in L proteins of non-segmented RNA viruses[27]. They can be grouped into two groups, RNA synthesis regions and cap formation regions. The regions for cap formation are also required for RNA synthesis. The L protein gene has a 68 nucleotide overlap region with the M2 gene [28]. This overlap and a second open reading frame (ORF) allow the M2 gene to code for two separate proteins, the M2-1 and M2-2 proteins[29].

The M2-1 protein is 194 amino acids long and is transcribed from the first ORF. It acts as a transcription elongation factor[30]. It co-localizes with the M, N, and P proteins in the cytoplasm. M2-1 forms an unusual CCCH zinc-finger motif[31]. It competes with the P protein for mRNA binding[32]. M2-1 is also an antitermination factor and increases the readthrough frequency[29]. The M2-2 protein is coded by the second ORF and is 83 to 90 amino acids in length[29]. It is co-localized with ribosomal fractions and the endoplasmic reticulum (ER)[33]. M2-2 interacts with cellular proteins to negatively regulate translation, protein folding, and mRNA splicing. From analysis of MS/MS data, Scudero *et al.* found 72 potential host

proteins that M2-2 could interact with. Bermingham & Collins[29] report in their work that M2-2 has a negative regulatory effect on RNA transcription but a positive effect on RNA replication. It is also suggested that M2-2 may have an antagonistic effect on the host's activation of eIF α stress response, this however needs to be validated[33].

1.2.3 Envelop Structure

The matrix (M) protein is a 256 amino acid long protein[8]. It plays a role in the assembly and budding of the virus[21]. During viral assembly, it acts as a bridge between the lipid bilayer envelope and the nucleocapsid[34]. It mediates the transport of ribonucleoprotein complexes (RNPs) to the assembly site at the cell surface. For this, the M protein interacts directly with polymerized actin. Oliveira *et al.*[21] showed that the M protein interacts with Tm to re-organize the cytoskeletal microfilaments (including actin) to aid virus budding. The M protein is localized in the nucleus during early infection[35][36] and is found in cytoplasmic inclusions in infected cells[30]. The import into the nucleus is facilitated by the amino acids 110-183 which include binding regions sufficient for nuclear targeting[35]. Importin $\beta 1$ directly recognizes the M protein and mediates its import alongside Ran. Localized to the nucleus it inhibits host transcription. To note, the import is not dependent on importin α . At later stages of infection, the M protein interacts with the RNPs to inhibit viral transcription and initiate viral assembly[30]. The M protein interacts with nucleophosmin (Npm)[21]. Npm plays an important role in the transport of ribosome components for ribosome assembly. It is indicated that Npm is recruited by the M protein for the correct assembly of the virion. The recent work of Cao *et al.*[36], suggests an interaction between the M protein and the receptor of activated C kinase 1 (RACK1). RACK1 is an adaptor protein that negatively regulates IRF3 and interferon regulatory factor 7 (IRF7) expression to inhibit IFN signaling.

1.2.4 Surface Proteins

RSV has three surface proteins, attachment (G) glycoprotein, fusion (F) glycoprotein, and the small hydrophobic (SH) protein. They are located on the outside of the envelope structure of the virion[11].

The G protein is between 292 and 319 amino acids long, depending on the strain[37]. It is comprised of three parts, an intracellular cytoplasmic tail from amino acids 1 to 37, a transmembrane domain from 38 to 66, and the extracellular domain ending. The transmembrane domain contains a second translation initiation site at amino acid 48[37]. Cleavage of the remaining 18 amino acids results in a G protein only containing the extracellular domain ending. This truncated G protein (sG) is secreted. The sG form binds to neutralizing antibodies, diverting them from their target, the full membrane-bound form of the G protein on the envelope surface, thereby reducing antibody-mediated neutralization of the virus and acting as an antibody decoy[37]. The complete G protein is a single-pass type II integral

membrane protein[11]. There are between 4 and 5 N-linked glycans¹ and 30 to 40 O-linked glycans across the variable parts of the protein. The glycans make up around 60% of the molecular mass of a mature G protein. The glycosylation varies depending on the producing cell type[11][37]. The outer-capsid domain can be split into three further domains. Two highly glycosylated variable mucin-like domains and a central conserved domain (CCD). The CCD contains the C-X3-C chemokine (CX3C) motif[38][39] and a heparin-binding domain (HBD)[40]. The CX3C motif allows the G protein to bind to the CX3C chemokine receptor (CX3CR1). This binding allows the virus to infect primary human airway epithelial cells. In addition, the CX3C motif implies an influence on the host's immune response[41]. Some host cell surfaces also contain glycosaminoglycans, which the HBD can bind to[40]. The G protein is (depending on the strain) an essential or non-essential protein. Deletion can reduce the level of replication or can stop replication altogether[41]. Mutations of the G protein CCD, especially the CX3C motif, showed many regulatory functions of the binding to CX3CR1[37]. The motif seems to decrease IFN- γ levels and reduce the trafficking of CX3CR1+CD4 and CD8+ cells to the lungs. When removing the cysteine residues in the CCD, the cytotoxic T-cell response to the infection was decreased. The G protein regulates infected neonatal regulatory B lymphocytes (nBreg) to generate immunosuppressive interleukin-10 (IL-10)[41]. Furthermore, the occurrence of RSV-infected nBreg cells in the respiratory tract of newborns can be used as a predictor of acute bronchiolitis severity.

The F protein is a class I fusion protein with a length of 574 amino acids and is highly conserved between strains with only a 25 amino acid difference[8][11]. It is first translated into one precursor F0 protein, which is co-translationally modified by the addition of 5 or 6 glycans (depending on the strain)[11][42]. This monomer is inactive. To be activated, the peptide 27 (p27, is a 27 amino acid long peptide[43]) needs to be removed by a furin-like protease followed by linking up to a compact trimer². The F protein facilitates viral entry into the cell and fusion between infected and non-infected cells[11]. In the pre-fusion state, the F protein cannot fuse to the membrane of the host. To become fusion-competent the F protein is cleaved into two subunits, F1 and F2[7]. After cleavage the F protein undergoes a conformational change, presenting the fusion peptides, which were initially hidden in a cavity in the F0 state, outward and initiating fusion. When the fusion peptides are inserted into the membrane, the protein undergoes another conformational change bringing the virus closer to the membrane which then leads to the fusion of the membranes. Tayyari *et al.*[44] showed that the interaction between F and nucleolin is necessary for efficient infection.

The SH protein is a 64 or 65 amino acids small type II transmembrane protein[45]. The SH protein is a viroporin[46]. It contains one N-linked glycosylation site and in infected cells, it exists in a non-glycosylated, N-linked glycosylated and polylactosaminoglycan-modified form[47]. Antibodies against SH protein specifically recognize the glycosylated and non-glycosylated forms[45]. Recent findings suggest that the glycosylation of SH could be a signal for the trafficking of glyco-

¹Glycans are chains of repeating disaccharides that are part of the cellular glycocalyx of the host.

²This can also be the other way around.

sylated SH into virus particles[47]. It is also speculated when the virus particles contain matured SH proteins that these interact with the inflammasome after virus (re-)entry. This concerns the fusion of new cells when the virus moves to new cells in the host. The SH protein inhibits tumor necrosis factor-alpha (TNF- α) signaling and subsequently inhibits NF- κ B activity and apoptosis [45]. The presence of SH is not essential for virion assembly[47]. However, it is suggested that the absence of SH protein can lead to viral attenuation[46].

1.2.5 Viral Life Cycle

To infect a cell, RSV needs to enter the host. For this, the virus attaches to the cell membrane via the G (mainly) and the F proteins. This initiates the fusion process. The F protein fuses the host's membrane with the viral membrane through multiple conformational changes. After fusion, the RNP enters the cytoplasm. Replication happens in the cytoplasm. The viral RdRp complex is in charge of transcription and translation. The assembly of virions happens at the host's membrane[11]. F proteins recruit M proteins which initiates filament budding and transportation of the viral parts for assembly. This leads to an outward deformation of the membrane and subsequent dissociation from the plasma membrane[11].

1.3 Clinical Relevance and Medical Intervention

1.3.1 Antibodies

Palivizumab (MEDI493) is a humanized immune globulin G1 monoclonal antibody. It binds to the pre- and postfusion protein[48]. It is comprised of human amino acid sequences and 5% amino acid sequences of mouse origin[49]. Palivizumab prevents conformational changes of F as well as cell-to-cell fusion of already infected cells. It is administered once a month as prophylaxis. The most frequently reported adverse events were: pyrexia, irritability, injection site reaction, diarrhea, rash, increased aspartate aminotransferase, upper respiratory tract infection, abnormal liver function, increased alanine aminotransferase, vomiting, coughing, rhinitis, and anaphylaxis from first or re-administration. It was approved in the United States of America (USA) in 1998[50] and in Europe in 1999[48].

Nirsevimab (MEDI8897) is a recombinant human immune globulin G1 kappa monoclonal antibody used as a prophylaxis[51]. It binds to the prefusion F0 protein. The half-life of nirsevimab was prolonged through a modified Fc region. This allows prevention with only one dose per RSV season. Its neutralizing effect is greatly enhanced compared to palivizumab[52]. Adverse events during the clinical trial occurred in very few infants and only one severe event was attributed to nirsevimab[53]. These events include local pain or swelling of the injection site, discomfort and pyrexia. The more severe adverse event was a rash 6 days after injection, which disappeared after 20 days without treatment. It received approval

in the European Union (EU) in November 3rd, 2022[54], in the United Kingdom (UK) in November 7th, 2022[54], and in the USA in July 17th, 2023[55].

Subtavumab (REGN2222) is a human monoclonal antibody. It targets the pre-fusion F protein. In microneutralization assays, it showed a 10-fold higher neutralization in RSV A and a 5-fold higher neutralization in RSV B[50]. Due to a mutation in the binding site of RSV B, subtavumab did not show any significant reduction in hospitalization rates or lower respiratory tract infections in its phase III clinical trial. Research and development for subtavumab was discontinued[41].

RespiGAM (RI-002³) is an intravenous immune globulin infusion preparation. It was used from the late 1990s to 2003. In 2003 it was discontinued and replaced by palivizumab, which had a higher potency, easier administration, less interference with other vaccines, and fewer adverse events[56].

1.3.2 Antiviral Drugs - Ribavirin

Ribavirin is a Food & Drug Administration (FDA) approved antiviral treatment against RSV[56]. It inhibits replication of RNA and DNA viruses directly by interfering with RNA capping, inhibition of polymerase, and lethal mutagenesis. It also inhibits indirectly by suppressing inosine monophosphate dehydrogenase and immunomodulatory effects. It is recommended to administer ribavirin as aerosolization for patients on mechanical ventilation and in oral form for pediatric lung transplant patients.

1.3.3 Vaccines

The development of vaccines for RSV was rapidly halted when in the 1960s an unsuccessful trial with a formalin-inactivated RSV vaccine resulted in a possible higher number of severe respiratory diseases and two deaths[57]. This hindered the development of new vaccines for years to come.

RSVPreF3 OA approved as Arexvy by GSK is the first vaccine against RSV to be accepted by the FDA and the European Medicines Agency (EMA) for adults in 2022[57]. A recombinant stabilized prefusion F protein is used as an antigen. The clinical trial for pregnant patients was halted after the rate of preterm births increased by about 2% compared to placebo.

RSVpreF approved as Abrysvo by Pfizer was approved in by the FDA and the EMA for adults in 2023[57]. It is still part of an ongoing phase III trial. Interim data of the study points to a possible safety issue due to the possibility of contracting Guillain Barré syndrome (GBS) as a serious adverse event. Other studies did not show any indications of GBS. It is the first vaccine against RSV to be approved for

³ID from DrugBank accessed 18.01.2024.

pregnant patients in the last trimester. With this vaccine, a non-significant rise in preterm rates was observed, so a causality between vaccination and preterm birth cannot be ruled out.

1.3.4 Host Factors

Most of the options for medical intervention mentioned above are prophylactic. The only option for intervention in a case of infection lies in the use of ribavirin. However, ribavirin is prone to resistance due to mutagenesis of the viral protein it acts upon. Thus, there is an urgent need for new therapeutic drugs. An alternative route in medical intervention would be not to interact with the virus directly. The proliferation of viruses like RSV depends on some host cell machinery, so-called host dependency factors (HDF). Their counterparts are host restriction factors (HRF). HDFs are factors (mostly proteins) that enhance viral growth. Usually, host factors (HF) can be determined through a knockdown or knockout screen (or loss of function screen), as in the screens from Hoffmann *et al.*[58] and Lingemann *et al.*[59]. In addition, Feng *et al.*[60] presented a comprehensive list of the HFs described in the literature.

Host Dependency Factors

Feng *et al.* write that HDFs include receptors binding to the F or G proteins of RSV, but also other proteins like pumps and regulators among others. The HDFs can be broadly grouped by their function. Insuline-like growth factor 1 receptor (IGF1R), Importin β 1, Cofilin 1 (Cof1), Exportin 1 (Crm1), ATPase NA $^+$ /K $^+$ transporting subunit alpha 1 (ATP1A1), and Ras-related protein Rab-5A (Rab5a) are factors related to a transporting function. The already mentioned CX3CR1 and Nucleolin are taking on functions for attachment and entry of the virus, together with Eidermal growth factor receptor (EGFR), Heparan sulfate proteoglycans (HSPGs), and Intercellular adhesion molecule 1 (ICAM-1). For the assembly of the virion, Actin-related protein 2 (ARP2), ATP binding cassette E 1 (ABCE1), and Caveolin 2 (Cav2) are necessary. Some proteins related to the immune response are known to be also essential for the virus, including Toll-like receptor 4 (TLR4), Monocyte differentiation antigen CD 14 (CD14), and MAP kinases (JNKs).

Host Restriction Factors

HRFs mostly include proteins inhibiting viral replication, transcription, entry, or inducing the immune response. Among these are the Ribosomal protein L13a, Interferon-induced transmembrane protein (IFITM), Interferon-induced protein 44 (IFI44), and Interferon induced protein 44-like (IFI44L), and Apolipoprotein B mRNA Editing Enzyme Catalytic Subunit 3G (A3G)[60]. There is also the CXC Chemokine Ligand 4 (CXCL4) protein, which blocks RSV from binding to HSPGs, and Guanylate Binding Protein 5 (GBP5), promoting the secretion of SH protein from cells.

1.4 Gene Regulatory Networks

During and after entry into the cell, RSV induces changes in the infected host cell. As mentioned above, the NS1 protein regulates host gene expression. Other proteins of the virus inhibit different signals. One method used to gain insights into these regulation processes is the analysis of gene regulatory networks (GRNs). GRNs are networks formed by the interactions between DNA, RNA, and/or proteins[61]. These networks can be described as directed graphs. The nodes of these graphs are the aforementioned molecules. The interactions between them are directed edges. Data for inferring these networks can come from different sources, gene expression data (from RNA sequencing or microarrays), protein-protein interaction data, protein-DNA interaction data, and others. The precise inference and analysis of such networks is not an easy-to-solve task. To tackle this task, many tools are available. Exemplary, ISMARA and MIPRIP are described. However, there are many more, such as ARACNE[62], GENIE3[63], or SCENIC[64].

The web-based tool ISMARA (Integrated System for Motif Activity Response Analysis) processes raw gene expression (microarray or RNA-seq) or chromatin state data (ATAC-seq or ChIP-seq) [65]. ISMARA utilizes a curated genomic database to predict 200 transcription factor (TF) binding sites (TFBSs) in promoters using the comparative genomic Bayesian methods. The provided expression or chromatin data is used to calculate signals for each promoter. Predictions and measured signals are used to solve the linear motif activity response analysis (MARA) model. The MARA model was proposed by the FANTOM Consortium[66]. The linear model infers the activities for each motif. Outputs include motif activity profiles, significance-sorted target promoters, interaction networks from the STRING database, and enriched Gene Ontology categories, along with predicted regulatory interactions for each motif.

The R-based tool MIPRIP (Mixed Integer linear Programming based Regulatory Interaction Predictor) [67] predicts with the help of a regression model the TFs responsible for the gene expression of a given target gene. It takes a gene regulatory network, gene expression data, a target gene and the maximal amount of TFs that should be used to predict the gene expression of the target gene. The regulatory network is composed of up to seven different data sources. Using the network and the gene expression data, the activity of the TFs in each sample is computed. After successful optimization of the regression model, the predicted counts of the target gene as well as the TFs used for the prediction are returned. MIPRIP was already used in multiple scenarios, including the prediction of telomerase regulators in yeast[68], in the modelling of TERT regulation in a pan-cancer data set[67] and in the prediction of factors for reprogramming M2-like macrophages into M1-like macrophages[69]. MIPRIP is in version 2.0 and allows for the prediction of the proteins that regulate/control the target gene in the given expression profiles. It also allows for testing in one, two, or multiple conditions. Through manipulation of the activities of the TFs it is possible to simulate knockouts (in some cases even knockins) of TFs.

1.5 Objectives

One part of the main target demographics, i.e. infants and young children, is not eligible for any real immunization. In case of an infection, only the treatment with ribavirin is left. Adverse effects of ribavirin as mentioned above are plentyfull and it is regularly discussed, if the benefits of medication outweigh the risks. That is why an alternative way of medical interventions for RSV infections is needed. The goal of this thesis was to utilize expression data to analyze the changes in expression during a RSV infection. From the identified changes in expression the underlying processes should be identified and grouped into broader categories. Then overlaps with HDFs and HRFs should lead to a decision on specific categories to further investigate. For these categories regression models should be constructed with the help of gene regulatory networks to identify TFs, that induce the reprogramming observed in the gene expression changes during RSV infection. These TFs should then be copared to literature to identify their potential role in the infection.

2. Materials & Methods

2.1 Datasets

2.1.1 Gene Expression Data

To analyse the gene regulation responsible for the changes during a RSV infection gene expression data was used. The data sets used are publicly available in the Gene Expression Omnibus (GEO) database [70] and briefly described in the following.

Rajan *et al.*[71] provide an RNA-seq data set including counts for more than 60,000 genes and 5 different conditions. They used A549 and HEp-2 cell lines. The A549 cell line is a cell line that was originally isolated from a 58-year-old Caucasian male with lung cancer in 1972[72]. It is an epithelial-like lung cancer cell line. The HEp-2 cell line is a cell line derived from the HeLa cell line through cross-contamination[73]. HeLa cells were originally isolated in 1951 from an African American 31-year-old female with adenocarcinoma. HEp-2 cells are classified as epithelial cells, while HeLa cells are cervix epithelial cells[74]. The cells were cultured in a minimum essential medium (MEM) from Corning (10-010-CM), 10% fetal bovine serum (FBS), 1% antibiotic/antimycotic mixture (10000 U/mL penicillin/streptomycin/15 µg/mL amphotericin B), 1% L-glutamine, and maintained in 5% CO₂ at 36°C[71]. When the medium was nearly covered with cells they were infected with one of four strains of RSV at a multiplicity of infection (MOI) of 0.01 for 1.5h. The strains were RSV A Tracy (GA1), RSV B 18537 (GB1), RSV A Ontario (ON) and RSV B Buenos Aires (BA). After 1.5h the inoculum was removed and the cells were washed with phosphate buffered saline (PBS). 2% of FBS/MEM was added and the cells were incubated for 24h, 48h, 72h, and 96h post-infection. At each time point, samples were taken, the cells were lysed and RNA was extracted using Qiagens Mini Viral RNA kit. The mRNA was enriched from total RNA using oligonucleotide beads. The GEO entry (GEO accession: GSE196385) for the experiments stated that an Illumina NovaSeq 6000 machine was used for sequencing. Rajan *et al.* carried out quality control of the reads, aligned them to the human genome hg38 (GRCh38) and generated the count matrices counting non-overlapping uniquely mapped features. The mean read count per sample was at around 47 million. The data is available with the GEO accession GSE196385.

Xu *et al.*[75] used human small airway epithelial cells (hSAEC). They immortalized them using the method described by Ramirez *et al.* in 2004[76] by ectopically expressing human telomerase and CDK4 to generate a continuously replicating hu-

man primary cell. hSAECs grew in a small airway epithelial cell growth medium by Lonza in a 5% CO₂ atmosphere. RSV long strain was grown in HEp-2 cells pre-infection before infecting the primary cells with an MOI of 1 for 24 hours. Total RNA was isolated using Qiagens RNeasy kit. Illumina HiSeq 2000 paired-end sequencing was used for RNA sequencing. The quality controlled reads were aligned to the human genome hg19 and mapped paired-end reads for genes and transcripts were counted. The mean read count per sample was at around 849.700. Data is available with the GEO accession GSE161846.

2.1.2 Host Factor Screening Data

To help in the process of deciding on which cellular processes the analysis should focus, HDFs and HRFs were identified. Those were retrieved from two knockdown screens.

Hoffmann et al.[58] produced a gene trap virus in 293T cells with a lentivirus. 293T cells are transformed cells derived from human embryonic kidney cells (HEK293)[77]. They infected Hap1 cells for 3 consecutive days with the gene trap retrovirus so that the cells could be mutagenized by the lentivirus and then passaged for 5 days. Hap1 cells are white blood cells with leukemia from a 40-year-old male[78]. After the 5 days passaging the cells were frozen. 1*10⁸ cells of the mutagenized cells were exposed to RSV-GFP at an MOI of 1. The green fluorescing RSV variant was based on the RSV strain A2. To identify the insertion sites in RSV-selected cells, DNA was isolated, amplified by PCR, and sequenced with an Illumina HiSeq 2000. The reads were mapped to hg19. Mutations caused by the insertions were counted per individual gene. Significant genes were determined using a binomial test and multiple testing correction was done calculating the false discovery rate (FDR) using the method by Benjamini-Hochberg[79].

Lingemann et al.[59] used A549 cells in an F12-K media supplemented with 10% FBS and L-glutamine. The cells were incubated at 37°C and 5%CO₂. They were engineered to express DsRed as a viability marker. An siRNA based knockdown screen was performed with the siRNA library Ambion Silencer Select Human Genome version 4 from Fischer Scientific, which targets around 21,500 human genes. The cells were infected by RSV-GFP at an MOI of 1 for 48h. In this study, the green fluorescing RSV variant was also based on the RSV strain A2. Images were taken through an ImageXpress Micro XL automated microscope. For normalization, a negative control with the Ambion Silencer Negative Control #2 was done. As a positive control, an siRNA targeting GFP resulted in a near-complete elimination of positive GFP cells. To counter outliers, the median absolute deviation (MAD) based z-score was computed for each siRNA, which was used to adjust for seed-based off-target effects. In this work all z-scores below -0.2 indicated HDFs, while all z-scores above 0.2 indicated HRFs. We considered Z-scores between -0.2 and 0.2 as too small of a change to have an effect on the ability of the virus to strengthen or weaken its interactions with the host cells.

2.2 Differential Gene Expression Analysis

2.2.1 Data Exploration

For the whole of the analysis, the R language 4.3.2[80] and RStudio Build 386[81] were used. For all plots, if not stated otherwise, the package ggplot2[82] was used. To explore the data a hierarchical clustering was done with pvclust[83]. The clustering method was set to average which corresponds to the unweighted pair group method with arithmetic mean (UPGMA). As a distance measure the Pearson's correlation coefficient was used. To achieve a uniform look throughout this work the tidygraph[84] and the dendextender[85] packages were used. They allow the extraction of the dendrogram structure from pvclust and to modify it in a ggplot graph. Additionally, a (PCA) was done. For the A549 and HEp-2 data sets the mock samples were included multiple times. A correlation test between these shows that they are the same samples. Thus they were removed, only one sample per time point remained for further analysis. Some gene IDs were represented multiple times. In that case, the mapping with the highest variance was kept. Additionally, the genes with a low count number were filtered out. All genes were kept that had an equal or higher count per million (CPM) than 1 for at least 8 samples for the A549 and HEp-2 data sets and 1 CPM for at least 4 samples for the hSAEC data set. Filtering in this way is recommended by the edgeR user guide[86]. CPM accounts for different library sizes and 8 (or 4) samples are recommended because the smallest condition size for the data set is 8. This resulted in 12,601 A549 genes, 11,802 HEp-2 genes, and 11,549 hSAEC genes for further analyses.

2.2.2 Normalization

Over the last decade, many normalization methods have been proposed. In comparisons two of them were suggested repeatedly, the median of ratios (TRM)[87] method and the trimmed mean of M-values (TMM)[88] method [89][90][91][92]. Both of these methods assume that most of the genes are not differentially expressed and those that are, are about equally distributed between up- and downregulated expression[91]. This assumption is however not always met and not often tested[91]. Kadota *et al.*[93] presented a method to use TMM without prior knowledge if the two previously mentioned assumptions are met. Their workflow first excludes all possible differentially expressed genes (DEGs) and then uses TMM on the remaining to calculate the normalization factors. Then these factors normalize the whole data set. The tag count comparison (TCC) R package[94] implements this method and expands it for use with TRM and TMM. Sun *et al.*[94] recommend such an iterative TMM mode followed by an analysis with the edgeR[95] package and three iterations of normalization and analysis for data sets with replicates. In this work, the recommendations of Sun *et al.*[94] were followed with the gene expression data that was filtered by CPM as described in the step before. TMM was used to make the first normalization step, then edgeR determined DEGs to be removed, and then TMM renormalized the data. This was repeated three times, resulting in normalized

gene expression data sets for A549, HEp-2 and hSAEC.

2.2.3 Gene Expression Analysis

Differential Gene Expression (DGE) analysis was performed with the edgeR[95] package. DGEList objects were created from raw counts, the library size (sum of all counts in a sample), the TCC calculated normalization factors, the groups as described in 3.1, and the gene symbols. As recommended by the edgeR users' guide[86] the common, trended, and tagwise dispersion was estimated. edgeR employs the Cox-Reid profile-adjusted likelihood (CR) method to estimate dispersions, which involves fitting generalized linear models (GLMs) using a designated design matrix based on the groups being compared. The design matrix is constructed to represent the specific contrasts between groups. Taking the estimated dispersions into account, a GLM is individually fitted for each gene (in the edgeR functions and in the user's guide "tag" is used). Since the correct distribution of the genes is unknown, a quasi-likelihood estimator is utilized in conjunction with the negative binomial dispersion to estimate the scaling factor for subsequent analysis of variance (ANOVA) testing. In this analysis, a quasi-likelihood F-test is employed to conduct an ANOVA. The quasi-likelihood F-test is applied to the fitted models, resulting in a list-object including a tag table with the genes, ranked by log2-fold change (LFC). The computed p-values for the genes are then FDR corrected with the Benjamini-Hochberg method. To determine the DEGs for a data set, the tag-table was filtered for tags with a p-value below 0.05 and an absolute LFC of at least 1. Upregulated genes have a LFC larger or equal to 1 and downregulated genes have an LFC smaller or equal to -1. Parameter values not based on the data sets or groups were set to the recommendations of the user guide. DEGs were determined for A549 and HEp-2 comparing the late time points to the early ones and hSAEC at 16h, and hSAEC at 24h, compared to uninfected controls at 0h, respectively.

2.3 Gene Set Enrichment Analysis Pipeline

2.3.1 Gene Set Enrichment Analysis

For each of the four sets of DEGs (A549, HEp-2, hSAEC 16h and 24h post infection), a separate gene set enrichment analysis (GSEA) for up- and downregulated genes, respectively, was performed with the gprofiler2 package[96] for R which is a wrapper for the web-based tool g:Profiler[97]. The last version of g:Profiler used for analysis was *e111_eg58_p18_30541362* with a database update on the 25th of January 2024¹. g:Profiler uses the databases from Ensembl[98] and Ensembl Genomes[99] to update its own database. From the various data sources (e.g. Gene Ontology Biological Process (GO:BP), Gene Ontology Molecular Function (GO:MF), Gene Ontology

¹The database used was downloaded as a Gene Matrix Transposed (GMT) file and is available in the supplementary files of the publication.

Cellular Component (GO:CC), KEGG and Reactome) g:Profiler provides the sub-ontology GO:BP from Gene Ontology (GO)[100][101] and was used to determine the enriched pathways. g:Profiler performs a cumulative hypergeometric test to calculate a value of significance which Kolberg *et al.* call an enrichment score for each gene set from the source (GO:BP in this case). The enrichment score is then corrected for multiple testing by the default method of g:SCS (Set Counts and Sizes). This is a method specifically developed to estimate thresholds in complex and structured data[102]. The resulting gene sets with a corrected p-value less than 0.05 were then filtered by their set size. All gene sets with a set size between 30 and 1000 genes were kept. In enrichment analysis, the size of a gene set may not always correspond to the intersection size. Gene set size represents the total number of genes associated with a gene set, while the intersection size denotes the overlap between the list of selected genes and genes in the gene set.

A term in the GO represents a gene set. The genes mapped to a term are all related to a function (GO:MF), a compartment (GO:CC) or a process (GO:BP). The gene sets have relations between them. For GO in particular these relations are described as *is a*, *part of*, *has part*, and *regulates*. Gene sets in GO and their relations can be represented in a directed acyclic graph. In this graph, gene sets are the vertices (or nodes), and the relations represent the edges. A node can have multiple parent nodes and multiple child nodes. Given the nature of these relations, gene sets closer to the root are broader and more overarching than those closer to the leaves.

2.3.2 Redundancy Reduction

Several enriched gene sets may have many overlapping genes. To reduce those gene sets that are similar, customized scripts were provided by Amol Kolte[103]. The overlap between two gene sets A and B was quantified by the Jaccard similarity coefficients[104]:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

For the different sets to reduce redundancy, different thresholds were used. For all the upregulated gene sets a threshold of 0.3 was used. For the downregulated gene sets different thresholds were used. For the A549 gene set a threshold of 0.6 was used, for the HEp-2 set 0.5 was used, 0.3 was used for the hSAEC gene sets at 16 hours, and for the hSAEC gene sets at 24 hours a threshold of 0.5 was also used². Using these thresholds, for each of the gene sets an undirected graph $G = (X, E)$ was created. X represents the gene sets as vertices, while E represents pairs of gene sets with $J(A, B)$ being larger or equal to the corresponding threshold. Formulating

²These thresholds were chosen by reducing the redundancy for all thresholds between 0.1 and 1 (in steps of 0.1). Due to elimination of gene sets, genes could eventually be lost. The threshold which would guarantee the remainder of 95% of all genes was used for the actual redundancy reduction.

a mixed-integer linear programming problem with the objective:

$$\text{Max} \sum_{i=1}^n w_i X_i \quad (2)$$

where

$$w_i = \frac{1 - \log_{10}(P)}{100} \quad (3)$$

such that for each edge at most one of the two vertices can be selected:

$$X_i + X_j \leq 1 \text{ for all } \{i, j\} \in E \quad (4)$$

$$X_i \in \{0, 1\} \text{ for } 1 \leq i \leq n \quad (5)$$

w_i is the weight of a gene set, determined by the p-value P of the respective gene set. n is the number of gene sets. This mixed integer programming (MIP) problem was solved using Gurobi™ (version 10.0) and the R interface in version 10.0-3 [105]. This MIP is equivalent to the graph's maximum weight independent set problem. This ensures that for every possible pair A and B, their Jaccard similarity coefficient is smaller than or equal to the corresponding threshold. The reduced gene sets were used to create one combined custom database for g:Profiler. This was done by using the complete database for the original GSEA and reducing it to only these terms that were left after the reduction. The resulting custom database was then uploaded to the g:Profiler server and a new GSEA was done to select the reduced gene sets.

2.3.3 Categorization of Enriched Pathways

To get a better overview, the gene sets were then classified into one of 18 categories. These categories represent higher-order systems and processes. The 18 categories are *cell adhesion*, *cell cycle*, *cell death*, *cellular component organization*, *ER/endosome/lysosome related process*, *gene regulation*, *immune system process*, *metabolic process*, *morphogenesis/development*, *multicellular organismal process*, *neural related process*, *other*, *protein modification/signaling*, *receptor related process*, *response to stimulus*, *stress response*, *transport*, and *viral related process*. A machine learning model was provided by Jonas Marx (Friedrich-Schiller-University, unpublished), which was trained to classify GO terms from g:Profiler into the 18 categories. The model is based on natural language processing (NLP) and random forest (RF). For this, a database of all the GO terms was set up. Columns of the database include term_name, parents, children, ancestors, definition, and genes. A column with groups/categories was added and for 3,475 gene sets (terms) manual annotations were added. The model was trained to infer the other 19,240 GO terms available in g:Profiler (however, through filtering much less are relevant for this work). The parents, children, ancestors, and definition columns were collapsed into one text_summary column. The text_summary column is then transferred to a document-term matrix. In this matrix, a row corresponds to an entry in the text_summary column of the database, the columns are all unique words that are present in the whole of the text summaries. An entry $M[i, j]$ in this matrix is filled

with the frequency of word i in text summary of gene set j . The document-term matrix was then reduced by removing columns with a sparse percentage³ of 99.5%, reducing the number of columns from 52,668 to 12,263. An RF model is then trained on this reduced document-term matrix and validated by 10x cross-validation. The model has an accuracy of approximately 82.2%. Marx used the following R packages for the construction of the model: GSA[106], tm[107][108], SnowballC[109], caTools[110], randomForest[111], caret[112], GOfuncR[113], and ontologyIndex[114]. The classifications of the RF model were then manually checked for correctness and then used to map the GO terms to the corresponding categories. The results were visualized with the R packages ggplot2[82] and circlize[115]. For the pie charts without HF information (see Figures 10 and 11) the categories were scaled to the percentage of genes occurring in the category. The entirety of observed genes equals to 100%. If a gene occurs in n categories, its weight for a single category is $1/n$. This scaling will be called *gene weight*. Due to the bar plot around the pie chart with HF information the extended pie chart could not be scaled according to *gene weight*.

2.4 Prediction of Transcription Factors using a Gene Regulatory Network Model

To predict regulators for the genes of the selected (sub-)categories the R package MIPRIP[67] was used. MIPRIP stands for Mixed Integer linear Programming based Regulatory Interaction Predictor. In its simplest applicationuse case, MIPRIP takes as input expression data, a gene regulatory network (GRN) and a target gene (TG). For the gene expression data, the filtered but unnormalized data sets were used. The count tables were then vsn normalized with the vsn R package[116], this provided better results in the prediction of regulators. Poos *et al.*[67] provide a generic human regulatory network (gHRN). It was compiled from of 7 different sources: MetaCore™[117], the ChIP Enrichment Analysis (ChEA) databas[118], ChIP data from ENCODE[119] (<http://www.genome.gov/Encode/>), human ChIP-seq and ChIP-ChIP data from hmChIP[120], experimentally verified interactions from the Human Transcriptional Regulation Interactions database (HTRIdb)[121], ChIP-seq data from ChIPBase[122], and from known TF binding site motifs using the Total Binding Affinity (TBA) method[123]. An interaction between a TF t and a TG g was assumed if it was included in MetaCore™ labeled as direct (MC_{dir}) or listed in ENCODE (enc). It was also included in the gHRN if it was in at least two out of MetaCore™ (with the label indirect) (MC_{indir}), ChEA ($chea$), TBA with a score ≥ 1.5 (tba), or HTRI ($htri$). If an interaction was listed in ChIPBase ($chip$) and hmChIP (hm) it was also included. Since not all sources were considered

³This refers to the percentage of zero values in a column.

equally reliable an entry es_{tg} in the gHRN is defined as an edge score

$$\begin{aligned} es_{tg} := & 2 \cdot MC_{dir_{tg}} + 0.5 \cdot enc_{tg} \\ & + a_{tg} \cdot (MC_{indir_{tg}} + chea_{tg} + htri_{tg} + tba_{tg}) \\ & + 0.25 \cdot hm_{tg} \cdot chip_{tg} \end{aligned} \quad (6)$$

with

$$a_{tg} := \begin{cases} 1, & \text{if } MC_{indir_{tg}} + chea_{tg} + htri_{tg} + tba_{tg} \geq 2 \\ 0, & \text{else} \end{cases} \quad (7)$$

and

$$MC_{dir_{tg}}, enc_{tg}, MC_{indir_{tg}}, chea_{tg}, htri_{tg}, tba_{tg}, hm_{tg}, chip_{tg} \in \{0, 1\} \quad (8)$$

The network, represented by a matrix, includes 1,160 TFs (columns) and 31,915 TGs (rows). As TGs, the DEGs from the selected pathways were chosen. DEGs were used to predict the regulatory factors responsible for changes in gene expression. The network was reduced to the relevant genes and TFs. The genes in the network were reduced to the number of genes present in the data sets and the TFs were reduced to these TFs that regulate more than 20 of the TGs. Each data set was analysed independently with MIPRIPs single mode, with up to 10 parameters, 5 repetitions, and 3x cross-validation. MIPRIP predicts the relevant regulators by predicting the gene expression \tilde{g} of a gene g in sample k through a linear model, i.e.:

$$\tilde{g}_{g,k} = \beta_0 + \sum_{t=1}^T \beta_t \cdot es_{t,g} \cdot act_{tk} \quad (9)$$

The covariates $es_{t,g} \cdot act_{tk}$ represent the potential TFs, with β_t being the optimization parameter for TF t . T is the number of all regulators that have a non-zero entry in the gHRN for target g . β_0 is an additive offset. act_{tk} is a calculated activity value. This is determined for each TF t and each sample k as the cumulative effect of a regulator on its TGs i , normalized by the combined edge score of the targets, i.e.

$$act_{tk} = \frac{\sum_{i=1}^n es_{ti} \cdot g_{ik}}{\sum_{i=1}^n es_{ti}} \quad (10)$$

With g_{ik} representing the gene expression value for gene i in sample k . The objective function is

$$\min \sum_{k=1}^g |g_{g,k} - \tilde{g}_{g,k}| = \sum_{k=1}^l e_{g,k} \quad (11)$$

Using Mixed Integer Programming, MIPRIP optimizes the absolute error between the observed expression value of a gene g and its predicted value. Within a cross-validation scheme MIPRIP also computes the Pearson correlation for the observed and the predicted gene expression as an estimation of the performance. To determine significant TFs, p-values were computed. This was done by simulating random predictions 1,000 times for each TG. Each simulation step sampled the outcome of 150 models, due to the generation of 150 models per gene modelled by MIPRIP. The number of models is calculated by the number of cross-validations times the

number of repeats times the number of parameters. This culminates in $3 \cdot 5 \cdot 10 = 150$. Each sampling was done by drawing, without replacement, the number of TFs needed from the whole list of possible TFs. Thus 150,000 samples were taken. Then, the probability of occurrence was calculated from the simulations for each of the predicted TFs. A two-sided permutation test was then performed based on these probabilities. P-values were corrected for multiple testing with the method of Benjamini-Hochberg[79]. Regulators with an FDR-corrected p-value equal to or higher than 0.05 were removed. Each TF may have been predicted to regulate more than one TG. Thus, a TF represents a set of genes it regulates. Similar to the reduction of redundancy in gene sets as described in section 2.3.2, the regulators were reduced by the custom scripts provided by Amol Kolte[103]. The TF would be considered a gene set, the genes it regulates were considered the genes in the gene set. As seen in equation (3) the computed p-value was used for the calculation of the weight of a gene set. The reduction was done with a threshold of 0.1 for Jaccard's coefficient, keeping approximately 90% of genes.

3. Results

3.1 Data Exploration Leads to the Removal of Duplicate Samples

The data set from Rajan et al.[71] was reduced from 60,738 total genes to only the protein-coding resulting in 19,988 genes left. The data was split into two sets. One set for the A549 cell line data (A549) and one for the HEp-2 cell line data (HEp-2). The data from Xu et al.[75] provided only protein-coding genes. Only the wild-type samples were kept in the analysis. Gene knock-out samples were removed. The logarithmized counts of the samples of each of the 3 data sets (with a pseudo-count added) were first clustered with pvclust to get an overview of the data and the samples. Figure 4A shows the clustering for the A549 data set. It can be seen that the first and most striking split of the data appears between the mock (control), the 24-hour, and the 48-hour samples as the early condition and the 72-hour and 96-hour samples as the late condition. For the HEp-2 data in Figure 4B, the groups are separated similarly. For the hSAEC data set Figure 4C shows that all 3 time points do split very distinct from each other. The control or first-time point (0 hours) is the first to split from the rest. Then the other two (16 hours and 24 hours) split. Additionally, a PCA was done for each of the data sets. When plotting the first principal component against the second it confirmed the findings of the hierarchical clustering. In the data from Rajan *et al.*, shown in Figures 6A, and B, the early and late conditions separated quite well. Figure 5C shows a clear separation for all time points in the data from Xu *et al.*. For further analysis, the mock samples of the A549 and the HEp-2 data sets were reduced to one replicate at each time point. They were also grouped in an early and late condition/group. The early group consists of the two mock samples taken 24 hours and 48 hours after starting the experiment, as well as the 4 samples taken 24 and 48 hours post infection. Leaving the early group with a size of 10 samples. The late group was then left with the 4 samples taken 72 and 96 hours post infection. Resulting in 8 samples for the late group. Since the hSAEC groups clustered well and the PCA also showed an excellent split between time points, these were chosen as groups. This included a group at 0 hours post infection, here used as a control group, one group 16 hours post infection and one 24 hours post infection. Recapitulating, the A549 and the HEp-2 data sets were grouped in an early and a late group. The early group is comprised of the mock samples, the samples 24 hours post infection and the samples 48 hours post

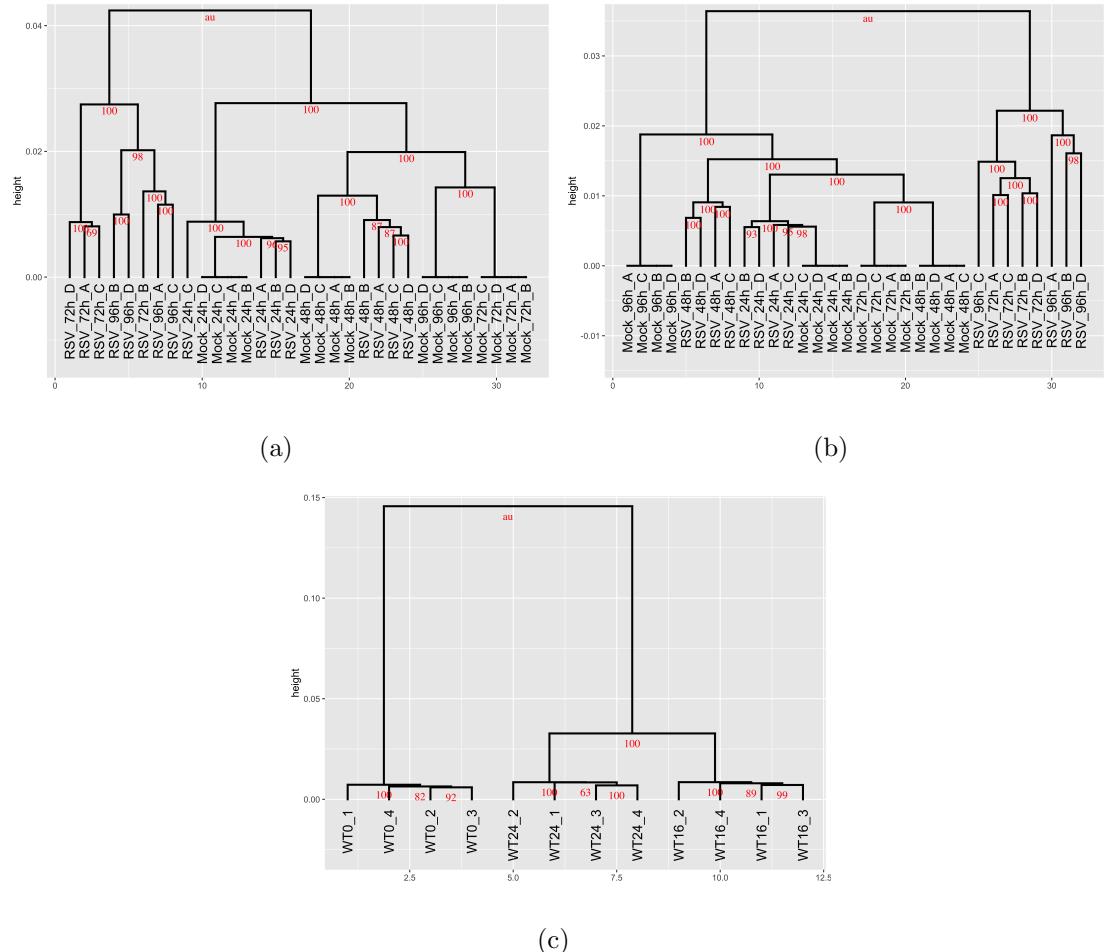


Figure 4: Shows the hierarchical clustering with UPGMA of the 3 data sets. The red text is the approximately unbiased p-value (au). Figures 1- 3 show all 3 Figures in full size.

(a) Shows the A549 data set. The most striking split is the first which splits the data set in two conditions, early (mock, 24h, and 48h) and late (72h and 96h). **(b)** Depicts the HEp-2 data set. It is also immediately split into two groups, early and late. **(c)** The hSAEC clustering shows a good split between all 3 time points. It is also noticeable that the earliest time point branches out first.

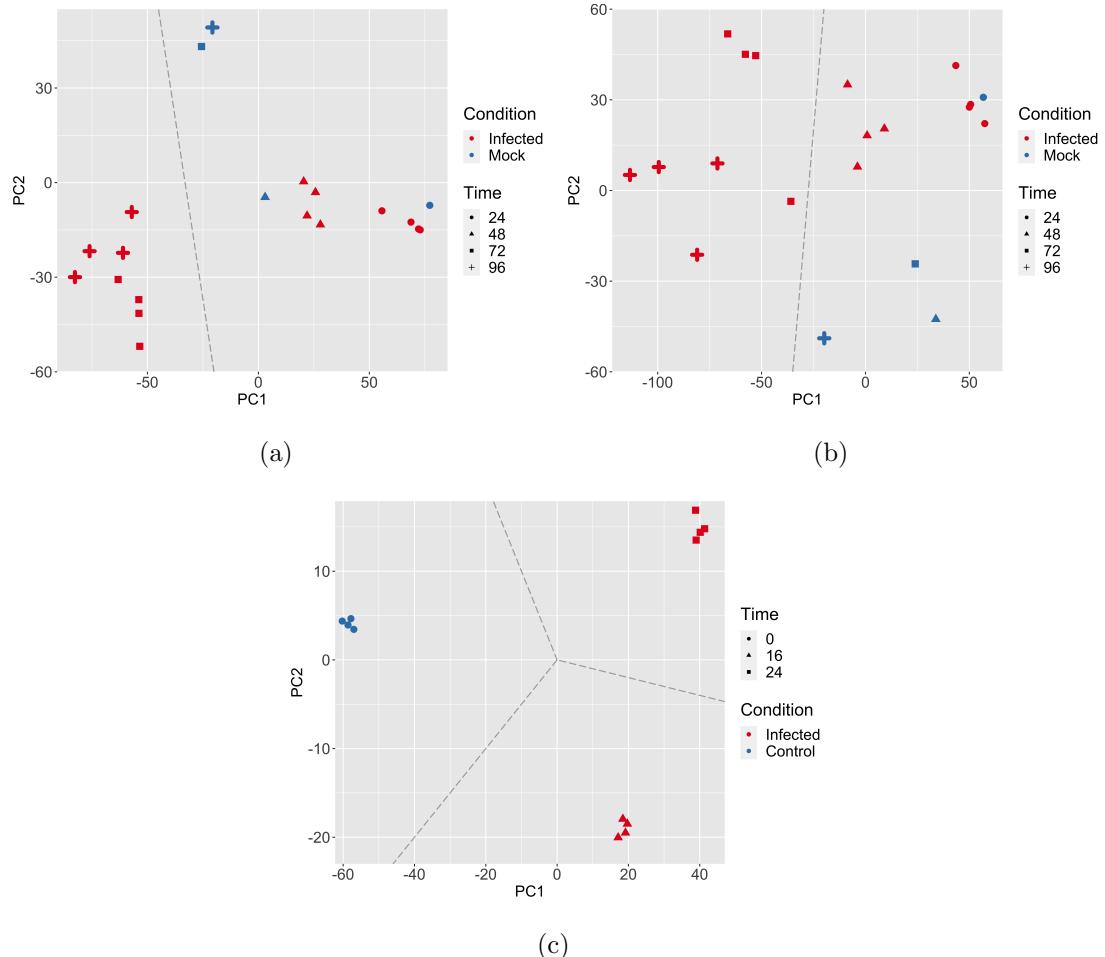


Figure 5: PC1 against PC2 for the 3 data sets. The dashed lines show the split between groups. Red data points represent infected conditions while blue represents the control conditions.

(a) and (b) show the plots for the A549 and HEp-2 data, respectively. Squares and crosses represent the time points 72 hours and 96 hours post-infection. Circles and triangles show the time points 24 hours and 48 hours. (c) shows a separation of all three time points while infected groups are closer together than the control group. Circles represent 0 hours, triangles 16 hours, and squares 24 hours post-infection.

infection. This group will be called the “control” group. The late group consists of the samples taken 72 hours and 96 hours post infection. The late group will be called “infected” group. The hSAEC data set was split into three groups. The “control” group including all of the samples taken at 0 hours post infection, the “early infected” group with all the samples taken 16 hours post infection and the “late infected” group including all the samples taken at 24 hours post infection.

3.2 Gene Expression Shows Upregulation of Host Restriction Factors

The normalized data was used to determine the DEGs. This was done with the three data sets for four DGE analysis experiments: A549 infected versus control, HEp-2 infected versus control, hSAEC early infected versus control and hSAEC late infected versus control. A gene is determined as upregulated if the change ($\log FC$) between a gene in the infected group and in the control group is larger than a threshold ($\log FC > 1$). A gene is downregulated if this change is lower than a given threshold ($\log FC < -1$). The results are visualized in so-called volcano plots. The $\log FC$ is plotted against the $-\log_{10}$ of the FDR-corrected p-value of the gene. This is shown exemplary in Figure 6 for the hSAEC late infected vs control experiment. The other three visualizations are included in the Supplemental Materials (4, 5, 6). DGE analysis determined 1,720 upregulated and 1,676 downregulated genes for the A549 data set. 1,608 gene were upregulated and 936 genes downregulated in the HEp-2 data set. In the hSAEC data set, we observed 1,651 upregulated and 1,386 downregulated genes when comparing the early infected versus control and 1,960 up- and 2,258 downregulated genes comparing late infected against control. Table 1 presents the up- and downregulated HDFs and HRFs listed by Feng *et al.*[60] in the analysed gene expression data.

Table 1: Up- and downregulated HDFs and HRFs in gene expression data.

Cell type	Up HDF	Down HDF	Up HRF	Down HRF
A549	ICAM1		IFI44, GBP5, EGFR, IFITM3, IFI44L	
HEp-2	ICAM1		IFI44, GBP5, EGF	
hSAEC early	ICAM1	ABCE1, NCL	IFI44, GBP5, XPO1 APOBEC3G, IFITM2	
hSAEC late	ICAM1	ABCE1, NCL	IFI44, GBP5, APOBEC3G, IFITM2	

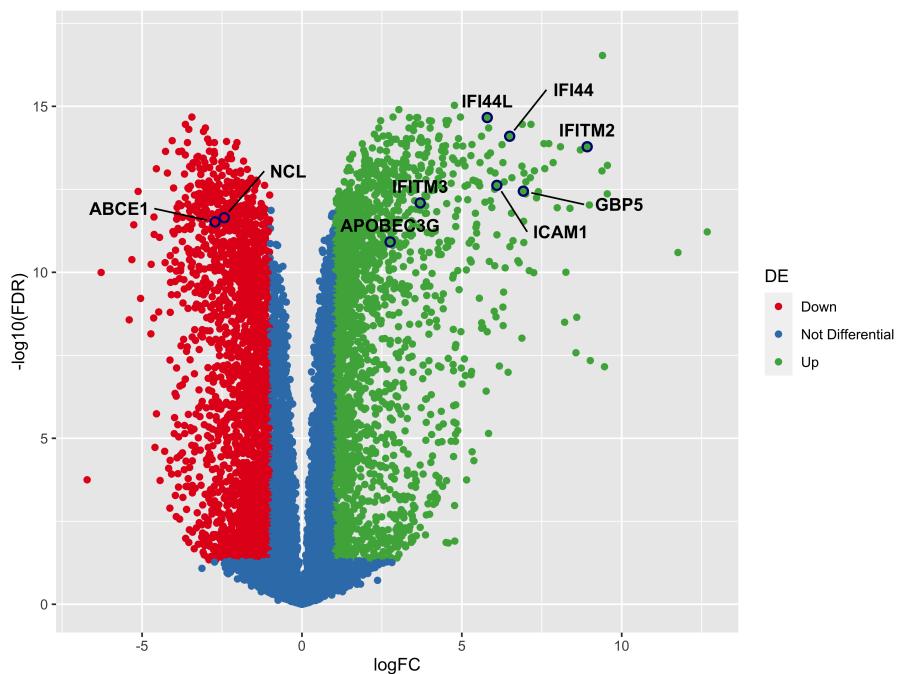


Figure 6: Volcano plot of the hSAEC late infected vs control groups. Red dots show the differentially downregulated genes, and green shows the differentially upregulated genes. Most HRFs are upregulated, while two HDFs are downregulated. Others are not differentially expressed (blue).

Table 2: Remaining genes after reduction of upregulated gene sets.

Jaccard Coef.	% A549	% HEp-2	% hSAEC 16h	% hSAEC 24h
0.1	76.16	80.46	80.11	78.2
0.2	93.4	93.24	92.23	90.23
0.3	96.35	96.12	97.05	96.53
0.4	98.21	98.1	98.28	98.30
0.5	98.91	98.35	99.1	99.08
0.6	99.15	98.76	99.26	99.22
0.7	99.38	99.67	99.18	99.5
0.8	99.61	99.67	99.59	99.58
0.9	99.61	99.92	100.0	99.93
1.0	100.0	100.0	100.0	100.0

3.3 Gene Set Enrichment Analysis Shows Upregulation of Genes Involved in Morphogenesis and Immune Response

Commonalities of data sets in function or of processes are very difficult to compare on the gene level, due to the sheer number of genes to compare. Gene sets give far better overview of what happens in the cells. The GSEA with g:Profiler[102] yielded 562 gene sets for the A549 upregulated genes and 217 gene sets for the A549 downregulated genes. The HEp-2 DEGs presented 580 upregulated gene sets and 142 downregulated gene sets. The hSAEC DEGs resulted in 584 and 616 upregulated and 176 and 265 downregulated gene sets for the early infected and late infected groups respectively. However, the nearly 600 upregulated and about 250 downregulated terms per GSEA are still quite unwieldy, thus a further reduction of gene sets is needed without loss in information given through the presented genes. To determine which gene sets to remove the Jaccard similarity coefficient was used.

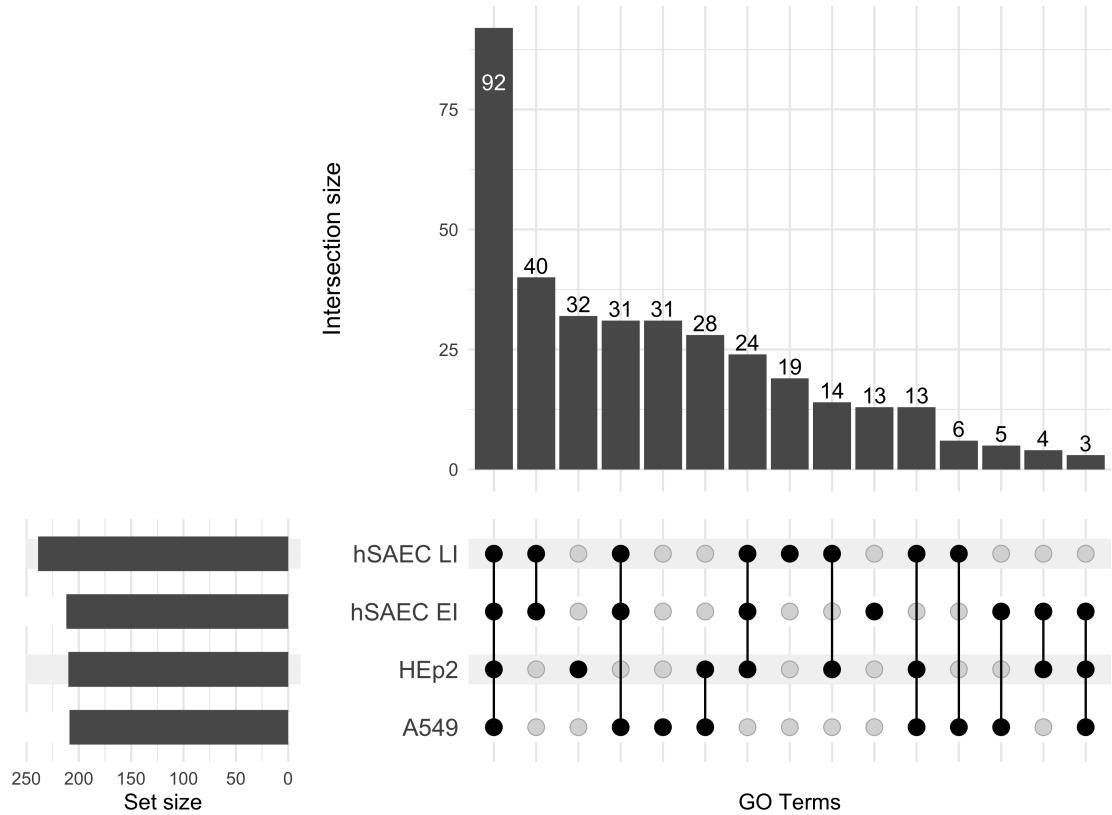


Figure 7: Upset plot of the upregulated gene sets. 150 gene sets are in at least 3 out of 4 groups. LI: *late infected* EI: *early infected*

For example, in the downregulated gene sets of A549 two GO terms have the same observed genes mapped to them. These two are the GO terms mitochondrial ATP synthesis coupled electron transport (GO:0042775) and ATP synthesis coupled electron transport (GO:0042773) (overlap shown in Figure 8). With the redundancy reduction, one of these terms is eliminated, without losing the information about the contribution of genes towards the processes inside of the data set. The results of the overlap-dependent filtering are shown in Tables 2 and 3 (plots produced as an overview of the percentage of remaining genes and of the number of remaining gene sets can be seen in the Figures 7-14). After the actual redundancy reduction a second GSEA was done (details are described in 2.3.2). From the 562 gene sets for the A549 upregulated genes, 209 upregulated remained, while from the 217 gene sets for the A549 downregulated genes, 117

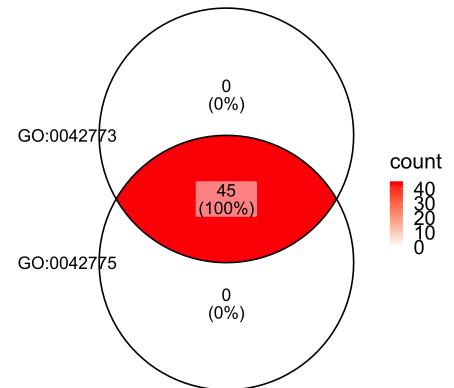


Figure 8: Venn diagram of the overlap of genes of GO:0042775 and GO:0042773.

Table 3: Remaining genes after reduction of downregulated gene sets.

Jaccard Coef.	% A549	% HEp-2	% hSAEC 16h	% hSAEC 24h
0.1	61.22	57.83	66.62	75.2
0.2	82.6	88.67	90.46	90.14
0.3	91.41	91.33	95.76	92.38
0.4	91.41	91.33	95.5	94.77
0.5	93.56	97.11	98.8	96.56
0.6	97.42	98.8	98.01	97.46
0.7	98.28	98.8	98.01	97.46
0.8	99.36	99.28	98.28	97.68
0.9	99.57	99.04	100.0	99.85
1.0	100.0	100.0	100.0	100.0

downregulated gene sets were left over. 210 upregulated and 66 downregulated gene sets remained for the HEp-2 gene sets. In the early infected hSAEC gene sets 212 up- and 89 downregulated gene sets remained and in the late infected hSAEC gene sets, 239 upregulated and 131 downregulated gene sets remained. The overlap in gene sets between the four upregulated data sets is at 150 gene sets from 355 in total (42.25%). For the downregulated data sets the overlap is at 75 gene sets of 170 total (44.11%). Overlap is here defined as a gene set that is present in at least 3 different data sets. This can be seen in the two upset plots in the Figures 7 and 9. The gene sets that were in at least 3 different data sets were further analyzed. To get a better overview, we categorized them into 18 different categoriess.

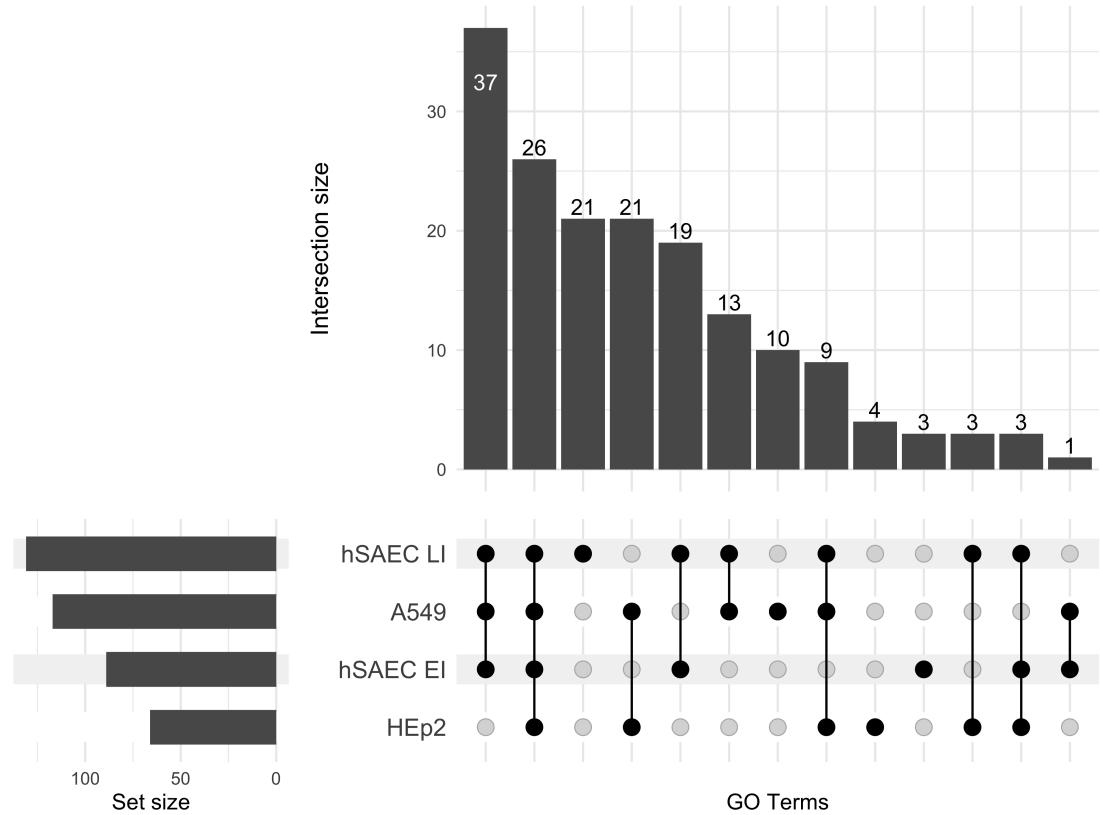


Figure 9: Upset plot of the downregulated gene sets. 75 gene sets are in at least 3 out of 4 groups. LI: *late infected* EI: *early infected*

Figures 10 and 11 show the up- and downregulated categories of the A549, HEp-2, hSAEC early (16h) infected and hSAEC late (24h) infected data sets. Percentages are calculated with the earlier described *gene weight*. The category with the highest number of upregulated genes is the *immune system process*. It is at nearly 20% across all upregulated data sets, 19.33% for A549, 16.21% in HEp-2, 22.79% at 16 hours, and 21.07% at 24 hours of the hSAEC data. The category *morphogenesis/development* is also highly upregulated. 17.04% for A549, 17.74% for HEp-2, and 11.52% and 10.93% in 16 and 24 hours of the hSAEC data set respectively. In A549 cells the downregulated *cell cycle* fills about 28.87% of the pie chart of all downregulated gene sets, whilst the *metabolic process* accounts for 38.07%. Of all the downregulated gene sets from the HEp-2 data set, 29.22% were *cell cycle* gene sets, and 51.65% of *metabolic process*. In the early infected data set of the hSAEC data the *cell cycle* is downregulated with 28.3% of all downregulated gene sets. *Metabolic process* represents 20.11% of all downregulated categories. In the late infected data set *cell cycle* stays at around the same level with 26.64% while *metabolic process* rises to 30.77%.

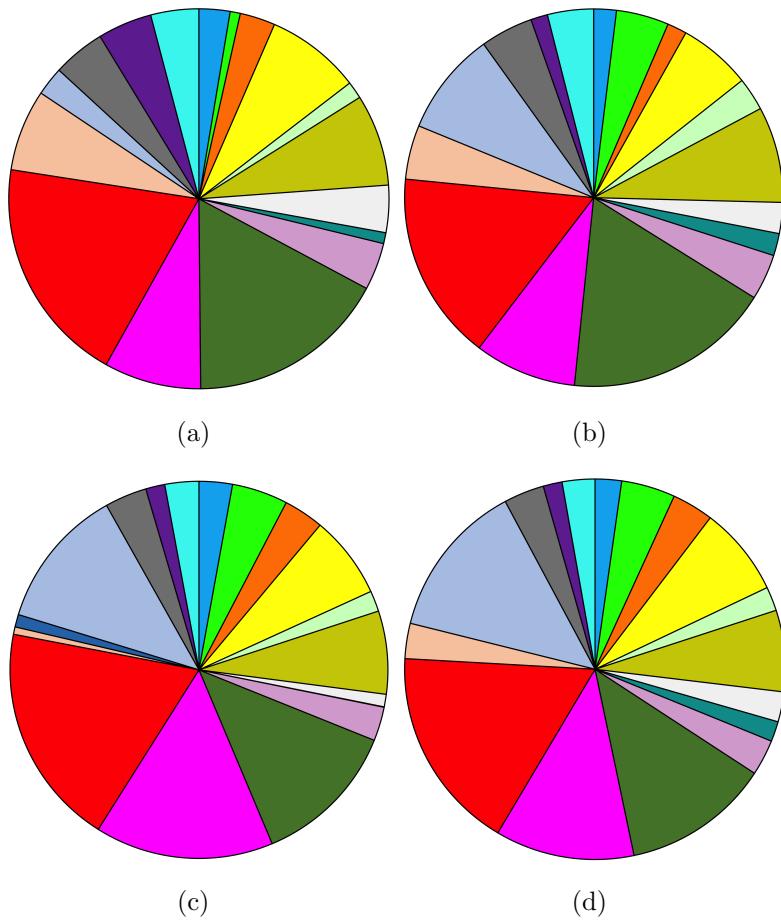


Figure 10: Pie charts of the gene sets being enriched with upregulated genes in **(a)** A549 and **(b)** HEp-2 cells late, **(c)** 16h (early) and **(d)** 24h (late) in hSAEC after infection. The gene sets were grouped into different categories. The groups are scaled according to *gene weight*.

Category	%A549	%HEp-2	%hSAEC 16h	%hSAEC 24h
cell adhesion	4.09	2.95	2.91	2.78
cell cycle	4.62	1.15	1.64	1.61
cell death	4.46	3.03	3.59	3.43
cellular component organization	2.44	9.13	12.13	13.3
ER/endosome/lysosome related process	0	0	1.08	0
gene regulation	6.94	3.56	0.63	2.99
immune system process	19.33	14.48	19.08	17.37
metabolic process	8.27	8.6	15.28	11.81
morphogenesis/ development	17.03	18.38	12.57	12.51
multicellular organismal process	4.02	4.31	2.91	3.04
neural related process	0.93	2.33	0.03	1.71
other	4.02	3.69	1.07	2.55
protein modification/ signaling	7.81	7.11	7.12	6.92
receptor related process	1.46	3.15	1.78	2.05
response to stimulus	8.07	8.5	7.1	7.6
stress response	3.01	2.93	3.44	3.53
transport	0.87	5.28	4.79	4.58
viral related process	2.63	1.42	2.85	2.24

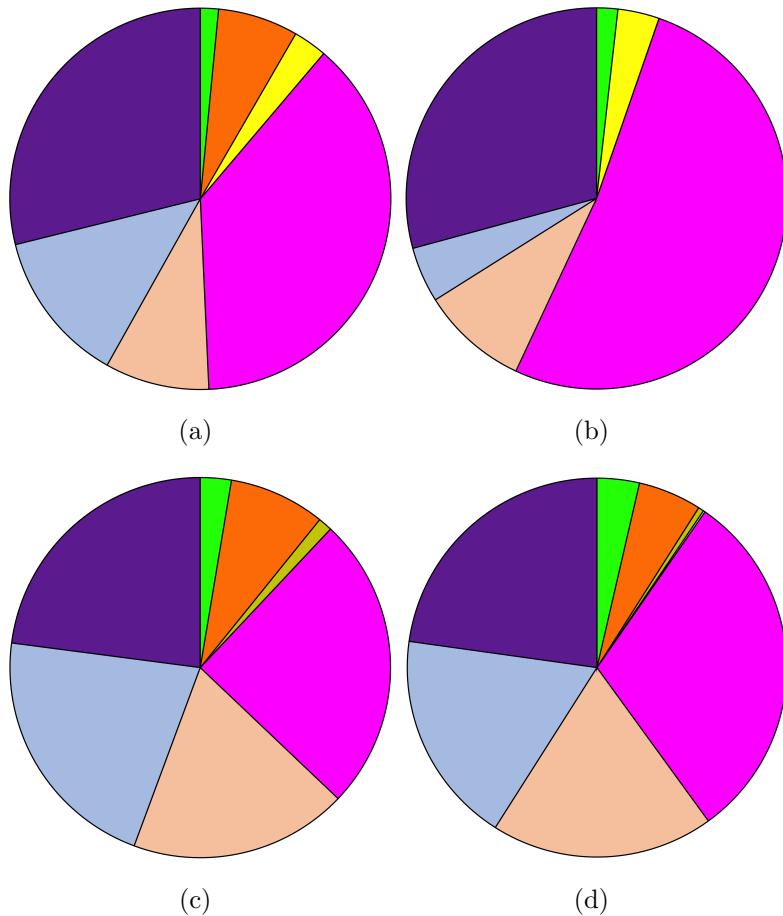


Figure 11: Pie charts of the gene sets being enriched with downregulated genes in (a) A549 and (b) HEp-2 cells late, (c) 16h (early) and (d) 24h (late) in hSAEC after infection. The gene sets were grouped into different categories. The groups are scaled according to *gene weight*.

Category	%A549	%HEp-2	%hSAEC 16h	%hSAEC 24h
cell cycle	28.87	30.8	22.93	22.79
cellular component organization	13	7.06	21.42	18.21
gene regulation	8.85	5.38	18.54	18.99
metabolic process	38.07	52.29	25.1	30.37
other	0	0	0	0.16
protein modification/ signaling	0	7.11	1.2	0.47
response to stimulus	2.86	3	0	0
stress response	6.82	0	8.17	5.41
transport	1.53	1.47	2.63	3.6

3.4 Comparing Gene Expression Profiles with Data from Host Factor Screens Shows High Overlap in Genes Involved in Cytokine Related Pathways and Cell Death

With the data provided by Hoffmann *et al.*[58] and Lingemann *et al.*[59], we compiled a comprehensive list of potential HDFs and HRFs. A gene listed in both datasets was considered as a host factor. The dataset from Lingemann *et al.*[59] also included a z-score for each gene. A negative z-score below -0.2 indicated less viral expression and those genes could thus be considered HDFs. A positive z-score above 0.2 indicated more viral expression and thus those genes were considered HRFs. Genes with a z-score between -0.2 and 0.2 were removed from the list of host factors. The subsequent intersection of these host factors with the upregulated genes in the hSAEC dataset resulted in the identification of 165 host factors, comprising 94 unique HDFs and 71 unique HRFs. The hSAEC dataset was chosen for further analysis, since infected primary cells are a more realistic model of an infected patient compared to infected cancer cell lines. To gain an insight into the distribution of host factors, the gene sets from both hSAEC time points were merged by combining all upregulated gene sets from the early infected data set and from the late infected data set. Duplicate gene sets were merged by considering all genes expressed in early and late infected stages of the gene set. The result is visualized in Figure 12 , displaying a pie chart representing the distribution of categories within the hSAEC data. The outer ring of the pie chart incorporates a bar chart, where each bar represents a gene set. The *immune system process* category emerged as the largest, containing 34.26% of all gene sets, followed by morphogenesis/development with 10.64%. A more detailed rankings based on the ratio of unique HDFs to unique genes within each group is presented in Table 4. Notably, the category *neural related process*, even though exhibiting a high ratio of HDFs to genes, was omitted from further analysis as to it only consisted of one gene set with only 19 genes. Further exploration focused on the categories *cell death* and *cytokine pathway*. The *cell death* category was selected based on its ranking. The *cytokine pathways* category, a sub-category of the *immune system process*, was selected as it was striking that the *immune system process* category showed so many HDFs, even though one would not expect that of the immune response. The *cytokine pathways* predominantly comprises pathways related to cytokine production and cytokine-mediated signaling (Table 5). The cell death category, as depicted in Table 6, includes the apoptosis pathway along with other terms such as pyroptosis and necrotic cell death.

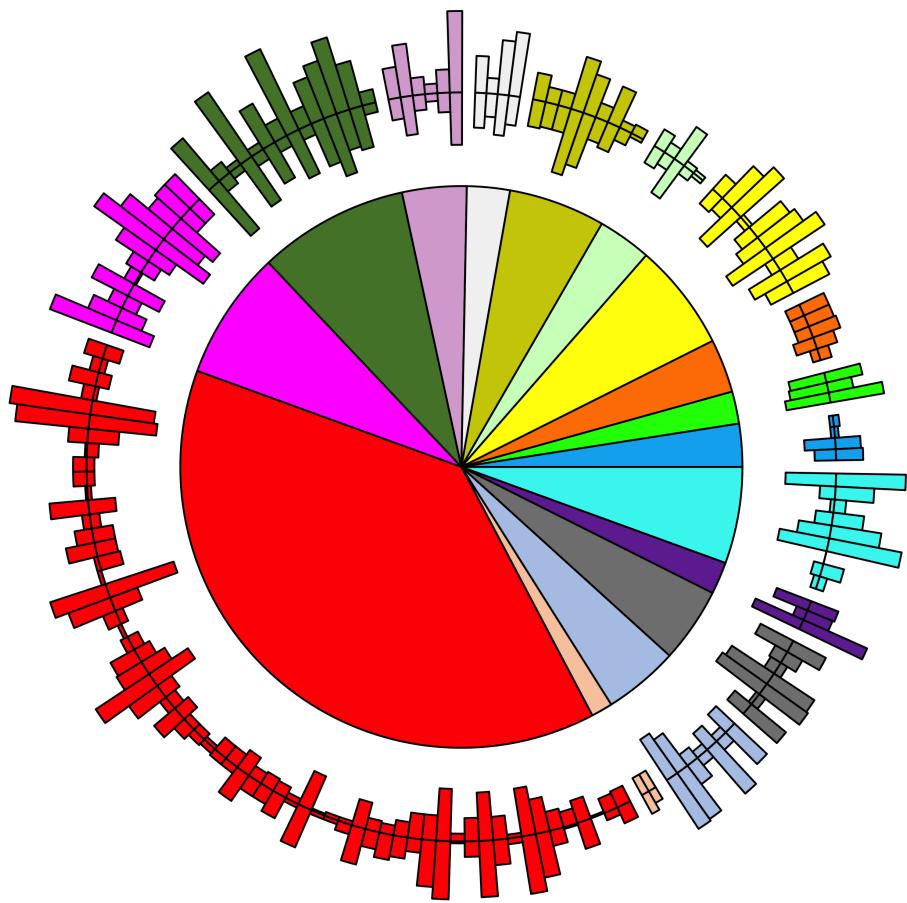


Figure 12: Pie chart of the intersection in gene sets in the upregulated hSAEC data. The bar plot around this pie chart represents the differentially expressed HFs. Bars showing inward represent HRFs and outward HDFs. The length of the bars is in relation to the number of HDFs and HRFs. The size/percentages of the pie slices are dependent on the number of gene sets, not on the gene impact values (see 2.3.3).

Category	%
cell adhesion	3.7
cell cycle	2.31
cell death	5.1
cellular component organization	8.8
gene regulation	0.93
immune system process	34.26
metabolic process	5.56
morphogenesis/ development	10.64
multicellular organismal process	2.78
neural related process	0.4
other	1.39
protein modification/ signaling	3.24
receptor related process	2.31
response to stimulus	7.88
stress response	4.63
transport	3.7
viral related process	2.31

Table 4: Ordered list of categories by the ratio of unique HDFs to genes

Category	HDFs/Genes	# HDFs
cell adhesion	6.07	15
cell cycle	8.92	19
cell death	10.37	25
cellular component organization	6.68	42
cytokine pathways	7.37	25
gene regulation	4.3	4
immune system process	7	50
metabolic process	7.74	39
morphogenesis/ development	6.72	44
multicellular organismal process	7.17	21
neural related process	10.53	2
other	9.89	18
protein modification/ signaling	8.75	28
receptor related process	7.39	19
response to stimulus	7.65	42
stress response	7.89	31
transport	7.42	25
viral related process	6.5	13

Table 5: Gene sets of the sub-category cytokine pathways.

Name	Number of HDFs	Up/Down regulated	p-value (from g:Profiler)
cytokine-mediated signaling pathway	6	183/33	3.3e-18
cytokine production	14	247/98	2.51e-25
cytokine production involved in immune response	3	41/12	1.71e-05
negative regulation of cytokine production	3	85/39	3.74e-05
negative regulation of response to cytokine stimulus	0	23/14	0.0079
positive regulation of cytokine-mediated signaling pathway	1	28/3	4.46e-05
positive regulation of cytokine production involved in immune response	1	32/9	0.0002
positive regulation of tumor necrosis factor superfamily cytokine production	1	37/13	0.0014
regulation of cytokine-mediated signaling pathway	2	58/18	5.33e-05
regulation of response to cytokine stimulus	2	61/20	2.69e-05
regulation of T cell cytokine production	2	18/2	9.05e-06
response to cytokine	6	318/115	6.53e-33
tumor necrosis factor superfamily cytokine production	2	51/18	0.0015

Table 6: Gene sets of the category cell death.

Name	Number of HDFs	Up/Down regulated	p-value (from g:Profiler)
activation of cysteine-type endopeptidase activity involved in apoptotic process	4	28/14	0.0027
apoptotic signaling pathway	14	176/114	5.53e-11
extrinsic apoptotic signaling pathway	6	75/40	5.16e-10
intrinsic apoptotic signaling pathway	10	93/57	0.0015
negative regulation of apoptotic process	11	246/159	1.5e-08
negative regulation of apoptotic signaling pathway	6	67/50	0.0006
negative regulation of extrinsic apoptotic signaling pathway	2	33/22	0.0014
negative regulation of programmed cell death	12	251/165	5.47e-08
positive regulation of apoptotic signaling pathway	3	48/21	0.0023
positive regulation of programmed cell death	10	173/96	8.03e-15
regulation of cysteine-type endopeptidase activity involved in apoptotic process	7	57/38	0.0024
regulation of endopeptidase activity	12	102/49	1.08e-08
release of cytochrome c from mitochondria	4	21/12	0.0114
pyroptosis	1	15/1	0.0211
macroautophagy	7	92/41	0.0001
programmed necrotic cell death	4	18/11	0.0017
apoptotic mitochondrial changes	4	29/27	0.0068
necroptotic process	4	16/9	0.0178

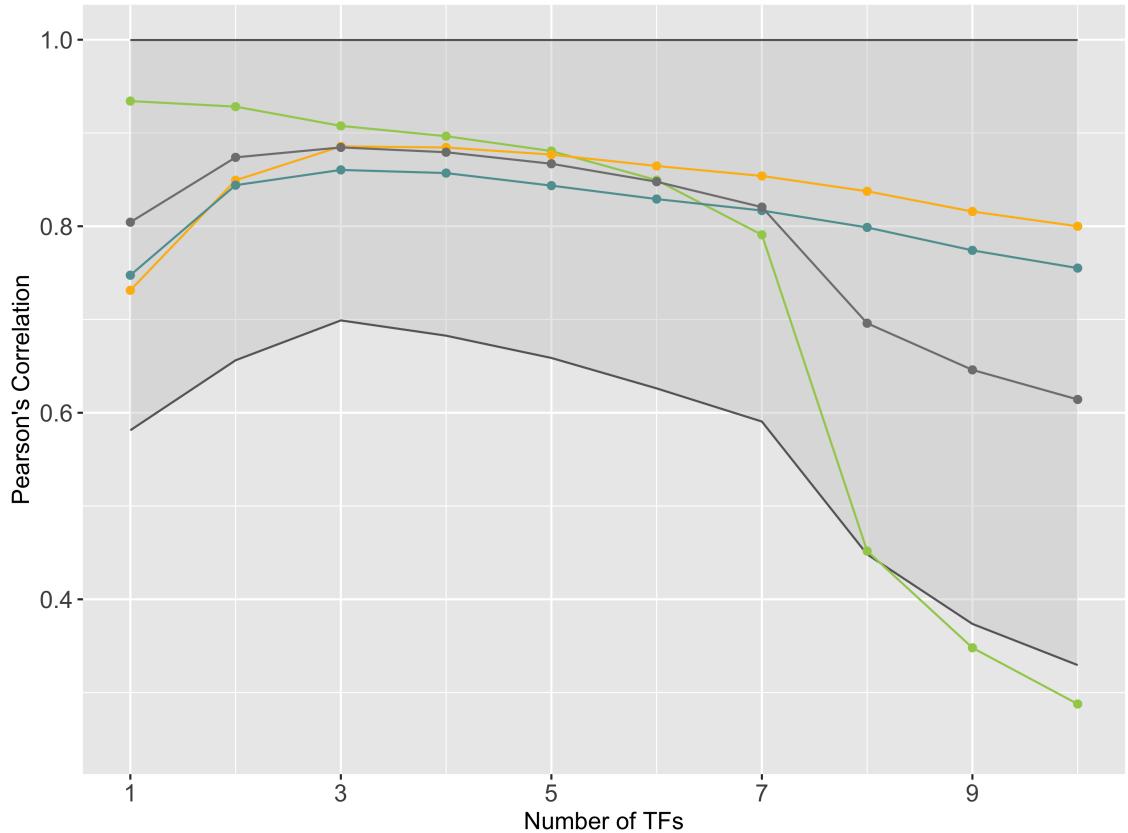


Figure 13: Correlation between the predicted and the observed gene expression for the different model sizes. The green line shows the correlation in the hSAEC data, gold for A549 and blue for the HEp-2 data. The grey line shows the mean correlation the transparent ribbon behind the mean line shows the standard deviation around the mean.

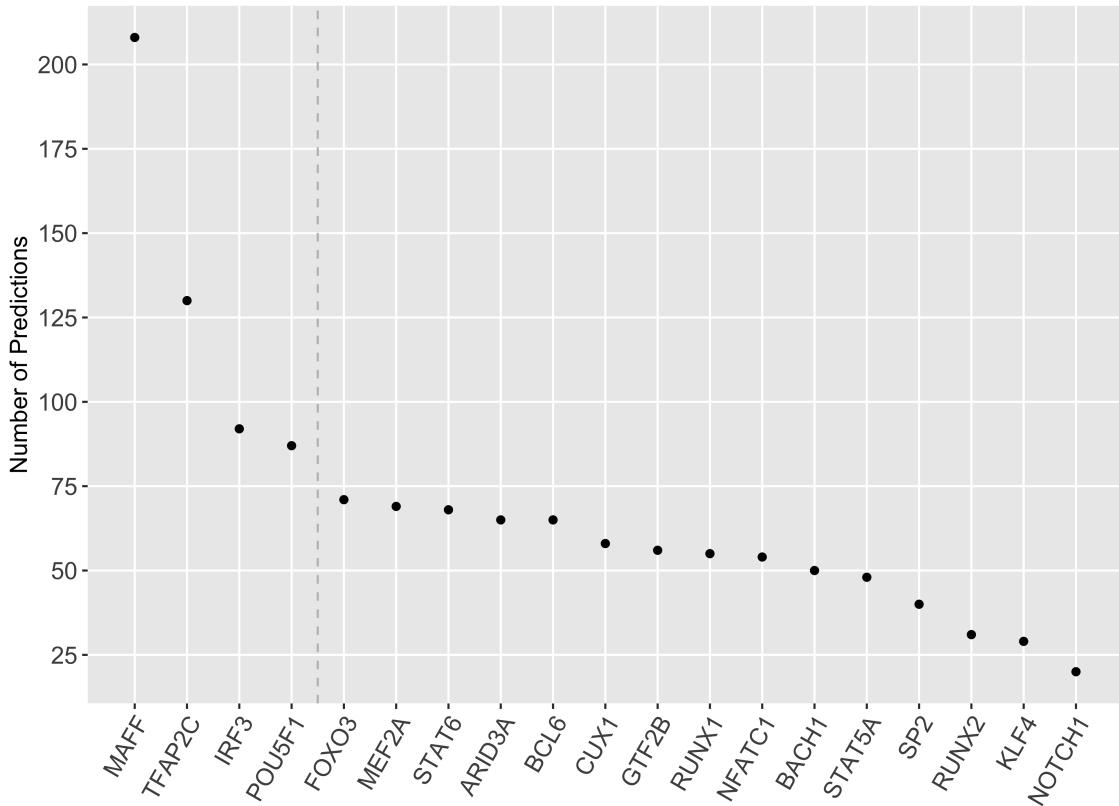


Figure 14: Shows the frequency with which a specific TF for *cytokine pathways* was used for predictions in models with 3 to 7 TFs. The dashed line shows the chosen threshold.

3.5 Gene Regulatory Models Predict MAFF, TFAP2C, IRF3 and POU5F1 Regulating Cytokine Related Pathways due to Virus Infection

To predict the TFs regulating a gene, we employed the MIPRIP tool[67] on each of the genes present in all data sets. This was done to increase the data points for predictions. Utilizing this approach, we optimized 150 distinct linear regression models for each of the 575 genes for each of the 3 expression data sets. Only upregulated genes were considered, but the TFs selected by the models did not need to be upregulated. These models were refined through a mixed integer programming (MIP) framework, employing 5 repetitions, 3 cross-validations, and 10 models per iteration, with 1 to 10 TFs considered in each model. This accumulated to 258.750 models. The objective was to develop models capable selecting TFs that predict the gene expression of the TG. To evaluate the performance of the models, we computed the difference between the predicted TG expression and the observed TG expression, which served as the parameter to minimize. Additionally, we conducted Pearson's

correlation tests between predicted and observed TG expressions (Figure 13) to further assess the quality of predictions. MIPRIP predicted 154 TFs for all 575 genes. Pearson's correlation tests revealed a decrease in prediction accuracy for models incorporating more than 7 TFs in the hSAEC expression dataset. Similarly, predictions on the A549 and HEp-2 expression datasets exhibited weaker correlation when 1 or 2 TFs were included in a single model (Figure ??). Consequently, only TFs predicted in models comprising 3 to 7 TFs were retained for further analysis. For these TFs, p-values were calculated and FDR corrected performing a permutation test (see Materials & Methods section 2.4). Only TFs with an FDR corrected p-value below 0.05 were retained. The remaining TFs underwent further reduction using the same redundancy reduction method used prior for gene sets (solving a stable set problem, see Materials & Methods, section 2.4), resulting in 49 TFs. Among these 49 TFs, 25 were found to be upregulated in all the data sets (A549, HEp-2, hSAEC early and hSAEC late) combined. This restriction was done due to it being easier to knockout a TF in an experiment instead of knocking it in. Upon closer examination, 19 out of the 25 upregulated TFs were predicted as regulators of the 339 upregulated genes associated with cytokine-related pathways. Analysis of the frequency distribution of these TFs across all models containing 3 to 7 TFs highlighted four predominant TFs: MAFF, TFAP2C, IRF3, and POU5F1 (see Figure 14). MAFF emerged as the most frequently predicted TF, occurring in n=208 models, followed by TFAP2C (n=130), IRF3 (n=92), and POU5F1 (n=87). A comprehensive examination of the genes within the cytokine pathways was conducted. A. W. Burgess collected a comprehensive list of known growth factors and cytokines in 2015[124]. M. J. Cameron and D. J. Kelvin maintained the Cytokine and Chemokine database in the Madame Curie Bioscience Database until 2013[125]. By comparing these resources with the complete list of TGs (Table 7) regulated by the four selected TFs, the following associations between TFs and growth factors/cytokines are highlighted. MAFF is predicted to directly regulate IL15, IL36B and IL6. IRF3 is predicted to regulate CXCL1, CXCL10, IFNB1, IFNL1, IFNL2, IFNL3, IL15, IL23A, IL33, IL6, IRF3, IRF7 and TNF. POU5F1 is predicted to regulate VEGFC. TFAP2C is predicted to regulate IL24, TNF, VEGFA and VEGFC. All these regulations are present in the gHRN[67]. Not all of these genes regulated by the four TFs are upregulated in the data used. The upregulated genes are marked bold in the Table 7.

To estimate the regulatory impact of the TFs, we calculated the Pearson's correlation coefficient (PCC) between the predicted gene expression and the observed expression of the hSAEC expression data. For each of the four TFs, we determined the PCC for all genes within the cytokine pathways category, as well as exclusively for the TGs regulated by the respective TF. The median of all positive PCCs and the median of all negative PCCs are summarized in Table 8, displaying the median PCC values for each TF across all genes and the median PCC value for the TGs.

Table 7: Genes regulated by the 4 predicted transcription factors related to cytokine pathways. Bold genes are cytokines, chemokines or growth factors. (These interactions are all in the gHRN with at least a edge strength value of 0.5.)

TF	Regulated Genes
MAFF	ACKR3, ACTN4, ADAM10, ADAM9, ADIPOR2, AFF3, AGPAT1, AKAP12, AKAP6, ANKRD42, ANXA4, APP, ARRB1, ATF2, BBS4, BCAT2, BRCA1, BTN2A2, BTN3A1, BTN3A3, C1QTNF3, CAPN2, CARD11, CARD8, CAV1, CCBE1, CD226, CD276, CD2AP, CD36, CD44, CD46, CD47, CNOT7, CREB1, CSF1R, CTBP2, CTH, CUL3, CYLD, DAB2IP, DDX1, DDX21, DDX56, DENND1B, DOK1, DUSP1, EIF2AK2, ELF4, ETV3, EZH2, F2RL1, FBLN1, FER, FLNB, FOXO3, FOXP1, FXR1, GAS6, GBP5, GCLC, GCLM, GPNMB, GSK3B, GSN, HCLS1, HDAC4, HDAC9, HIPK1, HLA-E, HOMER2, IFNAR2, IGF2BP1, IGF2BP2, IGF2BP3, IL15 , IL18R1, IL1A , IL1RAP, IL1RAPL2, IL1RL1, IL1RL2, IL1RN, IL36B , IL4R, IL6 , IL6R, IQGAP1, IRAK2, ITGAV, ITGB8, ITIH4, JARID2, KEAP1, KLHL22, KYNU, LAMP3, LAPTOM4B, LEPR, LGR4, LIFR, LIMS1, LRCH1, LRP8, LRRK2, LSM14A, LTBP1, LYN, MAP2K5, MAP3K5, MAP3K7, MAP4K3, MAPKBP1, MAST2, MED1, MERTK, MFAP3, MME, MT1X, MTF2, MYNN, NAV3, NDFIP1, NFAT5, NFE2L2, NFKB1, NKIRAS1, NLRC5, NLRP3, NMI, NMNAT3, NOD1, NPR2, NPTN, NR5A2, NRP2, NUMBL, OAS1, OAS3, OSMR, OTUD4, OTULIN, OXSR1, OXTR, PCSK5, PDE4D, PFKP, PHC1, PIBF1, PIK3CD, PIK3R1, PLA2R1, PLCB1, PLCG2, PLD3, PML, POLR3B, PPARA, PPM1B, PRDM5, PRKCA, PRLR, PTAFR, PTK2B, PTPN1, PTPN11, PTPN2, PTPN22, PTPRJ, PT-PRS, RAD23B, RBM47, RC3H1, REL, RFL, RFTN1, RFX2, RHOA, RIF1, RIPK1, RNF125, RNF135, RNF216, ROBO1, ROCK2, RORA, RPL13A, RPS6KA5, RPS6KB1, RTN4, RUNX1, SCAMP5, SIRT1, SLC2A10, SLC37A4, SLC7A5, SLIT2, SMAD7, SNX10, SOCS5, SOD1, SOS1, SPPL2A, SPTBN1, SRC, STAT1, STAT3, STAT4, STK39, STXBP1, STXBP3, STXBP4, SULF1, TAF9, TFRC, TGFB2, TIA1, TLE4, TMEM106A, TNF , TNFRSF1A, TP53, TRAF3IP2, TRAF4, TRAF5, TRAF6, TREX1, TRIM27, TRIM44, TYMS, UBXN2A, ULK1, USP10, VIM, WBP1L, WNK1, XRCC5, YTHDC2, YTHDF2, ZBTB20, ZP3
IRF3	B2M, BTN2A2, CCL5 , CD36, CITED1, CXCL1 , CXCL10 , DDOST, FER, G3BP1, HMGB2, IFNB1, IFNL1, IFNL2, IFNL3, IKBKB, IL12A , IL15 , IL1A , IL23A , IL33 , IL6 , IRF3 , IRF7 , JARID2, KLF5, MAVS, NR1H2, NR1H3, PML, PPARG, PPP3CB, PRKACA, RBX1, RELA, RTN4, SMAD3, SRF, STIP1, TICAM2, TNF , TRAF2, TYK2, ZC3HAV1
POU5F1	ACKR3, ACTG1, ACTN4, ACTR2, ACTR3, ADAM9, AKAP12, AKT1, ARRB1, BAD, BANF1, BBS4, BCL3, BMI1, BRCA1, BST2, CASP3, CDH3, CDK4, COMMD7, CUEDC2, DLL1, DPYSL3, DUSP1, EPHA2, ETV3, EXT1, FLNB, FOS, FOXA2, GATA6, GLDC, GPSM3, GSTP1, HDAC4, HES1, HIC2, HK2, HMGB2, HSP90AA1, HSPB1, HSPD1, IFitm2, IGF2BP1, IGF2BP2, IGF2BP3, IL1RL1, ILK, IQGAP1, ISL1, JAK2, JARID2, KDM3A, KLF2, KLF4, KLF5, KRT18, LGR4, LIFR, LIMS1, LYN, MAP2K7, MAST2, MFAP3, MIF, MMP2, MRAS, MSC, MTF2, MUL1, MYBL2, NFKB2, NR5A2, PARK7, PATZ1, PCK2, PDE4D, PHC1, PIK3CD, PLD3, PLSCR1, PML, POLR3G, PPARG, PTPN2, RARA, RIF1, RNF125, RPS3, RTN4, SFRP1, SHMT1, SIRPA, SIRT1, SKIL, SLC2A10, SLIT2, SMAD3, SMAD7, SPRY2, SPTBN1, SRGN, STAT3, STAT4, SULF2, TFRC, TLE4, TMSB4X, TNFRSF19, TNIP2, TP53, TPR, TRIM27, ULK1, UPF1, VEGFC , XRCC5, YAP1, ZC3HAV1, ZFP36L1
TFAP2C	ACTN4, AGPAT2, ANXA4, APP, ASB1, AZI2, B3GNT2, BBS4, BCL3, CCDC88B, CD14, CD55, CD58, CD81, CDC34, CDC42EP4, CDH3, CSF2, DAB2IP, DAPK3, ECM1, EPS8, ETV3, F3, FASN, FLNB, FLOT1, FOS, FURIN, GSK3B, GSTP1, HDAC7, HK1, HK2, HOMER2, HPSE, HSP90AA1, HSPA5, IL24 , IL4R, IL6R, IRAK2, JAK1, KEAP1, KLF4, KRT18, KYNU, MAPK3, MAPK9, MAST2, MBP, MRPL15, NR1H3, OXSR1, PDE4D, PRKCA, PRKCZ, PRLR, PTGS2, PTPRS, RARA, RARG, RFTN1, RIF1, RUNX1, SFRP1, SH2B2, SMAD3, SOD1, SPHK1, SPRY2, STAT1, STX4, SULF1, SULF2, TNF , TNFRSF19, TNFRSF1A, TNIP2, TRAF3, TRIM27, TRIM56, UBE2G2, VEGFC , XBP1, ZC3H15, ZYX

Table 8: PCC between the TF activity and gene expression in the hSAEC data. The first line in each row represents the median PCC for all PCCs greater than 0. The second line represents the median PCC for all PCCs lower than 0.

TF	PCC all genes	PCC TGs
MAFF	PCCs > 0:	0.611
	PCCs < 0:	-0.956
IRF3	PCCs > 0:	0.582
	PCCs < 0:	-0.932
POU5F1	PCCs > 0:	0.643
	PCCs < 0:	-0.979
TFAP2C	PCCs > 0:	0.459
	PCCs < 0:	-0.597

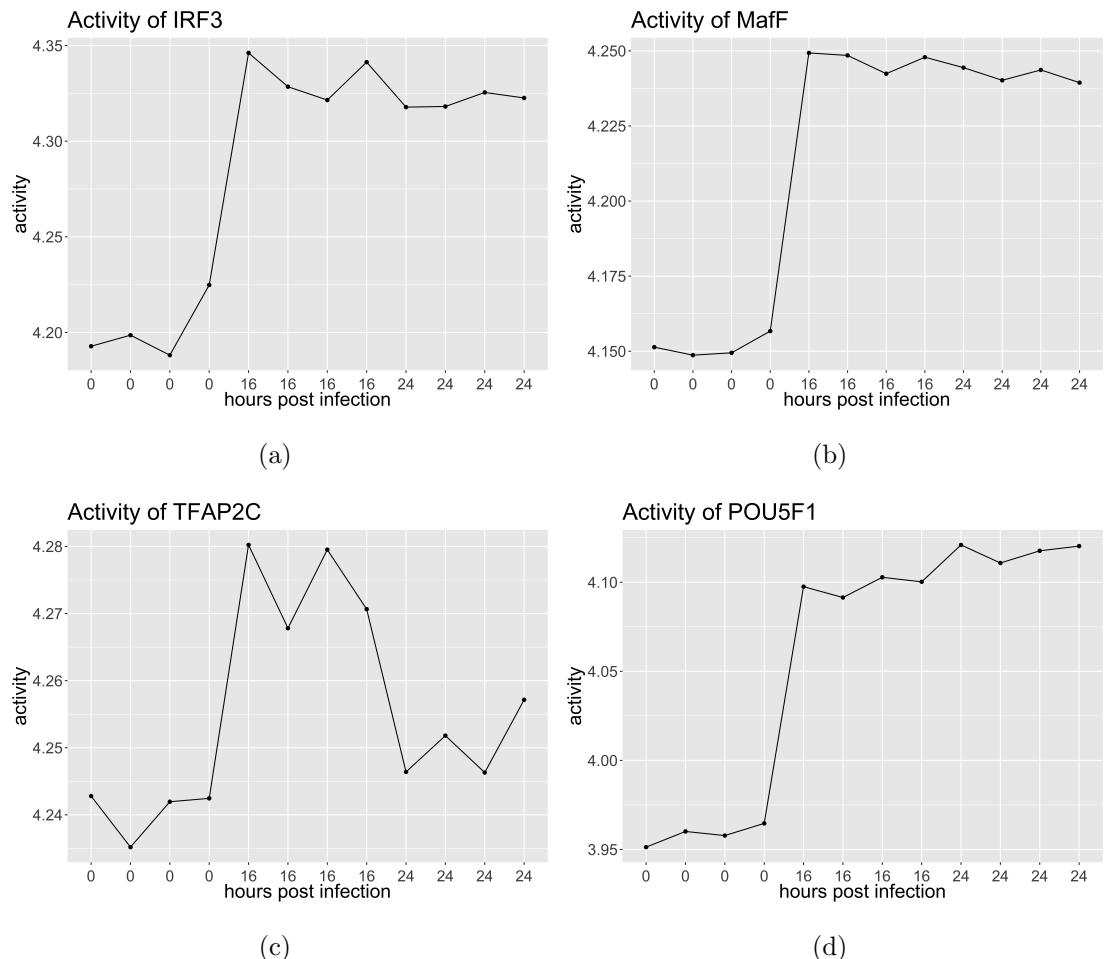


Figure 15: Activity of the four TFs IRF3, MafF, TFAP2C and POU5F1 over time for the hSAEC data.

Figure 16 illustrates the interactions between the predicted TFs and the aforementioned cytokines, chemokines, and growth factors.

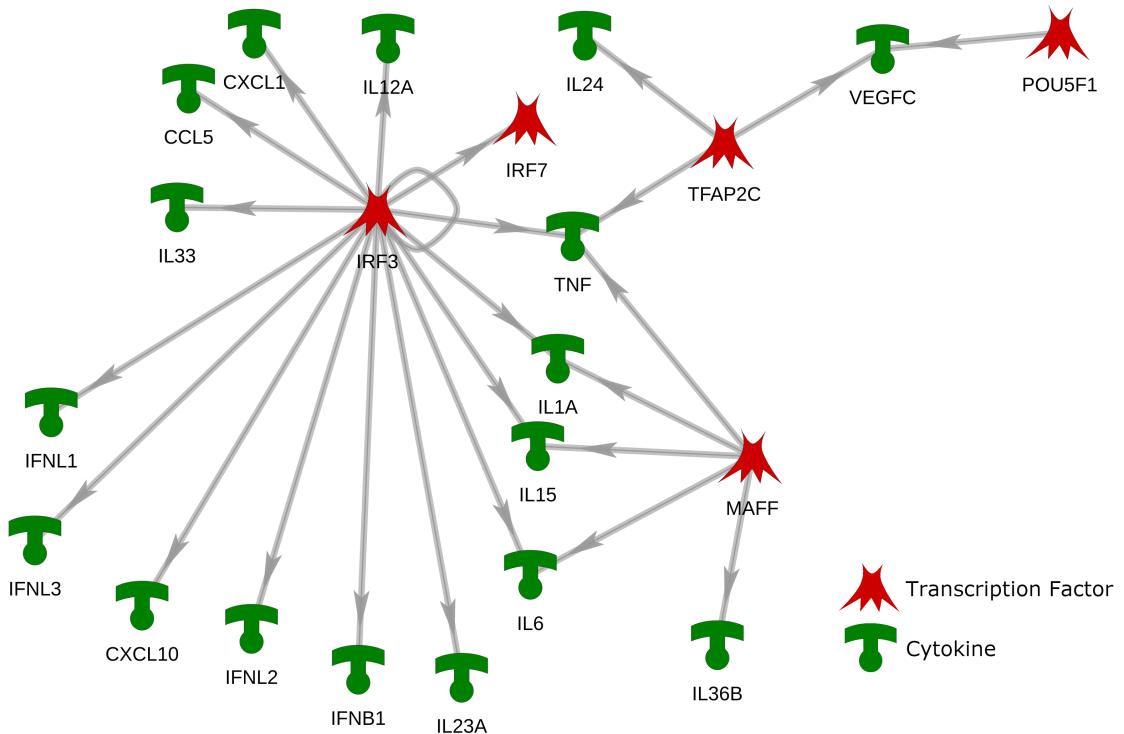


Figure 16: Regulation of cytokines, chemokines and growth factors by MAFF, TFAP2C, IRF3 and POU5F1. Red symbols represent TFs, green symbols are ligands. The ligands are all classified as cytokines. The arrows represent transcriptional regulation. In the network are gene products (proteins) visualized, however the interactions are between TF and genes. The network was visualized with MetaCore™[117], the edges are from the gHRN[67].

3.6 Gene Regulatory Models Predict STAT2, ATF2, TFAP2C and SIX5 Regulating Cell Death Related Pathways

From the initial list of 25 upregulated transcription factors (TFs) predicted by the models (see section 3.5), 12 were identified as potential regulators of the 368 upregulated genes associated with the cell death category. Figure 17 illustrates the frequency of utilization of these TFs across the previously selected models. Among them, STAT2, ATF2, TFAP2C, and SIX5 emerged as the most abundant TFs. Specifically, STAT2 and ATF2 were prominently predicted in 217 and 212 models, respectively, while TFAP2C and SIX5 were predicted in 130 and 117 models, re-

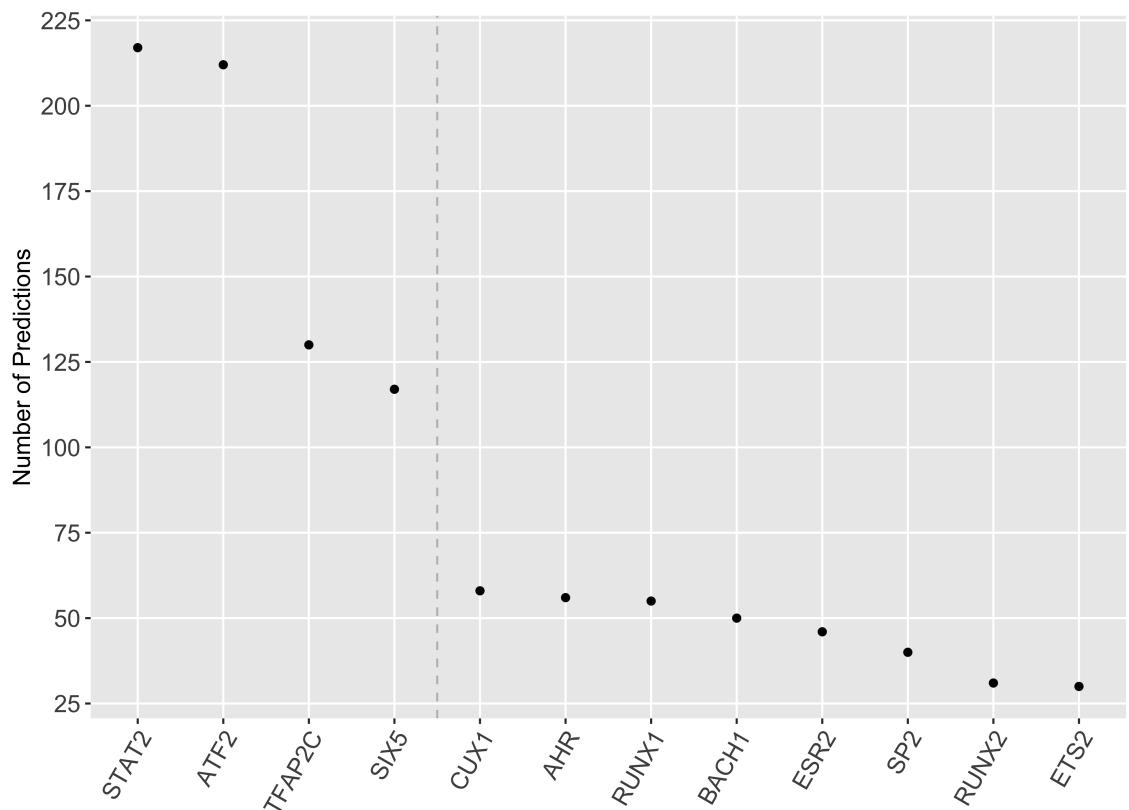


Figure 17: Shows the frequency with which a specific TF for *cell death* was used for predictions in models with 3 to 7 TFs. The dashed line shows the chosen threshold.

spectively. Table 8 provides a detailed list of the TGs regulated by these TFs within the *cell death* category.

Table 9: TGs regulated by the 4 predicted transcription factors related to *cell death*.
 Bold TGs are key genes in apoptosis and autophagy.

TF	Regulated Genes
TFAP2C	ATP6V0D1, BAG3, BCL11B, BCL3, BIK , BTC, CDKN1A, CSF2, CST3, DAPK3, DDR2, DEDD2 , F3, FOXA1, FURIN, HK2, HSH2D, HSPA5, IL6R, KLF11, KLF4, LYPD3, MMP9, NOTCH2, NR3C1, PRLR, PTGS2, S100A8, S100A9, SMURF1, SOD2, SPHK1, SPRY2, SQSTM1, STX4, TBC1D10A, TLE1, TNF , TNFAIP8, TNIP2, VEGFA, ZFYVE1
STAT2	ADAMTSL4, ADAR, AIM2, AKT2, ANP32B, BAK1, BCL2 , CASP8 , CAV1, CD274, CHMP5, DNAJA1, EIF2AK2, ERP29, FBXW7, FOXP1, FYCO1, GBP1, GBP2, GCLM, GRINA, HAX1, HMOX1, ICAM1, IFI16, IFI27, IFI6, IFIT2, IFIT3, IL12A, IL6, IRF7, KLF11, LGALS1, MDK, MLKL , MUC1, NBN, NPPC, NR3C1, NUP62, PML, PPARA, PRKAA1, PSMA3, PSMB8, PTK2B, RBCK1, RICTOR, SH3GLB1, SLFN12, SMCR8, SP100, ST6GAL1, STAT3, STX12, STX17, TNF , TNFSF10 , TPT1, TRIAP1, TRIM21, TXNIP, USP15, YWHAZ, ZFYVE26
ATF2	ADAR, ADNP, ANXA1, ANXA5, APH1A, API5, ARF4, ARFIP2, ATF2, ATF3, ATF4, ATG10 , ATG5 , ATG9A , ATP2A2, ATP6V0A1, ATP6V0C, BAD , BAG6, BCAP31, BCL2 , BCL2L13, BDNF, BIRC2 , BNIP1, CALM1, CALR, CARD8, CASP10 , CASP6 , CCAR2, CCL5, CCND1, CCND2, CDIP1, CDKN1A, CFL1, CHD8, CHMP3, CHMP7, CITED1, CITED2, CNTF, COL2A1, COPS5, CREB3, CTTN, CUL4A, CYCS, DDIAS, DDIT3, DDX5, DFFA , DNAJC3, DNM2, DPP9, DSTYK, EIF5A, ELL3, EP300, FOXO1, GADD45A, GADD45B, GAL, GCLC, GOLPH3, HAX1, HERPUD1, HIGD2A, HIP1, HIPK1, HMOX1, HSP90AB1, HSP90B1, HSPA9, HSPB1, HSPD1, HSPE1, IFNB1, IFT20, IFT57, IGF2R, IL6, IRS2, ITGA5, ITGB1, KDM1A, KLF11, LRSAM1, LYN, MAP3K5, MAP3K7, MAPK3, MAPK8IP3, MAPK9, MCOLN1, MED1, MLH1, MMP2, MMP9, MTOR, MYO18A, NAA35, NAA38, NACA, NCSTN, NDUFS1, NF1, NFKB1, NPPC, NR3C1, OGT, OPA1, P4HB, PARP2, PCID2, PDCD10, PDPK1, PHB2, PIH1D1, PIK3R3, PIKFYVE, PLAUR, PMAIP1 , PPARG, PPARC1A, PPIA, PPP1R10, PPP2R5C, PRKCD, PRR5, PSMA3, PTGS2, PTK2B, PTPN1, QRICH1, RAB7A, RB1, RELA, RFFL, RGL2, RIPK2, RNF103-CHMP3, RNPS1, RPL10, RPL11, RPS3A, RPS6, RRM2B, SCFD1, SEC22B, SERPINB5, SERPINH1, SFN, SHISA5, SIAH1, SIN3A, SIRT5, SLC7A11, SLC9A1, SNX5, SOX9, STAM2, STAT3, STX4, STYXL1, SUPV3L1, TFRC, TGFB1, TGFB2, TIMP1, TM2D1, TMED10, TMEM39A, TMEM41B, TMF1, TNF , TOM1, TRIAP1, TRIB3, TSC22D1, UBB, UBE2B, UBQLN1, USP30, USP36, VDAC1, VDAC2, VDR, VMP1 , VPS25, VPS37B, WDR24, XRCC2, YME1L1, YWHAZ
SIX5	ABL1, ADAR, ADNP, AKIRIN2, ALKBH1, AMBRA1 , APH1A, API5, ARAF, AREL1, ARF4, ARFIP2, ARHGEF7, ARL8B, ASAHI, ATAD5, ATF6, ATG101 , ATG13 , ATG14 , ATG16L1 , ATG16L2, ATG4B , ATG5 , ATG7 , ATP13A2, ATP6V0B, ATP6V0C, AUP1, AXL, BACE1, BAD , BAX , BCL2L12, BCL6, BIRC5, BIRC6, BNIP1, BRCA1, BRMS1, CAAP1, CALCOCO2, CAPN10, CARD8, CARM1, CCAR2, CDC34, CDIP1, CDK19, CDK5, CFDP1, CFL1, CHEK2, CHMP2A, CHMP2B, CHMP3, CHMP4A, CHMP4B, CHMP6, CIAPIN1, CLEC16A, COPS5, CPEB4, CRADD, CREB3, CSNK2A1, CTNNBL1, CTNS, CUL2, CUL4A, CYCS, CYLD, DAXX, DCUN1D3, DDIAS, DEDD, DFFA , DIDO1, DNAJA3, DNAJB6, DNM1L, DNM2, DNMT1, DPP8, DPP9, DSTYK, EHMT2, EI24, EIF2B5, EIF5A, ERBB2, EXOC1, EXOC4, EXOC8, EYA1, EYA2, FADD , FAF1, FASTK, FBXW7, FEM1B, FHIT, FIS1, FOSL1, FYCO1, FYN, GHITM, GLO1, GNAI2, GSK3A, HDAC10, HDAC2, HGS, HIGD2A, HIPK1, HMGB1, HSP90B1, HSPA9, HSPD1, HSPE1, HTRA2 , HYAL2, IFT20, ING4, INTS1, IRF3, ITCH, JTB, KATNB1, KLHL20, LAMTOR5, LATS1, LGMN, LIG4, LMNA, LRSAM1, MADD, MAP2K5, MAP3K10, MAP3K11, MAP3K7, MAPK7, MAPK8IP1, MAPK8IP3, MAPKAP1, MARK4, MBP, MCOLN1, MFF, MFN2, MITF, MLKL , MLST8, MNAT1, MSH2, MTCH1, MTM1, MTMR14, MUL1, MUTYH, MVB12A, NAA38, NACA, NBR1, NCK1, NCSTN, NDUFA13, NDUFS1, NDUFS3, NFE2L2, NLE1, NLRP3, NMT1, NR3C1, NSFL1C, OPA1, ORMDL3, P4HB, PAK2, PALB2, PARK7, PCID2, PDCD10, PDCD2, PDCD6, PDIA3, PELI1, PGAM5, PHB2, PHF23, PIH1D1, PIK3R3, PIP4K2B, PLAA, PLAGL2, POLDIP2, PPP1R10, PPP2R1A, PRDX2, PRKCI, PSENEN, PSMC5, PSMD10, PSME3, PTK2B, PTMA, RAB19, RAB1B, RAB5A, RAB7A, RAF1, RAG1, RBM5, RHOA, RHOT2, RNF103-CHMP3, RNF34, RNPS1, ROBO1, RPL11, RPS3, RPS3A,

TF	Regulated Genes (Continued)
SIX5	RPS6KA2, RPS6KB1, RRP1B, SCRIB, SEC22B, SENP1, SESN1, SESN2, SIRT2, SIRT4, SIRT5, SIVA1, SLC25A27, SLC25A5, SLC27A4, SMCR8, SNAP29, SNAPIN, SNF8, SNW1, SNX14, SNX18, SPATA2, SPHK2, SRPK2, ST3GAL1, STAM, STAMB, STK25, STK40, STRADB, STX4, STYXL1, SUPT5H, SUPV3L1, SYVN1, TAF9, TBC1D14, TFAP4, TFPT, THEM4, THOC6, TIMM50, TM2D1, TMED10, TMEM109, TMEM161A, TMEM39A, TOMM7, TOP2A, TOPORS, TP53 , TPT1, TRAF7, TRIAP1, TRIM39, TXNDC12, UBB, UBE2B, UBE2M, UBQLN1, UBXN6, UFC1, UFM1, USP15, USP30, VPS13D, VPS16, VPS25, VPS28, VPS37A, VPS41, VPS4B, VTA1, VTI1A, VTI1B, WDR24, WDR45B , WDR81, WWOX, YME1L1, ZC3HC1, ZFYVE1 , ZFYVE26, ZNF16, ZNF205, ZNF304, ZNF622, ZNF830

Following a literature search on Google Scholar (<https://scholar.google.com>) using keywords "cell death review," "apoptosis review," and "autophagy review," six highly cited publications were identified and selected. From these publications, two key gene lists were compiled: one comprising key genes for apoptosis[126][127] and the other consisting of key genes for autophagy[127][128][129][130]. These lists were subsequently refined by including the in the literature discussed genes with the genes present in the hSAEC dataset. Table 9 presents the regulatory interactions of the selected genes , while Figure 19 illustrates a network depicting the interactions between the TFs and the selected TGs. To evaluate the performance of the TFs , we calculated the PCC between the predicted gene expression and the observed expression of the hSAEC expression data. For each of the four TFs, we determined the PCC for all genes within the cell death category, as well as exclusively for the TGs regulated by the respective TF. The median of all positive PCCs and the median of all negative PCCs are summarized in Table 11, displaying the median PCC values for each TF across all genes and the median PCC value for the TGs. The PCCs seem not as evident as the PCCs of the *cytokine pathways*.

Table 10: Regulatory interactions between TFs and selected TGs of apoptosis and autophagy.

TF	Apoptosis Genes
MAFF	BIK, DEDD2, TNF
IRF3	BCL2, CASP8, MLKL, TNF, TNFSF10
POU5F1	BAD, BCL2, BIRC2, CASP10, CASP6, DFFA, PMAIP1, TNF
TFAP2C	BAD, BAX, DFFA, FADD, HTRA2, MLKL, TP53
	AMBRA1, A

Table 11: PCC between the TF activity and gene expression in the hSAEC data. The first line in each row represents the median PCC for all PCCs greater than 0. The second line represents the median PCC for all PCCs lower than 0.

TF	PCC all genes	PCC TGs
ATF2	PCCs > 0:	0.68
	PCCs < 0:	-0.957
STAT2	PCCs > 0:	0.194
	PCCs < 0:	-0.578
SIX5	PCCs > 0:	0.569
	PCCs < 0:	-0.944
TFAP2C	PCCs > 0:	0.204
	PCCs < 0:	-0.565

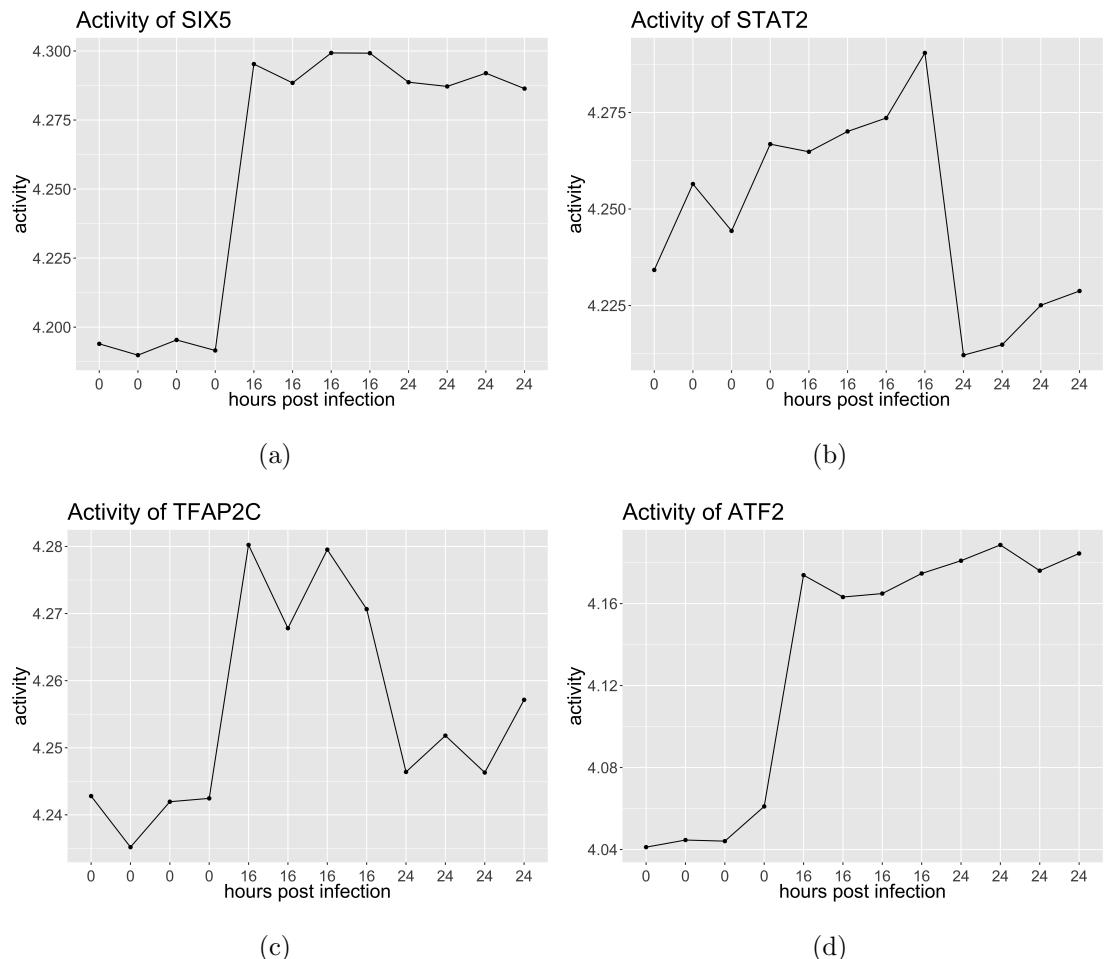


Figure 18: Activity of the four TFs SIX5, STAT2, TFAP2C and ATF2 for the hSAEC data..

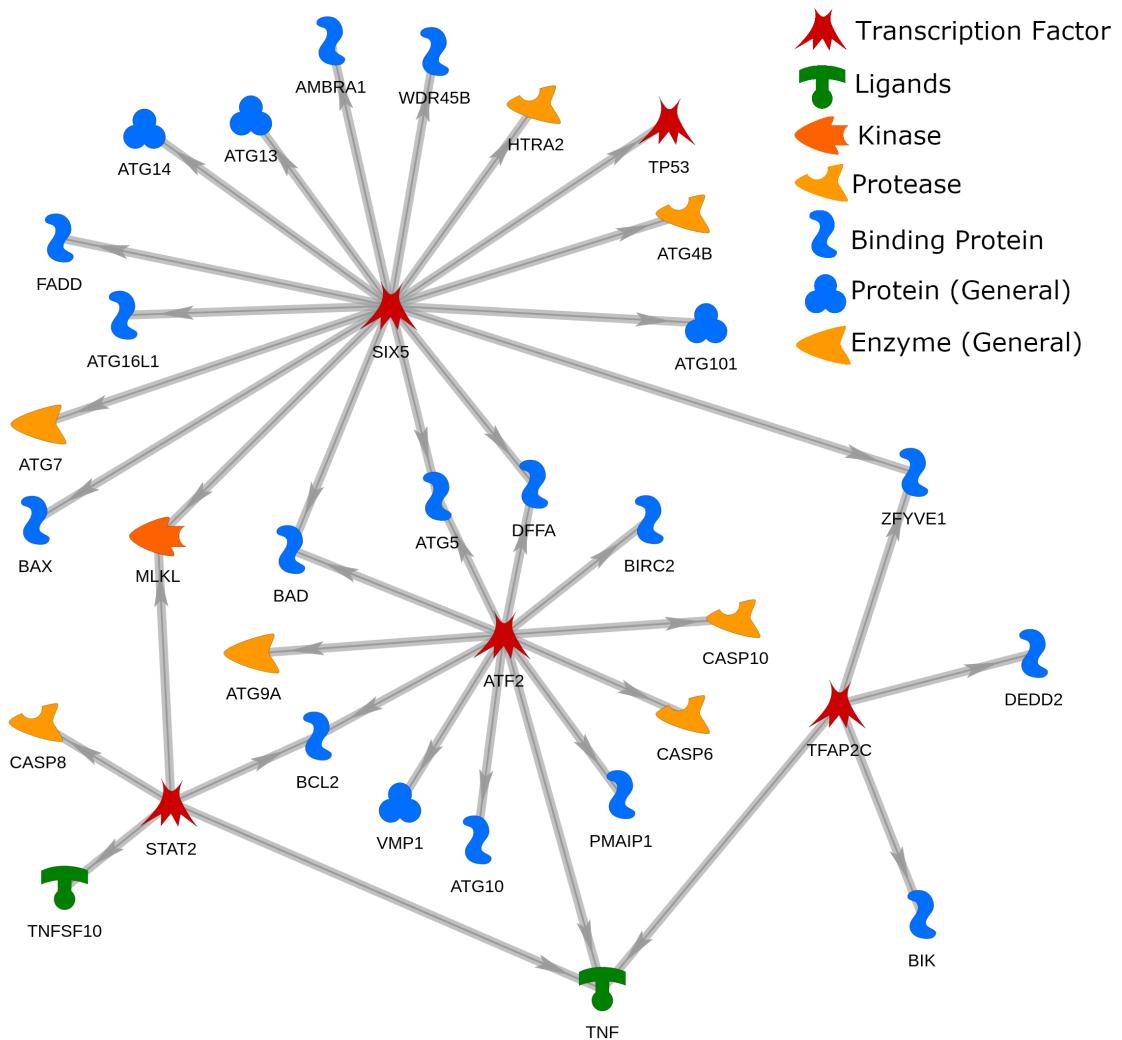


Figure 19: Regulation of gene products by STAT2, ATF2, TFAP2C and SIX5. Red symbols represent TFs (blue circle marks the four TFs mentioned), green symbols are ligands, blue “S” shapes are binding proteins, three blue dots shapes are proteins in general, dark orange is a kinase and light orange is a protease. The arrows represent transcriptional regulation. In the network are gene products (proteins) visualized, however the interactions are between TF and genes. The network was visualized with MetaCore™[117], the edges are from the gHRN[67].

4. Discussion

In this work, transcription profiles of two infected and non-infected cell lines (A549 and HEp-2) and primary cells (hSAEC) of two publications were used to conduct a DGEA to determine up and downregulated genes in comparison to early/uninfected conditions. The DEGs were then used to do a GSEA, which, after reducing redundancy and categorization, revealed a strong upregulation of *immune system processes* and *morphogenesis/development* in all cells, as well as a strong downregulation of *cell cycle* and *metabolism*. A comparison of the ratios of HDFs to genes and the number of HDFs per category lead to *cytokine related pathways* and *cell death* as categories for further analysis. Regulators were then predicted for the upregulated (sub-) categories of *cell death* and *cytokine related pathways*. For *cell death*, the TFs STAT2, ATF2, TFAP2C and SIX5 were predicted, while for *cytokine related pathways* IRF3, MAFF, POU5F1 and TFAP2C were predicted.

4.1 Predicted Transcription Factors

4.1.1 Cytokine Pathways are regulated by MAF BZIP Transcription Factor F

The MAF BZIP Transcription Factor F gene (MAFF) was predicted most often across all regulatory models of the cytokine pathways. It encodes the TF Maf F (Maff), which is a small Maf protein. Maff is a basic leucine zipper TF[131]. It lacks a transactivating domain. On its own, it acts as a repressor[132]. Ibrahim *et al.*[133] found that Maff is an HRF for the Hepatitis B virus (HBV). They also state that Maff is induced by the TNF, this suggests an activating relation between TNF and MAFF in the network shown above, which confirms the strong correlation between the activity of Maff and the expression of TNF (PCC: 0.964). In the network, it is seen, that Maff is predicted to regulate different interleukins (IL). These predicted regulations are intercations with the interleukins IL-6, IL-36B, IL-15. Interleukin-1 β (IL-1 β) was shown elsewhere to induce Maff expression[133]. The expression of IL-1 β is negative correlated (PCC: -0.512) with the activity of MafF. Interestingly, Ibrahim *et al.*[133] state that Maff competitively inhibits Hepatocyte Nuclear Factor 4 α (HNF-4 α) binding to an HBV enhancer sequence. Due to HBV being a DNA virus and HNF-4 α binds to DNA, has this limited application to RSV, which is an RNA virus. However, Ibrahim *et al.*[133] show that Maff plays a central role in the

regulation of HBV suggesting competitiveness with HNF-4 α . HBV evades the immune response leading to minimal induction of the interferon system[133]. As stated above and can be seen in Table 6, Maff may repress several interleukins and the IL-6 receptor, possibly leading to a dampening of the immune response of the host. It seems that Maff has a role in the regulation of pathways related to cytokine signaling and cytokine production. A reason for the appearance of Maff could be, that the expression of TNF and IL1- β strongly correlate with the expression of Maff. Both, TNF and IL1- β , are crucial factors in the immune response system.

4.1.2 Cytokine Pathways are regulated by Interferon Regulatory Factor 3

The TF IRF3 is part of the Interferon-regulatory protein family. It is found in an inactive form in the cytoplasm until it is transported in a phosphorylated bound complex with the CREB Binding Protein (CREBBP) to the nucleus[134]. As can be seen in Table 7 and in the literature, IRF3 is a key transcriptional regulator for cytokines. It is responsible for the type I IFN-dependent immune response. Thus, and through the activation of many more cytokines and chemokines, IRF-3 is a critical part of the cytokine pathway category, as predicted by the regulatory network model. Its activity (Figure 15) indicates a strong immune response of the host. As mentioned in 1.2.1 the NS1 and NS2 proteins of RSV suppress the transport of IRF3 into the nucleus, thus reducing its activity[19]. Ren *et al.*[135] showed that NS1 alone inhibits IFN- β synthesis by binding to the promoter site of IFN- β competitively to IRF-3. They also showed, as did Spann *et al.*, binding of NS1 to IRF-3 and to CREBBP interferes with the binding between IRF-3 and CREBBP, thus inhibiting its translocation to the nucleus[19]. The further upregulation of IRF-3 seems to be a promising step to reduce the viral burden on the host.

4.1.3 Cytokine Pathways are regulated by POU Class 5 Homeobox 1

The TF encoded by the POU class 5 homeobox 1 gene (POU5F1) contains a POU homeodomain[136]. It is mostly associated with embryonic development and stem cell pluripotency. In line with this association, upregulation of POU5F1 in adults is associated with tumorigenesis. The cytokine effects of POU5F1 are shown to regulate the cancer stem cell-like properties of colorectal cancer[137] and in lung cancer[138]. In colorectal cancer cells, POU5F1 regulates the enhanced cancer stem cell-like properties through the cytokines IL-8 and IL-32. Upregulation of IL-32 in lung cancer cells can be an indicator for the progression of cancer towards a metastatic phenotype. IL-32 is a pro-inflammatory cytokine, capable of inducing other cytokines, i.e. TNF- α and IL-1 β . POU5F1 induced cytokine signaling. It also recruits M2 macrophages through Macrophage colony-stimulating factor (M-CSF) and through inhibition of the Macrophage Migration Inhibition Factor (MIF). POU5F1 was also predicted by our models to regulate vascular endothelial growth

factors (VEGFs). VEGFs are reported to be upregulated by RSV to induce permeability of the bronchial airway epithelial monolayer[139]. Regulation of VEGFC through POU5F1 in mice is documented by multiple studies[140][141][142][143]. The only proof of interactions in human cells is a study by Li *et al.* from 2017[144]. However, expasy computed a POU5F1 binding site 858 base pairs (bp) downstream of the transcription starting site (TSS)[145]. The activity of POU5F1 (Figure 15) is strongly correlated with the expression of VEGFC (PCC: 0.984). The high ranking of POU5F1 could be because two of the three data sets used in this analysis were cancer cell lines. Expression of POU5F1 itself does not result in cancerous cells, however high expression can indicate cancerous cells. As expected, the expression profiles affirm this claim. In the A549 and HEp-2 cell lines POU5F1 was upregulated. In the hSAEC cells POU5F1 was not differentially expressed, but it was also predicted by our models for hSAEC.

4.1.4 Cytokine Pathways and Cell Death are predicted to be regulated by Transcription Factor AP-2 γ

The main responsibilities of the Transcription Factor AP-2 γ (TFAP2C), the gene product (the actual TF) is AP-2, are in the development of body walls, eyes, face, limbs and neural tube[146]. The Literature as well as the models show that the AP2 family of TFs regulates VEGFs too. They regulate VEGFs through competitive binding to the promoter regions against the SP3 TF[147], thus inhibiting the expression of VEGFs. AP2 has been shown to regulates VEGFs indirectly through the suppression of X-linked inhibitor of apoptosis (XIAP)-associated factor 1 (XAF1)[148]. XAF1 has been shown to suppress VEGFs[149]. XAF1 also induces autophagy via the induction of Beclin1[150]. As reported in section 1.2.1, NS2 also induces autophagy through the impairment of the ISGylation of Beclin1. Besides this, TFAP2C upregulates IL-24 through TF binding sites near the TSS[145][151]. AP2 is not only involved in cell death through the induction of autophagy but also through the regulation of several other factors. BIK and Death Effector 2 (DEDD2) are two pro-apoptosis TGs[152][153]. They are also predicted TGs in our model. BIK belongs to the Bcl-2 family[153]. The Bcl-2 family contains a range of proteins, pro- as well as anti-apoptotic. BIK belongs to the pro-apoptosis category. The AP2 family seems to be post-transcriptionally regulated by TNF[154]. TNF regulates AP2 through the induction of caspase 6 cleavage. The models predicted a direct regulation of TNF by AP2, this suggests that there is some form of feedback loop in place. Additionally, the PCC of 0.713 between the activity of TFAP2C and DEDD2, as well as the interactions with a member of the pro-apoptotic category of the Bcl-2 family and interactions with TNF, strongly hints for AP2 to be a pro-apoptotic TF. Its activity (Figures ?? and ??) seems to drop off in the late infected time points at 24 hours post infection, suggesting that regulators later in the infection cycle may suppress TFAP2C. The downregulation of apoptosis has been observed to lead to a milieu enhancing RSV proliferation[155].

4.1.5 Cell Death is regulated by Signal Transducer and Activator of Transcription 2

The Signal Transducer and Activator of Transcription 2 (STAT2) is a member of the Signal Transducer and Activator of Transcription (STAT) family of TFs. After signaling from cytokines and growth factors, it is phosphorylated and acts as an activating TF[156]. Whelan *et al.*[157] show that the NS2 protein represses the expression of STAT2. The loss of STAT2 due to NS2 led to a reduction in viral replication. STAT2 increases the expression of IL-6[158]. In combination with STAT1 and IRF9, it forms the Interferon Stimulated Gene Factor 3 (ISGF3)[159]. ISGF3 activates TNF. In cooperation with IFN- α , STAT2 activates Tumor Necrosis Factor Superfamily member 10 (TNFSF10 or TRAIL)[160]. This interaction is also predicted in our models. TNFSF10 or TRAIL is a so-called death ligand, binding to TNFRSF10A/TRAILR1, TNFRSF10B/TRAILR2. Both these receptors are death receptors[161]. STAT1 and STAT2 repress Bcl-2, which is an inhibitor of apoptosis[162]. The predicted regulatory network[67] also shows a regulation of Caspase 8 by STAT2 (PCC of 0.67 between the activity of STAT2 and the expression of CASP8), which is an essential factor in apoptosis[163]. These positive regulations of apoptosis are in line with the activity of STAT2 in the hSAEC data with apoptosis being suppressed in RSV infected cells. RSV also targets STAT2 through the NS1 protein, which contains binding sequences for cullin 2 and elongin C. These two proteins interact with NS1 to form a ubiquitin ligase (E3 ligase). STAT2 is regulated by NS1 through degradation through the E3 ligase formed by NS1, cullin 2 and elongin C[164]. Thus, the NS1 protein directly helps to evade the type I IFN response (e.g. the regulation of IL-6, formation of ISGF3 and activation of TNFSF10).

4.1.6 Cell Death is regulated by Activating Transcription Factor 2

The Activating Transcription Factor 2 (ATF2), a member of the leucine zipper family of DNA binding proteins, is recognized as a transcription factor with a broad regulatory capacity across various genes with distinct functions[165]. Notably, ATF2 binds to the Cyclic AMP-responsive element (CRE), exerting regulatory control over cell proliferation and differentiation. The ATF2/CRE complex positively regulates TNF[166]. In our models the activity of ATF2 and the expression values of TNF have a PCC of 0.965, also suggesting a positive regulation in the models. Moreover, ATF2 binds to a CRE binding site within the Bcl-2 promoter region, with a PCC of -0.99 between ATF2 activity and BCL2 expression values suggesting a negative regulatory role or suppression of BCL2 expression. Additionally, ATF2 participates in CRE-dependent transcriptional regulation by forming heterodimers with c-Jun[165]. Van Royen *et al.*[167] showed that phosphorylation of the ATF2/c-Jun heterodimer and NF- κ B triggers translocation of these factors to the nucleus , facilitating the transcriptional activation of cytokines and chemokines, thereby playing a pivotal role in the innate immune response to RSV infection. Notably, ATF2 concurrently regulates numerous genes involved in both apoptotic and autophagic

pathways, underscoring its multifaceted role in cellular processes.

4.1.7 Cell Death is regulated by SIX Homeobox 5

Six Homeobox 5 (SIX5) is a homeodomain-containing transcription factor[168]. It is implicated in the predicted regulatory network with several TGs associated with apoptosis regulation, including BAD, BAX, DFFA, FADD, HTRA2, MLKL, and TP53[67]. The predominantly negative PCC between the activity of SIX5 and the expression values of these TGs suggests a role in negative regulation of apoptosis. Conversely, SIX5 exhibits predominantly positive PCC values with pro-autophagy genes such as AMBRA1, ATG101, ATG13, ATG14, ATG16L1, ATG4B, ATG5, ATG7, WDR45B, and ZFYVE1, indicating a potential involvement in promoting autophagy processes. Moreover, SIX5 has been shown to bind to the ATP1A1 regulatory element (ARE) binding site of the ATP1A1 promoter[169]. ATP1A1 is an RSV HDF as shown in the work of Lingemann *et al.*[59]. It was shown that ATP1A1 activation led to c-Src-kinase signaling and therefore activation of EGFR. This resulted in the formation of macropinosomes, which Lingemann *et al.*[59] identified to be essential for RSV entry into the respiratory epithelial cells. SIX5 could be a potential target for downregulation of ATP1A1.

4.2 Limitations of the Study

4.2.1 Handling of the Data Sets

The datasets utilized in this study originated from two distinct publications: Rajan *et al.*[71] used two cancer cell lines, whereas Xu *et al.*[75] infected primary cells (hSAECs). Discrepancies in the methodology between the two datasets are evident, including variations in MOIs and time points. Moreover, the data presentation methods differed: Rajan *et al.* provided raw count data for all mapped genes, while Xu *et al.* presented filtered and normalized counts. Additional were the reads mapped (and quantified) with different methods. To facilitate a meaningful comparison between the datasets, it is imperative to re-map the sequence data from the publication of Xu *et al.* using the same tools as employed for the Rajan *et al.* data set or re-map both data sets with the same methods. In an effort to encompass both cell lines and primary cells, the gene sets were combined. To enhance the prediction results more samples/data sets seem to be imminent. This observation is highlighted in Figure 13, where models tend to exhibit overfitting beyond 7 parameters. It is reasonable to expect that a larger dataset would lead to more robust predictions and mitigate overfitting in regression models with 7 or more parameters. The initial situation after the predictions of TFs by MIPRIP was not favorable, 154 predicted TFs out of 154 possible TFs. This strengthens the consideration that the *immune system process* category is highly complex. This result could also be attributed to the method of predicting TFs. The method of reducing the TFS to the final 7 TFs through the scripts by Kolte[103] was not

created for this purpose. Methods with this purpose in mind should be considered in further analysis.

4.2.2 HDFs versus Expressed HDFs

In our analysis, we focused on comparing the ratio of unique HDFs to HRFs within each category. However, this comparison only considered significantly upregulated HDFs and HRFs. We opted for this approach because we aimed to explore the potential of inducing cell reprogramming through TF knockout or downregulation. It is reasonable to assume that targeting genes with consistent expression levels for knockout or downregulation may not effectively lead to the reprogramming outcomes we would want to induce. While our selection process primarily emphasized HFs, it is important to note that they were not the direct targets. There is merit in considering a broader analysis that includes all HDFs and HRFs within a category, regardless of their expression levels. When expanding our scope to encompass more host factors the role of downregulated host factors needs to be considered. The ranking of categories could then change.

4.2.3 Current Network Model and Prediction of TFs

The network in its current state provides a starting point for analysis. Some of the used sources, which were used to compile the network, are no longer available (htri and hmCHIP databases are no longer accessible¹) or have been updated (e.g. ChEA and ChIPBase both received major updates[170][171], since releasing the network). Some of the binding informations from MetaCore™ were only proven in mice or rats and one publication proving a binding is being actively investigated by the publisher. Inaccessible data makes it very complicated to comprehend the quality of a regulation. For example, considering all the regulations for cell death by SIX5 309 interactions are predicted. Considering only those regulations for which the es value from Equation 6 is greater than 1.0, only 2 interactions are left. From the es value itself it is not clear what quality the interaction has, due to the fact that multiple combinations of databases could result in the same es value. For the TF MAFF from 241 TGs only 5 remain after filtering for es values greater than 1.0. However, filtering the network by quality of the interaction, could lead to more significant results in prediction. Additionally, the prediction of TFs was done for one gene at a time. This is however inferior to a model predicting the TFs for all genes at a time. Predicting all genes is probably too computationally demanding to be solved in a single MIP so alternatives must be found. Hörhold et al.[69] adjusted MIPRIP enabling it to predict regulations of up to five genes at a time. Applying this approach in the future may lead to more realistic results. Reducing the redundancy of the set of TFs as if they were gene sets, may also have removed some significant regulators. Since the reduction by a set of genes to regulate, does not follow any biological rules and only suffices the idea that all genes must be regulated by at least one TF. Nevertheless, given the number of TFs, a reduction was necessary.

¹As of 29.02.2024

5. Conclusion and Future Outlook

The exploration of gene regulation across various RSV-infected and non-infected cell lines and primary cells has culminated in the prediction of seven distinct TFs (ATF2, IRF3, MAFF, POU5F1, SIX5, STAT2, and TFAP2C), demonstrating significant involvement in *cell death* and *cytokine related pathways*. Our GSEA analysis, coupled with subsequent gene set categorization, has corroborated existing literature, highlighting pronounced upregulation in *morphogenesis and development*, and a potent immune response (*immune system process*). The selection of *cell death* and *cytokine pathways* for investigation was guided by the ratio of unique HDFs and HRFs, further underscoring their biological relevance. The identified regulators serve as pivotal starting points for future investigations. Subsequent studies should aim to expand the analytical scope by exploring additional categories using the described methodologies, thereby enriching our understanding of cellular regulation during RSV infection. Moreover, efforts to change the mode of MIPRIP into "multi-mode", combining all available datasets to enhance prediction accuracy, warrants consideration, although this may come along with other challenges. Adaptations to MIPRIP enabling the simultaneous prediction of multiple genes could further advance predictive capabilities. However, for comprehensive gene predictions, alternative methods employing Deep Learning may prove more suitable. The predicted regulators, both those presented here and potentially newly discovered ones, hold promise for generating in-silico reprogramming models of infected cells. These models could simulate gene expression profiles akin to control samples, paving the way for follow-up *in vitro* experiments to validate these predictions. The workflow presented in this thesis could serve as the foundation for integrative approaches, holding significant potential to unravel the intricate dynamics of RSV infection and advance therapeutic interventions.

Bibliography

- [1] European Centre for Disease Prevention and Control and Word Health Organization. *RSV Virus Expected to Add Pressure on Hospitals in Many EU/EEA Countries This Season*. European Centre for Disease Prevention and Control. Nov. 23, 2022. URL: <https://www.ecdc.europa.eu/en/news-events/rsv-virus-expected-add-pressure-hospitals-many-eueea-countries-season> (visited on 12/15/2023).
- [2] European Centre for Disease Prevention and Control. *European Respiratory Virus Surveillance Summary (ERVISS)*. Jan. 12, 2023. URL: <https://erviss.org/> (visited on 12/10/2023).
- [3] Centers for Disease Control and Prevention. *RSV-NET: Respiratory Syncytial Virus Hospitalization Surveillance Network*. Jan. 12, 2023. URL: <https://www.cdc.gov/rsv/research/rsv-net.html> (visited on 12/10/2023).
- [4] Michael N. Teng, Stephen S. Whitehead, Alison Birmingham, et al. “Recombinant Respiratory Syncytial Virus That Does Not Express the NS1 or M2-2 Protein Is Highly Attenuated and Immunogenic in Chimpanzees”. In: *Journal of Virology* 74.19 (Oct. 2000), pp. 9317–9321. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.74.19.9317-9321.2000. URL: <https://journals.asm.org/doi/10.1128/JVI.74.19.9317-9321.2000> (visited on 01/15/2024).
- [5] Dr Julian Witte, Alena Zeitler, and Jana Diekmannshemke. “DAK-KINDER-UND JUGENDREPORT 2023”. In: (2023).
- [6] Bert Rima, Peter Collins, Andrew Easton, et al. “ICTV Virus Taxonomy Profile: Pneumoviridae”. In: *Journal of General Virology* 98.12 (Dec. 1, 2017), pp. 2912–2913. ISSN: 0022-1317, 1465-2099. DOI: 10.1099/jgv.0.000959. URL: <https://www.microbiologyresearch.org/content/journal/jgv/10.1099/jgv.0.000959> (visited on 01/05/2024).

- [7] Peter L. Collins and José A. Melero. “Progress in Understanding and Controlling Respiratory Syncytial Virus: Still Crazy after All These Years”. In: *Virus Research* 162.1-2 (Dec. 2011), pp. 80–99. ISSN: 01681702. DOI: 10.1016/j.virusres.2011.09.020. URL: <https://linkinghub.elsevier.com/retrieve/pii/S016817021100373X> (visited on 01/05/2024).
- [8] Sara A. Taleb, Asmaa A. Al Thani, Khalid Al Ansari, et al. “Human Respiratory Syncytial Virus: Pathogenesis, Immune Responses, and Current Vaccine Approaches”. In: *European Journal of Clinical Microbiology & Infectious Diseases* 37.10 (Oct. 2018), pp. 1817–1827. ISSN: 0934-9723, 1435-4373. DOI: 10.1007/s10096-018-3289-4. URL: <http://link.springer.com/10.1007/s10096-018-3289-4> (visited on 01/16/2024).
- [9] Elena Margaret Thornhill and David Verhoeven. “Respiratory Syncytial Virus’s Non-structural Proteins: Masters of Interference”. In: *Frontiers in Cellular and Infection Microbiology* 10 (May 19, 2020), p. 225. ISSN: 2235-2988. DOI: 10.3389/fcimb.2020.00225. URL: <https://www.frontiersin.org/article/10.3389/fcimb.2020.00225/full> (visited on 01/14/2024).
- [10] Prabha L. Atreya, Mark E. Peeples, and Peter L. Collins. “The NS1 Protein of Human Respiratory Syncytial Virus Is a Potent Inhibitor of Minigenome Transcription and RNA Replication”. In: *Journal of Virology* 72.2 (Feb. 1998), pp. 1452–1461. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.72.2.1452-1461.1998. URL: <https://journals.asm.org/doi/10.1128/JVI.72.2.1452-1461.1998> (visited on 01/14/2024).
- [11] Michael B. Battles and Jason S. McLellan. “Respiratory Syncytial Virus Entry and How to Block It”. In: *Nature Reviews Microbiology* 17.4 (Apr. 2019), pp. 233–245. ISSN: 1740-1526, 1740-1534. DOI: 10.1038/s41579-019-0149-x. URL: <https://www.nature.com/articles/s41579-019-0149-x> (visited on 09/26/2023).
- [12] Michael N. Teng and Peter L. Collins. “Altered Growth Characteristics of Recombinant Respiratory Syncytial Viruses Which Do Not Produce NS2 Protein”. In: *Journal of Virology* 73.1 (Jan. 1999), pp. 466–473. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.73.1.466-473.1999. URL: <https://journals.asm.org/doi/10.1128/JVI.73.1.466-473.1999> (visited on 01/15/2024).
- [13] Jingjing Pei, Nina R. Beri, Angela J. Zou, et al. “Nuclear-Localized Human Respiratory Syncytial Virus NS1 Protein Modulates Host Gene Transcription”. In: *Cell Reports* 37.2 (Oct. 2021), p. 109803. ISSN: 22111247. DOI:

- 10.1016/j.celrep.2021.109803. URL: <https://linkinghub.elsevier.com/retrieve/pii/S2211124721012638> (visited on 01/14/2024).
- [14] Ian G Macara. “Transport into and out of the Nucleus”. In: *MICROBIOL. MOL. BIOL. REV.* 65 (2001).
- [15] Samer Swedan, Alla Musiyenko, and Sailen Barik. “Respiratory Syncytial Virus Nonstructural Proteins Decrease Levels of Multiple Members of the Cellular Interferon Pathways”. In: *Journal of Virology* 83.19 (Oct. 2009), pp. 9682–9693. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.00715-09. URL: <https://journals.asm.org/doi/10.1128/JVI.00715-09> (visited on 12/06/2023).
- [16] Kim Chiok, Swechha M. Pokharel, Indira Mohanty, et al. “Human Respiratory Syncytial Virus NS2 Protein Induces Autophagy by Modulating Beclin1 Protein Stabilization and ISGylation”. In: *mBio* 13.1 (Feb. 22, 2022). Ed. by Thomas Shenk, e03528-21. ISSN: 2150-7511. DOI: 10.1128/mbio.03528-21. URL: <https://journals.asm.org/doi/10.1128/mbio.03528-21> (visited on 01/14/2024).
- [17] Mengdi Zhang, Jingxian Li, Haiyan Yan, et al. “ISGylation in Innate Antiviral Immunity and Pathogen Defense Responses: A Review”. In: *Frontiers in Cell and Developmental Biology* 9 (Nov. 25, 2021), p. 788410. ISSN: 2296-634X. DOI: 10.3389/fcell.2021.788410. URL: <https://www.frontiersin.org/articles/10.3389/fcell.2021.788410/full> (visited on 03/04/2024).
- [18] Alexander Kotelkin, Igor M. Belyakov, Lijuan Yang, et al. “The NS2 Protein of Human Respiratory Syncytial Virus Suppresses the Cytotoxic T-Cell Response as a Consequence of Suppressing the Type I Interferon Response”. In: *Journal of Virology* 80.12 (June 15, 2006), pp. 5958–5967. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.00181-06. URL: <https://journals.asm.org/doi/10.1128/JVI.00181-06> (visited on 01/14/2024).
- [19] Kirsten M. Spann, Kim C. Tran, and Peter L. Collins. “Effects of Non-structural Proteins NS1 and NS2 of Human Respiratory Syncytial Virus on Interferon Regulatory Factor 3, NF- κ B, and Proinflammatory Cytokines”. In: *Journal of Virology* 79.9 (May 2005), pp. 5353–5362. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.79.9.5353-5362.2005. URL: <https://journals.asm.org/doi/10.1128/JVI.79.9.5353-5362.2005> (visited on 12/06/2023).

- [20] Ayse Agac, Sophie M. Kolbe, Martin Ludlow, et al. “Host Responses to Respiratory Syncytial Virus Infection”. In: *Viruses* 15.10 (Sept. 26, 2023), p. 1999. ISSN: 1999-4915. DOI: 10.3390/v15101999. URL: <https://www.mdpi.com/1999-4915/15/10/1999> (visited on 12/21/2023).
- [21] Andressa P. Oliveira, Fernando M. Simabuco, Rodrigo E. Tamura, et al. “Human Respiratory Syncytial Virus N, P and M Protein Interactions in HEK-293T Cells”. In: *Virus Research* 177.1 (Oct. 2013), pp. 108–112. ISSN: 01681702. DOI: 10.1016/j.virusres.2013.07.010. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0168170213002323> (visited on 01/14/2024).
- [22] Leonid Gitlin, Loralyn Benoit, Christina Song, et al. “Melanoma Differentiation-Associated Gene 5 (MDA5) Is Involved in the Innate Immune Response to Paramyxoviridae Infection In Vivo”. In: *PLoS Pathogens* 6.1 (Jan. 22, 2010). Ed. by Ralph S. Baric, e1000734. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1000734. URL: <https://dx.plos.org/10.1371/journal.ppat.1000734> (visited on 01/15/2024).
- [23] M. A. García, J. Gil, I. Ventoso, et al. “Impact of Protein Kinase PKR in Cell Biology: From Antiviral to Antiproliferative Action”. In: *Microbiology and Molecular Biology Reviews* 70.4 (Dec. 2006), pp. 1032–1060. ISSN: 1092-2172, 1098-5557. DOI: 10.1128/MMBR.00027-06. URL: <https://journals.asm.org/doi/10.1128/MMBR.00027-06> (visited on 01/15/2024).
- [24] Fernando M Simabuco, John M Asara, Manuel C Guerrero, et al. “Structural Analysis of Human Respiratory Syncytial Virus P Protein: Identification of Intrinsically Disordered Domains”. In: *Brazilian Journal of Microbiology* 42.1 (Mar. 2011), pp. 340–345. ISSN: 1517-8382. DOI: 10.1590/S1517-83822011000100043. URL: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1517-83822011000100043&lng=en&nrm=iso&tlang=en (visited on 01/14/2024).
- [25] Bin Lu, Chien-Hui Ma, Robert Brazas, et al. “The Major Phosphorylation Sites of the Respiratory Syncytial Virus Phosphoprotein Are Dispensable for Virus Replication In Vitro”. In: *Journal of Virology* 76.21 (Nov. 2002), pp. 10776–10784. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.76.21.10776-10784.2002. URL: <https://journals.asm.org/doi/10.1128/JVI.76.21.10776-10784.2002> (visited on 01/16/2024).

BIBLIOGRAPHY

- [26] Choi-Lai Tiong-Yip, Lisa Aschenbrenner, Kenneth D. Johnson, et al. “Characterization of a Respiratory Syncytial Virus L Protein Inhibitor”. In: *Antimicrobial Agents and Chemotherapy* 58.7 (July 2014), pp. 3867–3873. ISSN: 0066-4804, 1098-6596. DOI: 10.1128/AAC.02540-14. URL: <https://journals.asm.org/doi/10.1128/AAC.02540-14> (visited on 01/14/2024).
- [27] Benjamin Morin, Philip J Kranzusch, Amal A Rahmeh, et al. “The Polymerase of Negative-Stranded RNA Viruses”. In: *Current Opinion in Virology* 3.2 (Apr. 2013), pp. 103–110. ISSN: 18796257. DOI: 10.1016/j.coviro.2013.03.008. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1879625713000321> (visited on 01/16/2024).
- [28] Rachel Fearn and Peter L. Collins. “Model for Polymerase Access to the Overlapped L Gene of Respiratory Syncytial Virus”. In: *Journal of Virology* 73.1 (Jan. 1999), pp. 388–397. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.73.1.388-397.1999. URL: <https://journals.asm.org/doi/10.1128/JVI.73.1.388-397.1999> (visited on 01/16/2024).
- [29] Alison Bermingham and Peter L. Collins. “The M2-2 Protein of Human Respiratory Syncytial Virus Is a Regulatory Factor Involved in the Balance between RNA Replication and Transcription”. In: *Proceedings of the National Academy of Sciences* 96.20 (Sept. 28, 1999), pp. 11259–11264. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.96.20.11259. URL: <https://www.pnas.org/doi/full/10.1073/pnas.96.20.11259> (visited on 01/14/2024).
- [30] R. Ghildyal, J. Mills, M. Murray, et al. “Respiratory Syncytial Virus Matrix Protein Associates with Nucleocapsids in Infected Cells”. In: *Journal of General Virology* 83.4 (Apr. 1, 2002), pp. 753–757. ISSN: 0022-1317, 1465-2099. DOI: 10.1099/0022-1317-83-4-753. URL: <https://www.microbiologyresearch.org/content/journal/jgv/10.1099/0022-1317-83-4-753> (visited on 12/21/2023).
- [31] Molly R. Braun, Sarah L. Noton, Emmeline L. Blanchard, et al. “Respiratory Syncytial Virus M2-1 Protein Associates Non-Specifically with Viral Messenger RNA and with Specific Cellular Messenger RNA Transcripts”. In: *PLOS Pathogens* 17.5 (May 18, 2021). Ed. by Carolina B. Lopez, e1009589. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1009589. URL: <https://dx.plos.org/10.1371/journal.ppat.1009589> (visited on 01/14/2024).
- [32] Thi-Lan Tran, Nathalie Castagné, Virginie Dubosclard, et al. “The Respiratory Syncytial Virus M2-1 Protein Forms Tetramers and Interacts with RNA and P in a Competitive Manner”. In: *Journal of Virology* 83.13 (July 2009),

- pp. 6363–6374. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.00335-09. URL: <https://journals.asm.org/doi/10.1128/JVI.00335-09> (visited on 01/14/2024).
- [33] Orlando Bonito Scudero, Verônica Feijoli Santiago, Giuseppe Palmisano, et al. “The Respiratory Syncytial Virus M2-2 Protein Is Targeted for Proteasome Degradation and Inhibits Translation and Stress Granules Assembly”. In: *PLOS ONE* 18.7 (July 25, 2023). Ed. by Boyan Grigorov, e0289100. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0289100. URL: <https://dx.plos.org/10.1371/journal.pone.0289100> (visited on 01/14/2024).
- [34] Shadi Shahriari, Ke-jun Wei, and Reena Ghildyal. “Respiratory Syncytial Virus Matrix (M) Protein Interacts with Actin In Vitro and in Cell Culture”. In: *Viruses* 10.10 (Sept. 30, 2018), p. 535. ISSN: 1999-4915. DOI: 10.3390/v10100535. URL: <http://www.mdpi.com/1999-4915/10/10/535> (visited on 01/14/2024).
- [35] Reena Ghildyal, Adeline Ho, Kylie M. Wagstaff, et al. “Nuclear Import of the Respiratory Syncytial Virus Matrix Protein Is Mediated By Importin β 1 Independent of Importin α ”. In: *Biochemistry* 44.38 (Sept. 1, 2005), pp. 12887–12895. ISSN: 0006-2960, 1520-4995. DOI: 10.1021/bi050701e. URL: <https://pubs.acs.org/doi/10.1021/bi050701e> (visited on 12/21/2023).
- [36] Jingjing Cao, Menghan Shi, Lina Zhu, et al. “The Matrix Protein of Respiratory Syncytial Virus Suppresses Interferon Signaling via RACK1 Association”. In: *Journal of Virology* 97.10 (Oct. 31, 2023). Ed. by Bryan R. G. Williams, e00747–23. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/jvi.00747-23. URL: <https://journals.asm.org/doi/10.1128/jvi.00747-23> (visited on 01/14/2024).
- [37] Larry Anderson, Samadhan Jadhao, Clinton Paden, et al. “Functional Features of the Respiratory Syncytial Virus G Protein”. In: *Viruses* 13.7 (July 1, 2021), p. 1214. ISSN: 1999-4915. DOI: 10.3390/v13071214. URL: <https://www.mdpi.com/1999-4915/13/7/1214> (visited on 01/14/2024).
- [38] Sara M. Johnson, Beth A. McNally, Ioannis Ioannidis, et al. “Respiratory Syncytial Virus Uses CX3CR1 as a Receptor on Primary Human Airway Epithelial Cultures”. In: *PLOS Pathogens* 11.12 (Dec. 11, 2015). Ed. by Mark T. Heise, e1005318. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1005318. URL: <https://dx.plos.org/10.1371/journal.ppat.1005318> (visited on 01/17/2024).

BIBLIOGRAPHY

- [39] Tatiana Chirkova, Songbai Lin, Antonius G. P. Oomens, et al. “CX3CR1 Is an Important Surface Molecule for Respiratory Syncytial Virus Infection in Human Airway Epithelial Cells”. In: *Journal of General Virology* 96.9 (Sept. 1, 2015), pp. 2543–2556. ISSN: 0022-1317, 1465-2099. DOI: 10.1099/vir.0.000218. URL: <https://www.microbiologyresearch.org/content/journal/jgv/10.1099/vir.0.000218> (visited on 01/17/2024).
- [40] Steven A. Feldman, R. Michael Hendry, and Judy A. Beeler. “Identification of a Linear Heparin Binding Domain for Human Respiratory Syncytial Virus Attachment Glycoprotein G”. In: *Journal of Virology* 73.8 (Aug. 1999), pp. 6610–6617. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.73.8.6610-6617.1999. URL: <https://journals.asm.org/doi/10.1128/JVI.73.8.6610-6617.1999> (visited on 01/17/2024).
- [41] Ralph A. Tripp, Ultan F. Power, Peter J. M. Openshaw, et al. “Respiratory Syncytial Virus: Targeting the G Protein Provides a New Approach for an Old Problem”. In: *Journal of Virology* 92.3 (Feb. 2018). Ed. by Carolyn B. Coyne, e01302–17. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.01302-17. URL: <https://journals.asm.org/doi/10.1128/JVI.01302-17> (visited on 01/14/2024).
- [42] Steven A. Feldman, Susette Audet, and Judy A. Beeler. “The Fusion Glycoprotein of Human Respiratory Syncytial Virus Facilitates Virus Attachment and Infectivity via an Interaction with Cellular Heparan Sulfate”. In: *Journal of Virology* 74.14 (July 15, 2000), pp. 6442–6447. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.74.14.6442-6447.2000. URL: <https://journals.asm.org/doi/10.1128/JVI.74.14.6442-6447.2000> (visited on 01/14/2024).
- [43] Wanderson Rezende, Hadley E. Neal, Rebecca E. Dutch, et al. “The RSV F P27 Peptide: Current Knowledge, Important Questions”. In: *Frontiers in Microbiology* 14 (June 21, 2023), p. 1219846. ISSN: 1664-302X. DOI: 10.3389/fmicb.2023.1219846. URL: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1219846/full> (visited on 12/05/2023).
- [44] Farnoosh Tayyari. “Identification of Nucleolin as a Cellular Receptor for Human Respiratory Syncytial Virus”. In: *nature medicine* 17.9 (2011).
- [45] Sandra Fuentes, Kim C. Tran, Priya Luthra, et al. “Function of the Respiratory Syncytial Virus Small Hydrophobic Protein”. In: *Journal of Virology* 81.15 (Aug. 2007), pp. 8361–8366. ISSN: 0022-538X, 1098-5514. DOI:

BIBLIOGRAPHY

- 10.1128/JVI.02717-06. URL: <https://journals.asm.org/doi/10.1128/JVI.02717-06> (visited on 01/14/2024).
- [46] Siok-Wan Gan, Edward Tan, Xin Lin, et al. “The Small Hydrophobic Protein of the Human Respiratory Syncytial Virus Forms Pentameric Ion Channels”. In: *Journal of Biological Chemistry* 287.29 (July 2012), pp. 24671–24689. ISSN: 00219258. DOI: 10.1074/jbc.M111.332791. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0021925820433198> (visited on 01/14/2024).
- [47] Tra Nguyen Huong, Laxmi Ravi Iyer, Jing Lui, et al. “The Respiratory Syncytial Virus SH Protein Is Incorporated into Infectious Virus Particles That Form on Virus-Infected Cells”. In: *Virology* 580 (Mar. 2023), pp. 28–40. ISSN: 00426822. DOI: 10.1016/j.virol.2023.01.013. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0042682223000132> (visited on 01/14/2024).
- [48] Lesley J. Scott and Harriet M. Lamb. “Palivizumab”. In: *Drugs* 58.2 (Aug. 1, 1999), pp. 305–311. ISSN: 1179-1950. DOI: 10.2165/00003495-199958020-00009. URL: <https://doi.org/10.2165/00003495-199958020-00009>.
- [49] Mary T. Caserta, Sean T. O’Leary, Flor M. Munoz, et al. “Palivizumab Prophylaxis in Infants and Young Children at Increased Risk of Hospitalization for Respiratory Syncytial Virus Infection”. In: *Pediatrics* 152.1 (July 1, 2023), e2023061803. ISSN: 0031-4005, 1098-4275. DOI: 10.1542/peds.2023-061803. URL: <https://publications.aap.org/pediatrics/article/152/1/e2023061803/192153/Palivizumab-Prophylaxis-in-Infants-and-Young> (visited on 01/22/2024).
- [50] Bernhard Resch. “Product Review on the Monoclonal Antibody Palivizumab for Prevention of Respiratory Syncytial Virus Infection”. In: *Human Vaccines & Immunotherapeutics* 13.9 (Sept. 2, 2017), pp. 2138–2149. ISSN: 2164-5515, 2164-554X. DOI: 10.1080/21645515.2017.1337614. URL: <https://www.tandfonline.com/doi/full/10.1080/21645515.2017.1337614> (visited on 01/18/2024).
- [51] M. Pamela Griffin, Yuan Yuan, Therese Takas, et al. “Single-Dose Nirsevimab for Prevention of RSV in Preterm Infants”. In: *New England Journal of Medicine* 383.5 (July 30, 2020), pp. 415–425. ISSN: 0028-4793, 1533-4406. DOI: 10.1056/NEJMoa1913556. URL: <http://www.nejm.org/doi/10.1056/NEJMoa1913556> (visited on 01/18/2024).

BIBLIOGRAPHY

- [52] Qing Zhu, Jason S. McLellan, Nicole L. Kallewaard, et al. “A Highly Potent Extended Half-Life Antibody as a Potential RSV Vaccine Surrogate for All Infants”. In: *Science Translational Medicine* 9.388 (May 3, 2017), eaaj1928. ISSN: 1946-6234, 1946-6242. DOI: 10.1126/scitranslmed.aaj1928. URL: <https://www.science.org/doi/10.1126/scitranslmed.aaj1928> (visited on 01/18/2024).
- [53] Laura L. Hammitt, Ron Dagan, Yuan Yuan, et al. “Nirsevimab for Prevention of RSV in Healthy Late-Premature and Term Infants”. In: *New England Journal of Medicine* 386.9 (Mar. 3, 2022), pp. 837–846. ISSN: 0028-4793, 1533-4406. DOI: 10.1056/NEJMoa2110275. URL: <http://www.nejm.org/doi/10.1056/NEJMoa2110275> (visited on 01/18/2024).
- [54] Susan J. Keam. “Nirsevimab: First Approval”. In: *Drugs* 83.2 (Feb. 2023), pp. 181–187. ISSN: 0012-6667, 1179-1950. DOI: 10.1007/s40265-022-01829-6. URL: <https://link.springer.com/10.1007/s40265-022-01829-6> (visited on 01/22/2024).
- [55] Food & Drug Administration. FDA Approves New Drug to Prevent RSV in Babies and Toddlers. July 17, 2023. URL: <https://www.fda.gov/news-events/press-announcements/fda-approves-new-drug-prevent-rsv-babies-and-toddlers> (visited on 01/18/2024).
- [56] Tiffany Turner, Benjamin Kopp, Grace Paul, et al. “Respiratory Syncytial Virus: Current and Emerging Treatment Options”. In: *ClinicoEconomics and Outcomes Research* (Apr. 2014), p. 217. ISSN: 1178-6981. DOI: 10.2147/CEOR.S60710. URL: <http://www.dovepress.com/respiratory-syncytial-virus-current-and-emerging-treatment-options-peer-reviewed-article-CEOR> (visited on 01/18/2024).
- [57] Xanthippi Topalidou, Alexis M. Kalergis, and Georgios Papazisis. “Respiratory Syncytial Virus Vaccines: A Review of the Candidates and the Approved Vaccines”. In: *Pathogens* 12.10 (Oct. 19, 2023), p. 1259. ISSN: 2076-0817. DOI: 10.3390/pathogens12101259. URL: <https://www.mdpi.com/2076-0817/12/10/1259> (visited on 01/19/2024).
- [58] H.-Heinrich Hoffmann, William M. Schneider, Vincent A. Blomen, et al. “Diverse Viruses Require the Calcium Transporter SPCA1 for Maturation and Spread”. In: *Cell Host & Microbe* 22.4 (Oct. 2017), 460–470.e5. ISSN: 19313128. DOI: 10.1016/j.chom.2017.09.002. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1931312817303906> (visited on 01/10/2024).

- [59] Matthias Lingemann, Thomas McCarty, Xueqiao Liu, et al. “The Alpha-1 Subunit of the Na⁺,K⁺-ATPase (ATP1A1) Is Required for Macropinocytic Entry of Respiratory Syncytial Virus (RSV) in Human Respiratory Epithelial Cells”. In: *PLOS Pathogens* 15.8 (Aug. 5, 2019). Ed. by Sean P. J. Whelan, e1007963. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1007963. URL: <https://dx.plos.org/10.1371/journal.ppat.1007963> (visited on 01/10/2024).
- [60] Ziheng Feng, Lili Xu, and Zhengde Xie. “Receptors for Respiratory Syncytial Virus Infection and Host Factors Regulating the Life Cycle of Respiratory Syncytial Virus”. In: *Frontiers in Cellular and Infection Microbiology* 12 (Feb. 25, 2022), p. 858629. ISSN: 2235-2988. DOI: 10.3389/fcimb.2022.858629. URL: <https://www.frontiersin.org/articles/10.3389/fcimb.2022.858629/full> (visited on 12/21/2023).
- [61] Yuji Zhang. “Gene Regulatory Networks: Real Data Sources and Their Analysis”. In: *Evolutionary Computation in Gene Regulatory Network Research*. John Wiley & Sons, Ltd, 2016, pp. 49–65. ISBN: 978-1-119-07945-3. DOI: 10.1002/9781119079453.ch3. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119079453.ch3>.
- [62] Adam A Margolin, Ilya Nemenman, Katia Basso, et al. “ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context”. In: *BMC Bioinformatics* 7.S1 (Mar. 2006), S7. ISSN: 1471-2105. DOI: 10.1186/1471-2105-7-S1-S7. URL: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-7-S1-S7> (visited on 02/19/2024).
- [63] VÂN ANH HUYNH-THU, Alexandre Irthum, Louis Wehenkel, et al. “Inferring Regulatory Networks from Expression Data Using Tree-Based Methods”. In: *PLoS ONE* 5.9 (Sept. 28, 2010). Ed. by Mark Isalan, e12776. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0012776. URL: <https://dx.plos.org/10.1371/journal.pone.0012776> (visited on 02/19/2024).
- [64] Sara Aibar, Carmen Bravo González-Blas, Thomas Moerman, et al. “SCENIC: Single-Cell Regulatory Network Inference and Clustering”. In: *Nature Methods* 14.11 (Nov. 2017), pp. 1083–1086. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.4463. URL: <https://www.nature.com/articles/nmeth.4463> (visited on 02/19/2024).

- [65] Piotr J. Balwierz, Mikhail Pachkov, Phil Arnold, et al. “ISMARA: Automated Modeling of Genomic Signals as a Democracy of Regulatory Motifs”. In: *Genome Research* 24.5 (May 2014), pp. 869–884. ISSN: 1088-9051. DOI: 10.1101/gr.169508.113. URL: <http://genome.cshlp.org/lookup/doi/10.1101/gr.169508.113> (visited on 01/21/2024).
- [66] The FANTOM Consortium and Riken Omics Science Center. “The Transcriptional Network That Controls Growth Arrest and Differentiation in a Human Myeloid Leukemia Cell Line”. In: *Nature Genetics* 41.5 (May 2009), pp. 553–562. ISSN: 1061-4036, 1546-1718. DOI: 10.1038/ng.375. URL: <https://www.nature.com/articles/ng.375> (visited on 01/22/2024).
- [67] Alexandra M. Poos, Theresa Kordaß, Amol Kolte, et al. “Modelling TERT Regulation across 19 Different Cancer Types Based on the MIPRIP 2.0 Gene Regulatory Network Approach”. In: *BMC Bioinformatics* 20.1 (Dec. 2019), p. 737. ISSN: 1471-2105. DOI: 10.1186/s12859-019-3323-2. URL: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-3323-2> (visited on 01/21/2024).
- [68] Alexandra M. Poos, André Maicher, Anna K. Dieckmann, et al. “Mixed Integer Linear Programming Based Machine Learning Approach Identifies Regulators of Telomerase in Yeast”. In: *Nucleic Acids Research* 44.10 (June 2, 2016), e93–e93. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkw111. URL: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw111> (visited on 06/08/2024).
- [69] Franziska Hörhold, David Eisel, Marcus Oswald, et al. “Reprogramming of Macrophages Employing Gene Regulatory and Metabolic Network Models”. In: *PLOS Computational Biology* 16.2 (Feb. 25, 2020). Ed. by Anders Wallqvist, e1007657. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1007657. URL: <https://dx.plos.org/10.1371/journal.pcbi.1007657> (visited on 01/21/2024).
- [70] R Edgar. “Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository”. In: *Nucleic Acids Research* 30.1 (Jan. 1, 2002), pp. 207–210. ISSN: 13624962. DOI: 10.1093/nar/30.1.207. (Visited on 01/23/2024).
- [71] Anubama Rajan, Felipe-Andrés Piedra, Letisha Aideyan, et al. “Multiple Respiratory Syncytial Virus (RSV) Strains Infecting HEp-2 and A549 Cells Reveal Cell Line-Dependent Differences in Resistance to RSV Infection”. In: *Journal of Virology* 96.7 (Apr. 13, 2022). Ed. by Mark T. Heise, e01904–21.

BIBLIOGRAPHY

- ISSN: 0022-538X, 1098-5514. DOI: 10.1128/jvi.01904-21. URL: <https://journals.asm.org/doi/10.1128/jvi.01904-21> (visited on 09/26/2023).
- [72] Donald J. Giard, Stuart A. Aaronson, George J. Todaro, et al. “In Vitro Cultivation of Human Tumors: Establishment of Cell Lines Derived From a Series of Solid Tumors2”. In: *JNCI: Journal of the National Cancer Institute* 51.5 (Nov. 1973), pp. 1417–1423. ISSN: 1460-2105, 0027-8874. DOI: 10.1093/jnci/51.5.1417. URL: <https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/51.5.1417> (visited on 01/23/2024).
- [73] T.R. Chen. “Re-Evaluation of HeLa, HeLa S3, and HEp-2 Karyotypes”. In: *Cytogenetic and Genome Research* 48.1 (1988), pp. 19–24. ISSN: 1424-8581, 1424-859X. DOI: 10.1159/000132579. URL: <https://www.karger.com/Article/FullText/132579> (visited on 01/23/2024).
- [74] ATCC. *HEp-2*. ATCC. URL: <https://www.atcc.org/products/ccl-23> (visited on 01/23/2024).
- [75] Xiaofang Xu, Dianhua Qiao, Morgan Mann, et al. “Respiratory Syncytial Virus Infection Induces Chromatin Remodeling to Activate Growth Factor and Extracellular Matrix Secretion Pathways”. In: *Viruses* 12.8 (July 26, 2020), p. 804. ISSN: 1999-4915. DOI: 10.3390/v12080804. URL: <https://www.mdpi.com/1999-4915/12/8/804> (visited on 01/23/2024).
- [76] Ruben D. Ramirez, Shelley Sheridan, Luc Girard, et al. “Immortalization of Human Bronchial Epithelial Cells in the Absence of Viral Oncoproteins”. In: *Cancer Research* 64.24 (Dec. 15, 2004), pp. 9027–9034. ISSN: 0008-5472, 1538-7445. DOI: 10.1158/0008-5472.CAN-04-3703. URL: <https://aacrjournals.org/cancerres/article/64/24/9027/512189/Immortalization-of-Human-Bronchial-Epithelial> (visited on 01/23/2024).
- [77] *BOSC-23 (CVCL_4401)*. May 10, 2023. URL: https://www.cellosaurus.org/CVCL_4401 (visited on 01/24/2024).
- [78] *HAP1 (CVCL_Y019)*. May 10, 2023. URL: https://www.cellosaurus.org/CVCL_Y019 (visited on 01/24/2024).
- [79] Yoav Benjamini and Yosef Hochberg. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 57.1 (Jan. 1995), pp. 289–300. ISSN: 0035-9246, 2517-6161. DOI: 10.1111/j.2517-6161.1995.tb02031.x. URL: <https://rss.onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1995.tb02031.x> (visited on 02/18/2024).

BIBLIOGRAPHY

- [80] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2023. URL: <https://www.R-project.org/>.
- [81] Posit Team. *RStudio: Integrated Development Environment for R*. Boston, MA: Posit Software, PBC, 2023. URL: <http://www.posit.co/>.
- [82] Hadley Wickham. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. URL: <https://ggplot2.tidyverse.org>.
- [83] Ryota Suzuki, Yoshikazu Terada, and Hidetoshi Shimodaira. *Pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling*. Version 2.2-0. 2019. URL: <https://CRAN.R-project.org/package=pvclust>.
- [84] Thomas Lin Pedersen. *Tidygraph: A Tidy API for Graph Manipulation*. Version 1.3.0. 2023. URL: <https://CRAN.R-project.org/package=tidygraph>.
- [85] Tal Galili. “Dendextend: An R Package for Visualizing, Adjusting and Comparing Trees of Hierarchical Clustering”. In: *Bioinformatics* 31.22 (Nov. 15, 2015), pp. 3718–3720. ISSN: 1367-4811, 1367-4803. DOI: 10.1093/bioinformatics/btv428. URL: <https://academic.oup.com/bioinformatics/article/31/22/3718/240978> (visited on 01/27/2024).
- [86] Yunshun Chen, Davis McCarthy, Pedro Baldoni, et al. “edgeR: Differential Analysis of Sequence Read Count Data User’s Guide”. In: (Oct. 18, 2023).
- [87] Simon Anders and Wolfgang Huber. “Differential Expression Analysis for Sequence Count Data”. In: (2010).
- [88] Mark D Robinson and Alicia Oshlack. “A Scaling Normalization Method for Differential Expression Analysis of RNA-seq Data”. In: *Genome Biology* 11.3 (2010), R25. ISSN: 1465-6906. DOI: 10.1186/gb-2010-11-3-r25. URL: <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2010-11-3-r25> (visited on 01/25/2024).
- [89] James H Bullard, Elizabeth Purdom, Kasper D Hansen, et al. “Evaluation of Statistical Methods for Normalization and Differential Expression in mRNA-Seq Experiments”. In: *BMC Bioinformatics* 11.1 (Dec. 2010), p. 94. ISSN: 1471-2105. DOI: 10.1186/1471-2105-11-94. URL: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-94> (visited on 01/25/2024).

- [90] Elie Maza, Pierre Frasse, Pavel Senin, et al. “Comparison of Normalization Methods for Differential Gene Expression Analysis in RNA-Seq Experiments: A Matter of Relative Size of Studied Transcriptomes”. In: *Communicative & Integrative Biology* 6.6 (Nov. 9, 2013), e25849. ISSN: 1942-0889. DOI: 10.4161/cib.25849. URL: <http://www.tandfonline.com/doi/abs/10.4161/cib.25849> (visited on 01/25/2024).
- [91] M.-A. Dillies, A. Rau, J. Aubert, et al. “A Comprehensive Evaluation of Normalization Methods for Illumina High-Throughput RNA Sequencing Data Analysis”. In: *Briefings in Bioinformatics* 14.6 (Nov. 1, 2013), pp. 671–683. ISSN: 1467-5463, 1477-4054. DOI: 10.1093/bib/bbs046. URL: <https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/bbs046> (visited on 01/25/2024).
- [92] Ciaran Evans, Johanna Hardin, and Daniel M Stoebel. “Selecting Between-Sample RNA-Seq Normalization Methods from the Perspective of Their Assumptions”. In: *Briefings in Bioinformatics* 19.5 (Sept. 28, 2018), pp. 776–792. ISSN: 1467-5463, 1477-4054. DOI: 10.1093/bib/bbx008. URL: <https://academic.oup.com/bib/article/19/5/776/3056951> (visited on 01/25/2024).
- [93] Koji Kadota, Tomoaki Nishiyama, and Kentaro Shimizu. “A Normalization Strategy for Comparing Tag Count Data”. In: *Algorithms for Molecular Biology* 7.1 (Dec. 2012), p. 5. ISSN: 1748-7188. DOI: 10.1186/1748-7188-7-5. URL: <https://almob.biomedcentral.com/articles/10.1186/1748-7188-7-5> (visited on 01/25/2024).
- [94] Jianqiang Sun, Tomoaki Nishiyama, Kentaro Shimizu, et al. “TCC: An R Package for Comparing Tag Count Data with Robust Normalization Strategies”. In: *BMC Bioinformatics* 14.1 (Dec. 2013), p. 219. ISSN: 1471-2105. DOI: 10.1186/1471-2105-14-219. URL: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-219> (visited on 01/25/2024).
- [95] Mark D. Robinson, Davis J. McCarthy, and Gordon K. Smyth. “edgeR : A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data”. In: *Bioinformatics* 26.1 (Jan. 1, 2010), pp. 139–140. ISSN: 1367-4811, 1367-4803. DOI: 10.1093/bioinformatics/btp616. URL: <https://academic.oup.com/bioinformatics/article/26/1/139/182458> (visited on 01/25/2024).

- [96] L Kolberg, U Raudvere, I Kuzmin, et al. “Gprofiler2 – an R Package for Gene List Functional Enrichment Analysis and Namespace Conversion Toolset g:Profiler [Version 2; Peer Review: 2 Approved]”. In: *F1000Research* 9.709 (2020). DOI: 10.12688/f1000research.24956.2.
- [97] Liis Kolberg, Uku Raudvere, Ivan Kuzmin, et al. “G:Profiler—Interoperable Web Service for Functional Enrichment Analysis and Gene Identifier Mapping (2023 Update)”. In: *Nucleic Acids Research* 51.W1 (July 5, 2023), W207–W212. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkad347. URL: <https://academic.oup.com/nar/article/51/W1/W207/7152869> (visited on 02/14/2024).
- [98] Fiona Cunningham, James E Allen, Jamie Allen, et al. “Ensembl 2022”. In: *Nucleic Acids Research* 50.D1 (Jan. 7, 2022), pp. D988–D995. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkab1049. URL: <https://academic.oup.com/nar/article/50/D1/D988/6430486> (visited on 02/14/2024).
- [99] Andrew D Yates, James Allen, Ridwan M Amode, et al. “Ensembl Genomes 2022: An Expanding Genome Resource for Non-Vertebrates”. In: *Nucleic Acids Research* 50.D1 (Jan. 7, 2022), pp. D996–D1003. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkab1007. URL: <https://academic.oup.com/nar/article/50/D1/D996/6427344> (visited on 02/14/2024).
- [100] Michael Ashburner, Catherine A. Ball, Judith A. Blake, et al. “Gene Ontology: Tool for the Unification of Biology”. In: *Nature Genetics* 25.1 (May 2000), pp. 25–29. ISSN: 1061-4036, 1546-1718. DOI: 10.1038/75556. URL: https://www.nature.com/articles/ng0500_25 (visited on 02/15/2024).
- [101] The Gene Ontology Consortium, Suzi A Aleksander, James Balhoff, et al. “The Gene Ontology Knowledgebase in 2023”. In: *Genetics* 224.1 (Mar. 2023), iyad031. ISSN: 1943-2631. DOI: 10.1093/genetics/iyad031. URL: <https://doi.org/10.1093/genetics/iyad031>.
- [102] Jüri Reimand, Meelis Kull, Hedi Peterson, et al. “G:Profiler—a Web-Based Toolset for Functional Profiling of Gene Lists from Large-Scale Experiments”. In: *Nucleic Acids Research* 35 (suppl_2 July 2007), W193–W200. ISSN: 1362-4962, 0305-1048. DOI: 10.1093/nar/gkm226. URL: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkm226> (visited on 02/14/2024).

- [103] Amol Kolte. “Improving the Outcome of the Critically Ill with Cortico- and Sex-Steroids Employing Clinico-Transcriptomics Approaches”. Jena: Friedrich-Schiller-University, 2020.
- [104] Paul Jaccard. “THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.” In: *New Phytologist* 11.2 (Feb. 1912), pp. 37–50. ISSN: 0028-646X, 1469-8137. DOI: 10.1111/j.1469-8137.1912.tb05611.x. URL: <https://nph.onlinelibrary.wiley.com/doi/10.1111/j.1469-8137.1912.tb05611.x> (visited on 02/16/2024).
- [105] LLC Optimization G. *Gurobi: Gurobi Optimizer 10.0 Interface*. 2023. URL: <https://www.gurobi.com>.
- [106] Brad Efron and R. Tibshirani. *GSA: Gene Set Analysis*. 2022. URL: <https://CRAN.R-project.org/package=GSA>.
- [107] Ingo Feinerer and Kurt Hornik. *Tm: Text Mining Package*. 2023. URL: <https://CRAN.R-project.org/package=tm>.
- [108] Ingo Feinerer, Kurt Hornik, and David Meyer. “Text Mining Infrastructure in R”. In: *Journal of Statistical Software* 25.5 (Mar. 2008), pp. 1–54. DOI: 10.18637/jss.v025.i05.
- [109] Milan Bouchet-Valat. *SnowballC: Snowball Stemmers Based on the C ‘lib-stemmer’ UTF-8 Library*. 2023. URL: <https://CRAN.R-project.org/package=SnowballC>.
- [110] Jarek Tuszynski. *caTools: Tools: Moving Window Statistics, GIF, Base64, ROC AUC, Etc.* 2021. URL: <https://CRAN.R-project.org/package=caTools>.
- [111] Andy Liaw and Matthew Wiener. “Classification and Regression by randomForest”. In: *R News* 2.3 (2002), pp. 18–22. URL: <https://CRAN.R-project.org/doc/Rnews/>.
- [112] Kuhn and Max. “Building Predictive Models in R Using the Caret Package”. In: *Journal of Statistical Software* 28.5 (2008), pp. 1–26. DOI: 10.18637/jss.v028.i05. URL: <https://www.jstatsoft.org/index.php/jss/article/view/v028i05>.
- [113] Steffi Grote. *GOfuncR: Gene Ontology Enrichment Using FUNC*. 2022. DOI: 10.18129/B9.bioc.GOfuncR. URL: <https://bioconductor.org/packages/GOfuncR>.

- [114] Daniel Greene, Sylvia Richardson, and Ernest Turro. “ontologyX: A Suite of R Packages for Working with Ontological Data”. In: *Bioinformatics* 33.7 (Apr. 1, 2017). Ed. by John Hancock, pp. 1104–1106. ISSN: 1367-4803, 1367-4811. DOI: 10.1093/bioinformatics/btw763. URL: <https://academic.oup.com/bioinformatics/article/33/7/1104/2843897> (visited on 02/17/2024).
- [115] Zuguang Gu, Lei Gu, Roland Eils, et al. “Circlize Implements and Enhances Circular Visualization in R”. In: *Bioinformatics* 30.19 (2014), pp. 2811–2812.
- [116] Wolfgang Huber, Anja Von Heydebreck, Holger Sültmann, et al. “Variance Stabilization Applied to Microarray Data Calibration and to the Quantification of Differential Expression”. In: *Bioinformatics* 18 (suppl_1 July 1, 2002), S96–S104. ISSN: 1367-4811, 1367-4803. DOI: 10.1093/bioinformatics/18.suppl_1.S96. URL: https://academic.oup.com/bioinformatics/article/18/suppl_1/S96/231881 (visited on 02/18/2024).
- [117] *Early Research Intelligence and Solutions—Metacore—Cortellis*. URL: [https://clarivate.com/cortellis/solutions/early-research-intelligence-solutions/..](https://clarivate.com/cortellis/solutions/early-research-intelligence-solutions/)
- [118] Alexander Lachmann, Huilei Xu, Jayanth Krishnan, et al. “ChEA: Transcription Factor Regulation Inferred from Integrating Genome-Wide ChIP-X Experiments”. In: *Bioinformatics* 26.19 (Oct. 1, 2010), pp. 2438–2444. ISSN: 1367-4811, 1367-4803. DOI: 10.1093/bioinformatics/btq466. URL: <https://academic.oup.com/bioinformatics/article/26/19/2438/230298> (visited on 02/17/2024).
- [119] Yunhai Luo, Benjamin C Hitz, Idan Gabdank, et al. “New Developments on the Encyclopedia of DNA Elements (ENCODE) Data Portal”. In: *Nucleic Acids Research* 48.D1 (Jan. 8, 2020), pp. D882–D889. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkz1062. URL: <https://academic.oup.com/nar/article/48/D1/D882/5622708> (visited on 02/17/2024).
- [120] Li Chen, George Wu, and Hongkai Ji. “hmChIP: A Database and Web Server for Exploring Publicly Available Human and Mouse ChIP-seq and ChIP-chip Data”. In: *Bioinformatics* 27.10 (May 15, 2011), pp. 1447–1448. ISSN: 1367-4811, 1367-4803. DOI: 10.1093/bioinformatics/btr156. URL: <https://academic.oup.com/bioinformatics/article/27/10/1447/260530> (visited on 02/17/2024).

- [121] Luiz A Bovolenta, Marcio L Acencio, and Ney Lemke. “HTRIdb: An Open-Access Database for Experimentally Verified Human Transcriptional Regulation Interactions”. In: (2012).
- [122] Jian-Hua Yang, Jun-Hao Li, Shan Jiang, et al. “ChIPBase: A Database for Decoding the Transcriptional Regulation of Long Non-Coding RNA and microRNA Genes from ChIP-Seq Data”. In: *Nucleic Acids Research* 41.D1 (Jan. 1, 2013), pp. D177–D187. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gks1060. URL: <http://academic.oup.com/nar/article/41/D1/D177/1054545/ChIPBase-a-database-for-decoding-the> (visited on 02/17/2024).
- [123] Elena Grassi, Ettore Zapparoli, Ivan Molineris, et al. “Total Binding Affinity Profiles of Regulatory Regions Predict Transcription Factor Binding and Gene Expression in Human Cells”. In: *PLOS ONE* 10.11 (Nov. 24, 2015). Ed. by Stein Aerts, e0143627. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0143627. URL: <https://dx.plos.org/10.1371/journal.pone.0143627> (visited on 02/17/2024).
- [124] Antony W Burgess. “Growth Factors and Cytokines”. In: 1.2 (2015).
- [125] MJ Cameron and DJ Kelvin. *Cytokines, Chemokines and Their Receptors*. Madame Curie Bioscience Database. 2013. URL: <https://www.ncbi.nlm.nih.gov/books/NBK6294/> (visited on 04/22/2024).
- [126] Susan Elmore. “Apoptosis: A Review of Programmed Cell Death”. In: *Toxicologic Pathology* 35.4 (June 2007), pp. 495–516. ISSN: 0192-6233, 1533-1601. DOI: 10.1080/01926230701320337. URL: <http://journals.sagepub.com/doi/10.1080/01926230701320337> (visited on 04/24/2024).
- [127] Mark S D’Arcy. “Cell Death: A Review of the Major Forms of Apoptosis, Necrosis and Autophagy”. In: *Cell Biology International* 43.6 (June 2019), pp. 582–592. ISSN: 1065-6995, 1095-8355. DOI: 10.1002/cbin.11137. URL: <https://onlinelibrary.wiley.com/doi/10.1002/cbin.11137> (visited on 04/24/2024).
- [128] Weiya Cao, Jinhong Li, Kepeng Yang, et al. “An Overview of Autophagy: Mechanism, Regulation and Research Progress”. In: *Bulletin du Cancer* 108.3 (Mar. 2021), pp. 304–322. ISSN: 00074551. DOI: 10.1016/j.bulcan.2020.11.004. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0007455120305014> (visited on 04/25/2024).

BIBLIOGRAPHY

- [129] Noboru Mizushima. “Autophagy: Process and Function”. In: *Genes & Development* 21.22 (Nov. 15, 2007), pp. 2861–2873. ISSN: 0890-9369, 1549-5477. DOI: 10.1101/gad.1599207. URL: <http://genesdev.cshlp.org/lookup/doi/10.1101/gad.1599207> (visited on 04/24/2024).
- [130] Danielle Glick, Sandra Barth, and Kay F. Macleod. “Autophagy: Cellular and Molecular Mechanisms”. In: *The Journal of Pathology* 221.1 (May 2010), pp. 3–12. ISSN: 0022-3417, 1096-9896. DOI: 10.1002/path.2697. URL: <https://pathsocjournals.onlinelibrary.wiley.com/doi/10.1002/path.2697> (visited on 04/24/2024).
- [131] National Center for Biotechnology Information (US). *MAFF MAF bZIP Transcription Factor F [Homo Sapiens (Human)]*. Jan. 26, 2024. URL: <https://www.ncbi.nlm.nih.gov/gene/23764#gene-expression>.
- [132] O Johnsen. “Small Maf Proteins Interact with the Human Transcription Factor TCF11/Nrf1/LCR-F1”. In: *Nucleic Acids Research* 24.21 (Nov. 1, 1996), pp. 4289–4297. ISSN: 13624962. DOI: 10.1093/nar/24.21.4289. URL: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/24.21.4289> (visited on 02/28/2024).
- [133] Marwa K. Ibrahim, Tawfeek H. Abdelhafez, Junko S. Takeuchi, et al. “Maff Is an Antiviral Host Factor That Suppresses Transcription from Hepatitis B Virus Core Promoter”. In: *Journal of Virology* 95.15 (July 12, 2021). Ed. by J.-H. James Ou, e00767–21. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.00767-21. URL: <https://journals.asm.org/doi/10.1128/JVI.00767-21> (visited on 02/28/2024).
- [134] National Center for Biotechnology Information (US). *IRF3 Interferon Regulatory Factor 3 [Homo Sapiens (Human)]*. May 2, 2024. URL: <https://www.ncbi.nlm.nih.gov/gene/3661#summary>.
- [135] Junping Ren, Tianshuang Liu, Lan Pang, et al. “A Novel Mechanism for the Inhibition of Interferon Regulatory Factor-3-Dependent Gene Expression by Human Respiratory Syncytial Virus NS1 Protein”. In: *Journal of General Virology* 92.9 (Sept. 1, 2011), pp. 2153–2159. ISSN: 0022-1317, 1465-2099. DOI: 10.1099/vir.0.032987-0. URL: <https://www.microbiologyresearch.org/content/journal/jgv/10.1099/vir.0.032987-0> (visited on 02/28/2024).

BIBLIOGRAPHY

- [136] National Center for Biotechnology Information (US). *POU5F1 POU Class 5 Homeobox 1 [Homo Sapiens (Human)]*. Feb. 19, 2024. URL: <https://www.ncbi.nlm.nih.gov/gene/5460#summary>.
- [137] Charn-Jung Chang, Yueh Chien, Kai-Hsi Lu, et al. “Oct4-Related Cytokine Effects Regulate Tumorigenic Properties of Colorectal Cancer Cells”. In: *Biochemical and Biophysical Research Communications* 415.2 (Nov. 2011), pp. 245–251. ISSN: 0006291X. DOI: 10.1016/j.bbrc.2011.10.024. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0006291X11018213> (visited on 02/28/2024).
- [138] Chia-Sing Lu, Ai-Li Shiao, Bing-Hua Su, et al. “Oct4 Promotes M2 Macrophage Polarization through Upregulation of Macrophage Colony-Stimulating Factor in Lung Cancer”. In: *Journal of Hematology & Oncology* 13.1 (Dec. 2020), p. 62. ISSN: 1756-8722. DOI: 10.1186/s13045-020-00887-1. URL: <https://jhoonline.biomedcentral.com/articles/10.1186/s13045-020-00887-1> (visited on 02/28/2024).
- [139] Muna M. Kilani, Kamal A. Mohammed, Najmunnisa Nasreen, et al. “Respiratory Syncytial Virus Causes Increased Bronchial Epithelial Permeability”. In: *Chest* 126.1 (July 2004), pp. 186–191. ISSN: 00123692. DOI: 10.1378/chest.126.1.186. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0012369215329135> (visited on 03/03/2024).
- [140] Arvind Shakya, Catherine Callister, Alon Goren, et al. “Pluripotency Transcription Factor Oct4 Mediates Stepwise Nucleosome Demethylation and Depletion”. In: *Molecular and Cellular Biology* 35.6 (Mar. 1, 2015), pp. 1014–1025. ISSN: 1098-5549. DOI: 10.1128/MCB.01105-14. URL: <https://www.tandfonline.com/doi/full/10.1128/MCB.01105-14> (visited on 03/03/2024).
- [141] Xaosong Liu, Jinyan Huang, Taotao Chen, et al. “Yamanaka Factors Critically Regulate the Developmental Signaling Network in Mouse Embryonic Stem Cells”. In: *Cell Research* 18.12 (Dec. 2008), pp. 1177–1189. ISSN: 1001-0602, 1748-7838. DOI: 10.1038/cr.2008.309. URL: <https://www.nature.com/articles/cr2008309> (visited on 03/03/2024).
- [142] Steve Bilodeau, Michael H. Kagey, Garrett M. Frampton, et al. “SetDB1 Contributes to Repression of Genes Encoding Developmental Regulators and Maintenance of ES Cell State”. In: *Genes & Development* 23.21 (Nov. 1, 2009), pp. 2484–2489. ISSN: 0890-9369, 1549-5477. DOI: 10.1101/gad.

1837309. URL: <http://genesdev.cshlp.org/lookup/doi/10.1101/gad.1837309> (visited on 03/03/2024).
- [143] Alexei A Sharov, Shinji Masui, Lioudmila V Sharova, et al. “Identification of Pou5f1, Sox2, and Nanog Downstream Target Genes with Statistical Confidence by Applying a Novel Algorithm to Time Course Microarray and Genome-Wide Chromatin Immunoprecipitation Data”. In: *BMC Genomics* 9.1 (Dec. 2008), p. 269. ISSN: 1471-2164. DOI: 10.1186/1471-2164-9-269. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2590703/> (visited on 03/03/2024).
- [144] Chunguang Li, Maoling Zhu, Xiaoli Lou, et al. “Transcriptional Factor OCT4 Promotes Esophageal Cancer Metastasis by Inducing Epithelial-Mesenchymal Transition through VEGF-C/VEGFR-3 Signaling Pathway”. In: *Oncotarget* 8.42 (Sept. 22, 2017), pp. 71933–71945. ISSN: 1949-2553. DOI: 10.18632/oncotarget.18035. URL: <https://www.oncotarget.com/lookup/doi/10.18632/oncotarget.18035> (visited on 03/03/2024).
- [145] Séverine Duvaud, Chiara Gabella, Frédérique Lisacek, et al. “Expasy, the Swiss Bioinformatics Resource Portal, as Designed by Its Users”. In: *Nucleic Acids Research* 49.W1 (July 2, 2021), W216–W227. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkab225. URL: <https://academic.oup.com/nar/article/49/W1/W216/6225225> (visited on 03/03/2024).
- [146] National Center for Biotechnology Information (US). *TFAP2C Transcription Factor AP-2 Gamma [Homo Sapiens (Human)]*. Nov. 23, 2023. URL: <https://www.ncbi.nlm.nih.gov/gene/7022#summary>.
- [147] Maribelis Ruiz, Curtis Pettaway, Renduo Song, et al. “Activator Protein 2–Inhibits Tumorigenicity and Represses Vascular Endothelial Growth Factor Transcription in Prostate Cancer Cells”. In: () .
- [148] Jun Hou, Joachim Aerts, Bianca Den Hamer, et al. “Gene Expression-Based Classification of Non-Small Cell Lung Carcinomas and Survival Prediction”. In: *PLoS ONE* 5.4 (Apr. 22, 2010). Ed. by William C. S. Cho, e10312. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0010312. URL: <https://dx.plos.org/10.1371/journal.pone.0010312> (visited on 03/04/2024).
- [149] Li Ming Zhu, Dong Mei Shi, Qiang Dai, et al. “Tumor Suppressor XAF1 Induces Apoptosis, Inhibits Angiogenesis and Inhibits Tumor Growth in Hepatocellular Carcinoma”. In: *Oncotarget* 5.14 (July 30, 2014), pp. 5403–5415. ISSN: 1949-2553. DOI: 10.18632/oncotarget.2114. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4118700/> (visited on 03/04/2024).

- oncotarget.com/lookup/doi/10.18632/oncotarget.2114 (visited on 03/04/2024).
- [150] Ping Hu Sun, Li Ming Zhu, Min Min Qiao, et al. “The XAF1 Tumor Suppressor Induces Autophagic Cell Death via Upregulation of Beclin-1 and Inhibition of Akt Pathway”. In: *Cancer Letters* (July 2011), S0304383511003946. ISSN: 03043835. DOI: 10.1016/j.canlet.2011.06.037. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0304383511003946> (visited on 03/04/2024).
- [151] George W. Woodfield, Yizhen Chen, Thomas B. Bair, et al. “Identification of Primary Gene Targets of TFAP2C in Hormone Responsive Breast Carcinoma Cells”. In: *Genes, Chromosomes and Cancer* 49.10 (Oct. 2010), pp. 948–962. ISSN: 1045-2257, 1098-2264. DOI: 10.1002/gcc.20807. URL: <https://onlinelibrary.wiley.com/doi/10.1002/gcc.20807> (visited on 03/04/2024).
- [152] National Center for Biotechnology Information (US). *DEDD2 Death Effector Domain Containing 2 [Homo Sapiens (Human)]*. May 3, 2024. URL: <https://www.ncbi.nlm.nih.gov/gene/162989#summary>.
- [153] Justin Kale, Elizabeth J Osterlund, and David W Andrews. “BCL-2 Family Proteins: Changing Partners in the Dance towards Death”. In: *Cell Death & Differentiation* 25.1 (Jan. 2018), pp. 65–80. ISSN: 1350-9047, 1476-5403. DOI: 10.1038/cdd.2017.186. URL: <https://www.nature.com/articles/cdd2017186> (visited on 04/26/2024).
- [154] Okot Nyormoi, Zhi Wang, Dao Doan, et al. “Transcription Factor AP-2 α Is Preferentially Cleaved by Caspase 6 and Degraded by Proteasome during Tumor Necrosis Factor Alpha-Induced Apoptosis in Breast Cancer Cells”. In: *Molecular and Cellular Biology* 21.15 (Aug. 1, 2001), pp. 4856–4867. ISSN: 1098-5549. DOI: 10.1128/MCB.21.15.4856-4867.2001. URL: <https://www.tandfonline.com/doi/full/10.1128/MCB.21.15.4856-4867.2001> (visited on 03/04/2024).
- [155] Vira Bitko, Olena Shulyayeva, Barsanjit Mazumder, et al. “Nonstructural Proteins of Respiratory Syncytial Virus Suppress Premature Apoptosis by an NF- κ B-Dependent, Interferon-Independent Mechanism and Facilitate Virus Growth”. In: *Journal of Virology* 81.4 (Feb. 15, 2007), pp. 1786–1795. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.01420-06. URL: <https://journals.asm.org/doi/10.1128/JVI.01420-06> (visited on 12/06/2023).

- [156] National Center for Biotechnology Information (US). *STAT2 Signal Transducer and Activator of Transcription 2 [Homo Sapiens (Human)]*. Jan. 3, 2024. URL: <https://www.ncbi.nlm.nih.gov/gene/6773#summary>.
- [157] Jillian N. Whelan, Kim C. Tran, Damian B. Van Rossum, et al. “Identification of Respiratory Syncytial Virus Nonstructural Protein 2 Residues Essential for Exploitation of the Host Ubiquitin System and Inhibition of Innate Immune Responses”. In: *Journal of Virology* 90.14 (July 15, 2016). Ed. by D. S. Lyles, pp. 6453–6463. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.00423-16. URL: <https://journals.asm.org/doi/10.1128/JVI.00423-16> (visited on 03/04/2024).
- [158] Jing Nan, Yuxin Wang, Jinbo Yang, et al. “IRF9 and Unphosphorylated STAT2 Cooperate with NF- κ B to Drive IL6 Expression”. In: *Proceedings of the National Academy of Sciences* 115.15 (Apr. 10, 2018), pp. 3906–3911. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1714102115. URL: <https://pnas.org/doi/full/10.1073/pnas.1714102115> (visited on 03/04/2024).
- [159] Yu Qiao, Eugenia G. Giannopoulou, Chun Hin Chan, et al. “Synergistic Activation of Inflammatory Cytokine Genes by Interferon- γ -Induced Chromatin Remodeling and Toll-like Receptor Signaling”. In: *Immunity* 39.3 (Sept. 2013), pp. 454–469. ISSN: 10747613. DOI: 10.1016/j.immuni.2013.08.009. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1074761313003373> (visited on 03/04/2024).
- [160] Barbara Testoni, Christine Völlenkle, Francesca Guerrieri, et al. “Chromatin Dynamics of Gene Activation and Repression in Response to Interferon α (IFN α) Reveal New Roles for Phosphorylated and Unphosphorylated Forms of the Transcription Factor STAT2”. In: *Journal of Biological Chemistry* 286.23 (June 2011), pp. 20217–20227. ISSN: 00219258. DOI: 10.1074/jbc.M111.231068. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0021925819490881> (visited on 03/04/2024).
- [161] National Center for Biotechnology Information (US). *TNFSF10 TNF Superfamily Member 10 [Homo Sapiens (Human)]*. July 4, 2024. URL: <https://www.ncbi.nlm.nih.gov/gene/8743#summary>.
- [162] Huiying Li, Lei Xing, Nan Zhao, et al. “Furosine Induced Apoptosis by the Regulation of STAT1/STAT2 and UBA7/UBE2L6 Genes in HepG2 Cells”. In: *International Journal of Molecular Sciences* 19.6 (May 31, 2018), p. 1629. ISSN: 1422-0067. DOI: 10.3390/ijms19061629. URL: <http://www.mdpi.com/1422-0067/19/6/1629> (visited on 03/04/2024).

- [163] National Center for Biotechnology Information (US). *CASP8 Caspase 8 [Homo Sapiens (Human)]*. Jan. 3, 2024. URL: <https://www.ncbi.nlm.nih.gov/gene/841>.
- [164] Joanne Elliott, Oonagh T. Lynch, Yvonne Suessmuth, et al. “Respiratory Syncytial Virus NS1 Protein Degrades STAT2 by Using the Elongin-Cullin E3 Ligase”. In: *Journal of Virology* 81.7 (Apr. 2007), pp. 3428–3436. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.02303-06. URL: <https://journals.asm.org/doi/10.1128/JVI.02303-06> (visited on 03/04/2024).
- [165] National Center for Biotechnology Information (US). *ATF2 Activating Transcription Factor 2 [Homo Sapiens (Human)]*. Jan. 3, 2024. URL: <https://www.ncbi.nlm.nih.gov/gene/1386#summary>.
- [166] James H. Steer, Karen M. Kroeger, Lawrence J. Abraham, et al. “Glucocorticoids Suppress Tumor Necrosis Factor- α Expression by Human Monocytic THP-1 Cells by Suppressing Transactivation through Adjacent NF- κ B and c-Jun-Activating Transcription Factor-2 Binding Sites in the Promoter”. In: *Journal of Biological Chemistry* 275.24 (June 2000), pp. 18432–18440. ISSN: 00219258. DOI: 10.1074/jbc.M906304199. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0021925819832039> (visited on 03/04/2024).
- [167] Tessa Van Royen, Iebe Rossey, Koen Sedeyn, et al. “How RSV Proteins Join Forces to Overcome the Host Innate Immune Response”. In: *Viruses* 14.2 (Feb. 17, 2022), p. 419. ISSN: 1999-4915. DOI: 10.3390/v14020419. URL: <https://www.mdpi.com/1999-4915/14/2/419> (visited on 03/04/2024).
- [168] National Center for Biotechnology Information (US). *SIX5 SIX Homeobox 5 [Homo Sapiens (Human)]*. Jan. 3, 2024. URL: <https://www.ncbi.nlm.nih.gov/gene/147912#summary>.
- [169] S. E. Harris, C. L. Winchester, and K. J. Johnson. “Functional Analysis of the Homeodomain Protein SIX5”. In: *Nucleic Acids Research* 28.9 (May 1, 2000), pp. 1871–1878. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/28.9.1871. URL: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/28.9.1871> (visited on 03/04/2024).
- [170] Junhong Huang, Wujian Zheng, Ping Zhang, et al. “ChIPBase v3.0: The Encyclopedia of Transcriptional Regulations of Non-Coding RNAs and Protein-Coding Genes”. In: *Nucleic Acids Research* 51.D1 (Jan. 6, 2023), pp. D46–D56. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkac1067. URL: [https:](https://)

BIBLIOGRAPHY

- //academic.oup.com/nar/article/51/D1/D46/6833235 (visited on 03/04/2024).
- [171] Alexandra B Keenan, Denis Torre, Alexander Lachmann, et al. “ChEA3: Transcription Factor Enrichment Analysis by Orthogonal Omics Integration”. In: *Nucleic Acids Research* 47.W1 (July 2, 2019), W212–W224. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkz446. URL: <https://academic.oup.com/nar/article/47/W1/W212/5494769> (visited on 03/04/2024).

6. Supplemental Material

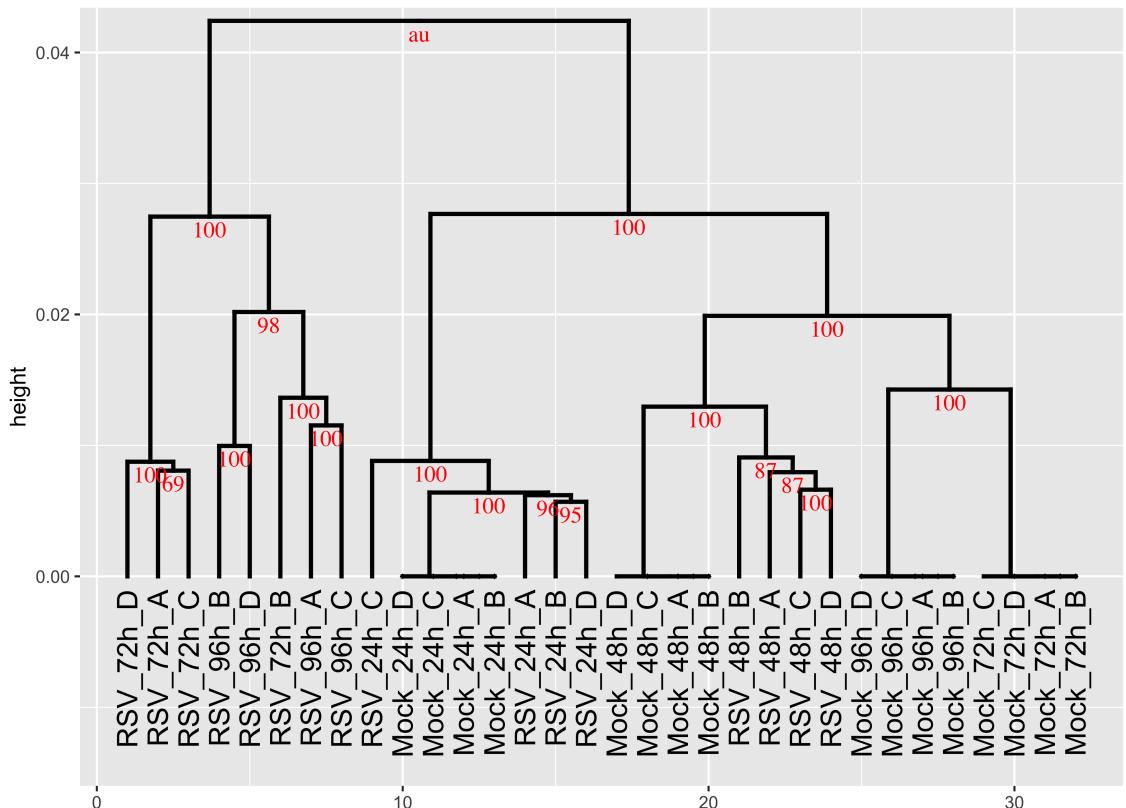


Figure S1: shows the full-size hierarchical clustering of the A549 data set. The two groups early/non-infected and late split directly at the first branch.

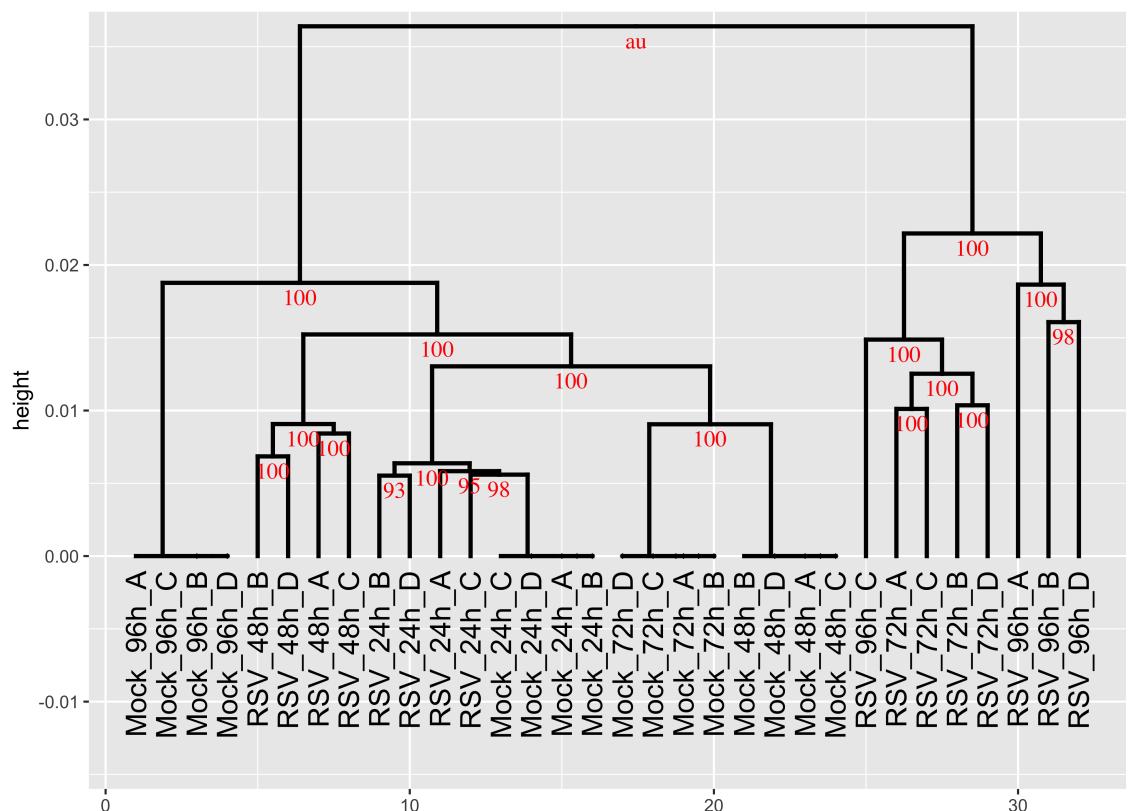


Figure S2: shows the full-size hierarchical clustering of the HEp-2 data set. The two groups early/non-infected and late split directly at the first branch.

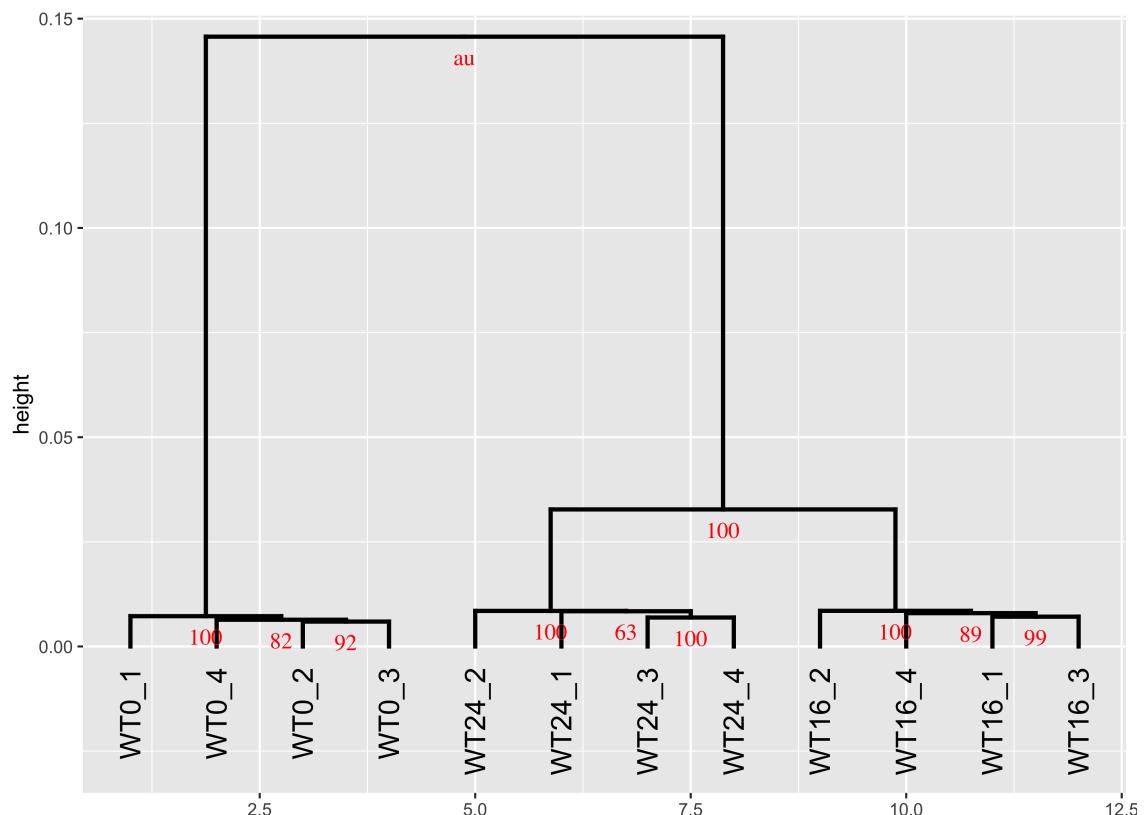


Figure S3: shows the full-size hierarchical clustering of the hSAEC data set. The 3 time points split nicely with the first time point being also the first to branch.

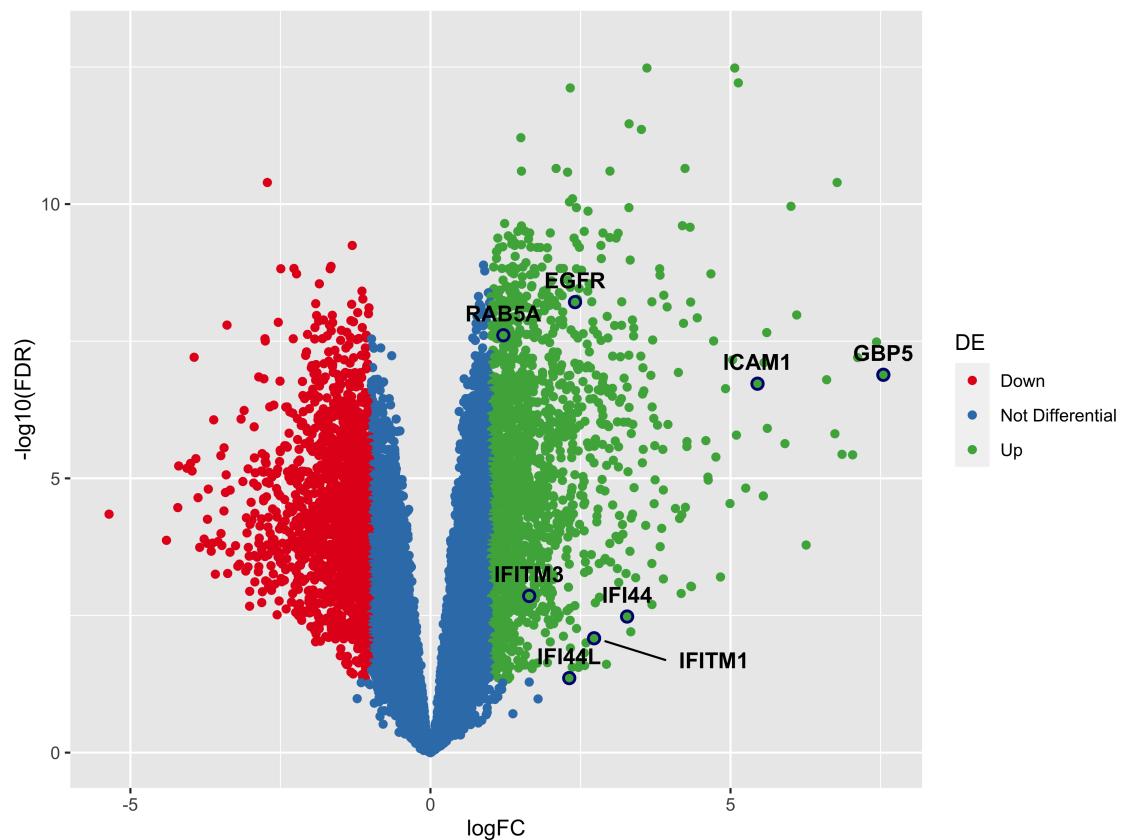


Figure S4: Volcano plot of the A549 data set infected vs control. Red dots show differentially downregulated genes, green show differentially upregulated genes. Explicitly depicted are some upregulated HRFs. Others are not differentially expressed (blue).

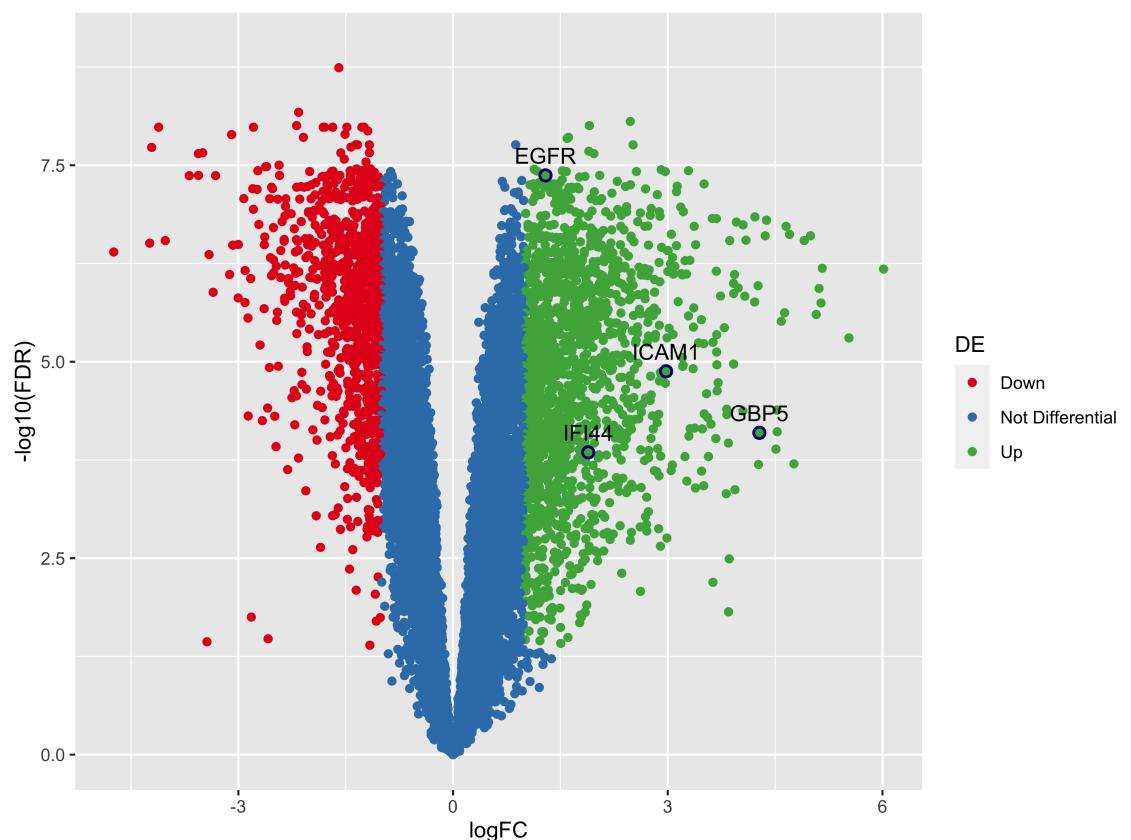


Figure S5: Volcano plot of the HEp-2 data set infected vs control. Red dots show differentially downregulated genes, green show differentially upregulated genes. Some HRFs are upregulated. Others are not differentially expressed (blue).

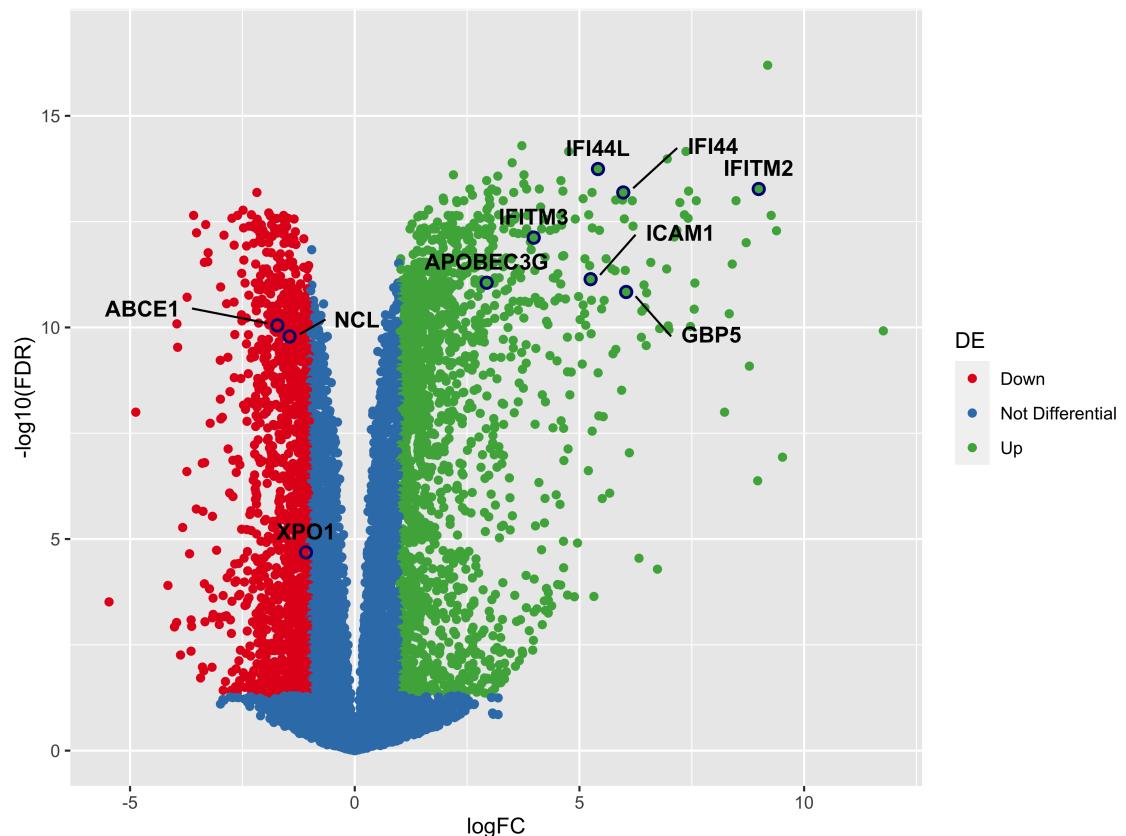


Figure S6: Volcano plot of the hSAEC data set 16 hours post infection vs control (0 hours). Red dots show the differentially downregulated genes, green show differentially upregulated genes. The most host factors are upregulated HRFs, while three HDFs are downregulated. Others are not differentially expressed (blue).

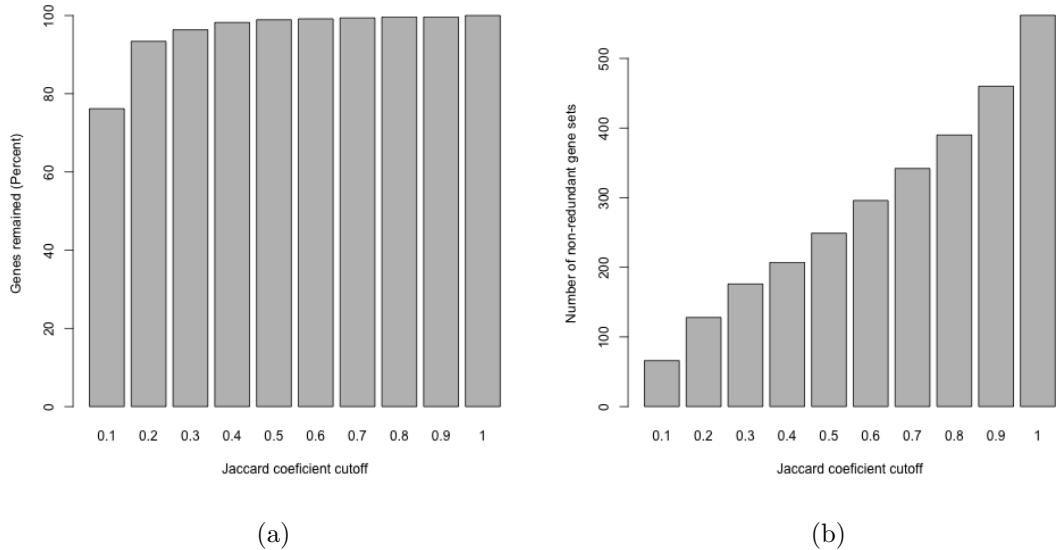


Figure S7: (a) Shows the percentage of remaining genes in the A549 upregulated gene sets for a reduction with different Jaccard coefficients. (b) Shows the number of remaining gene sets for a reduction with different Jaccard coefficients.

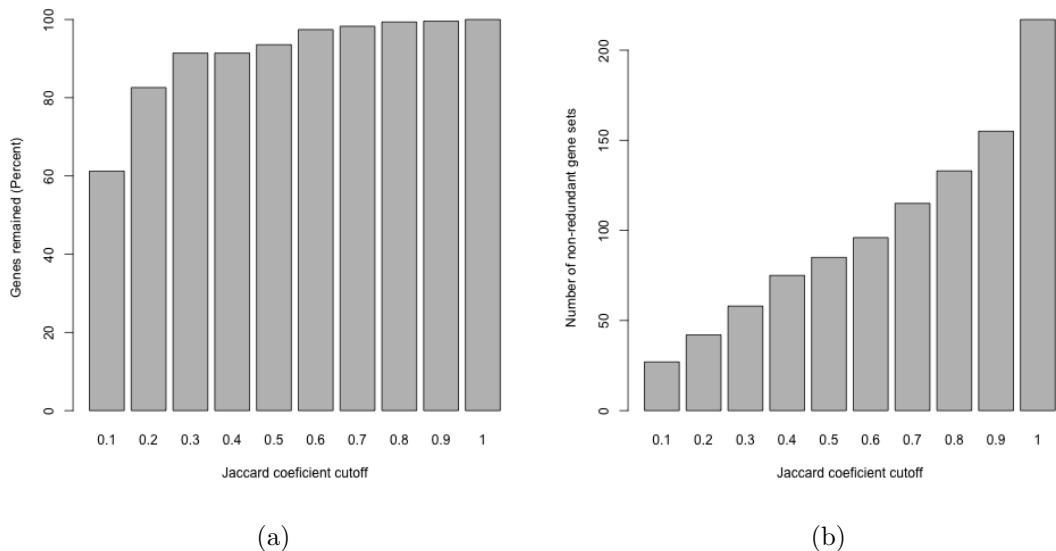


Figure S8: (a) Shows the percentage of remaining genes in the A549 downregulated gene sets for a reduction with different Jaccard coefficients. (b) Shows the number of remaining gene sets for a reduction with different Jaccard coefficients.

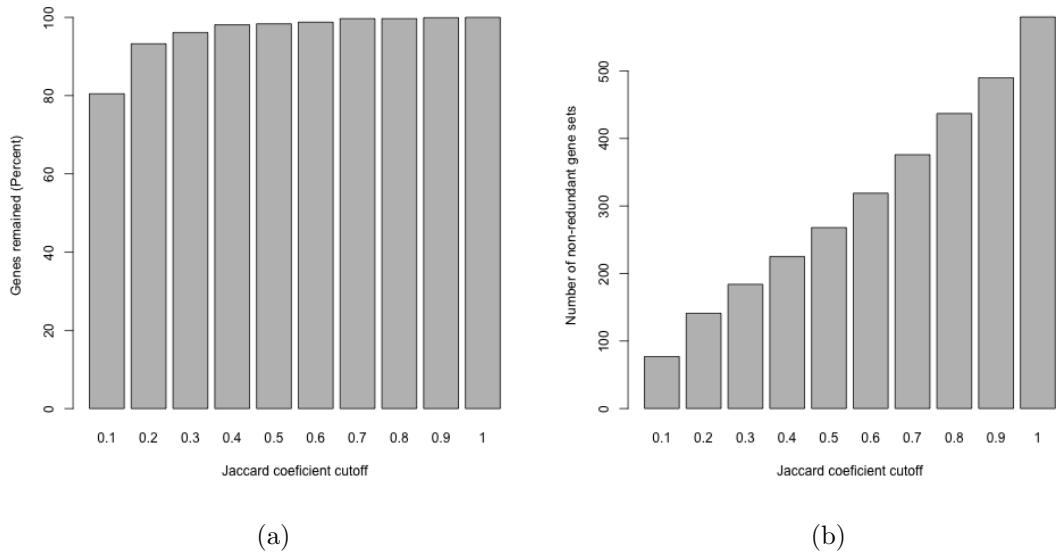


Figure S9: (a) Shows the percentage of remaining genes in the HEp-2 upregulated gene sets for a reduction with different Jaccard coefficients. (b) Shows the number of remaining gene sets for a reduction with different Jaccard coefficients.

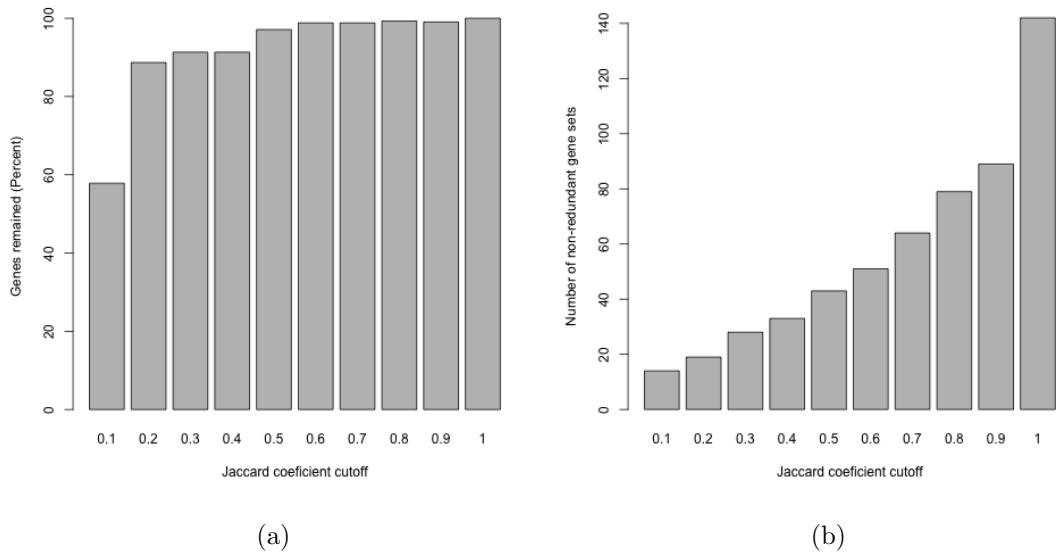


Figure S10: (a) Shows the percentage of remaining genes in the HEp-2 downregulated gene sets for a reduction with different Jaccard coefficients. (b) Shows the number of remaining gene sets for a reduction with different Jaccard coefficients.

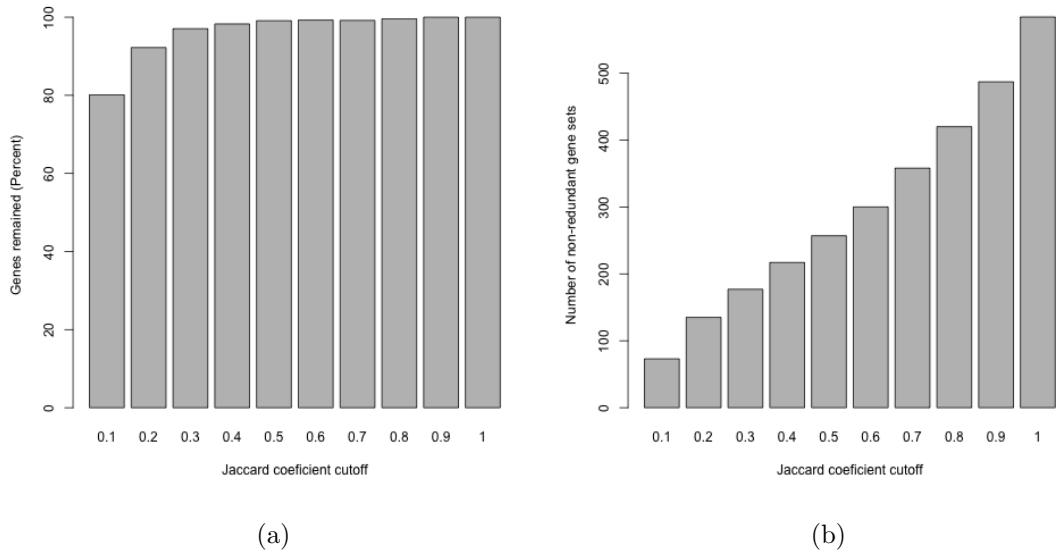


Figure S11: (a) Shows the percentage of remaining genes in the hSAEC early up-regulated gene sets for a reduction with different Jaccard coefficients. (b) Shows the number of remaining gene sets for a reduction with different Jaccard coefficients.

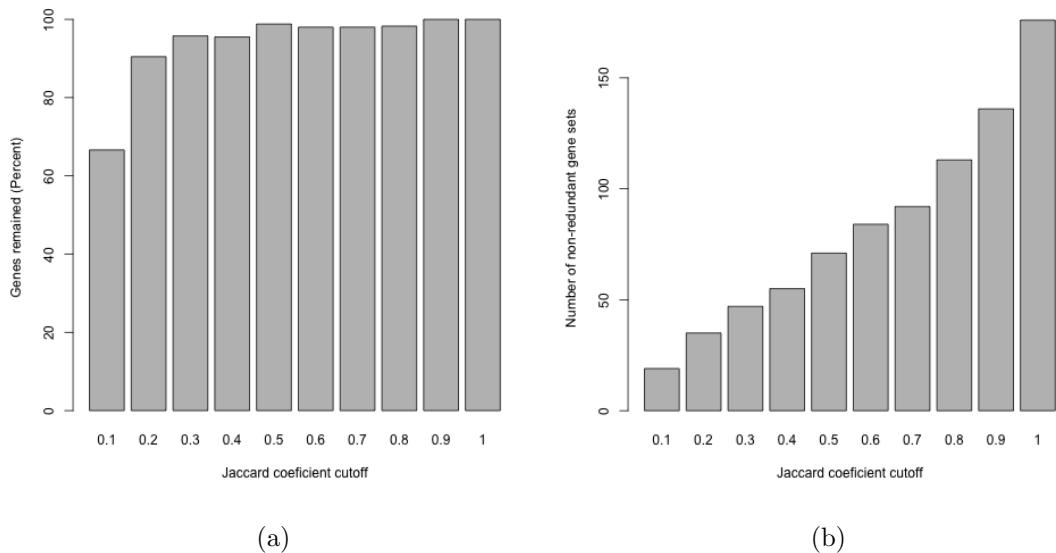


Figure S12: (a) Shows the percentage of remaining genes in the hSAEC early down-regulated gene sets for a reduction with different Jaccard coefficients. (b) Shows the number of remaining gene sets for a reduction with different Jaccard coefficients.

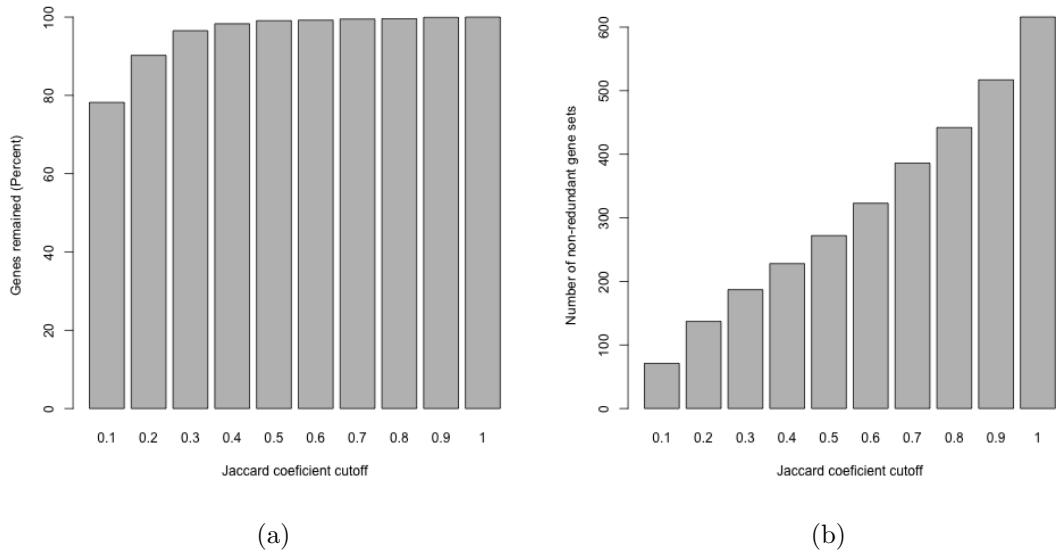


Figure S13: (a) Shows the percentage of remaining genes in the hSAEC late upregulated gene sets for a reduction with different Jaccard coefficients. (b) Shows the number of remaining gene sets for a reduction with different Jaccard coefficients.

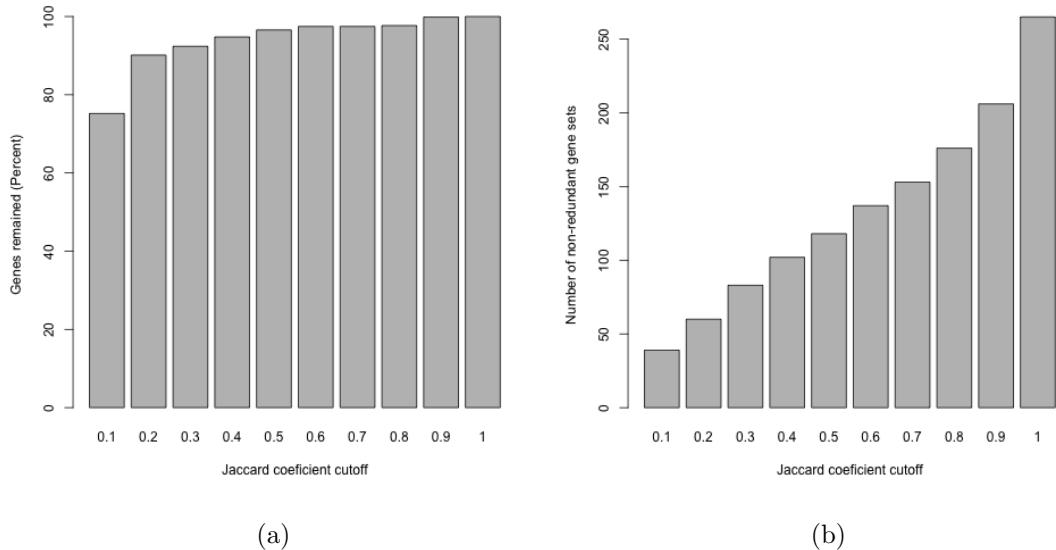


Figure S14: (a) Shows the percentage of remaining genes in the hSAEC late downregulated gene sets for a reduction with different Jaccard coefficients. (b) Shows the number of remaining gene sets for a reduction with different Jaccard coefficients.



Eigenständigkeitserklärung

1. Hiermit versichere ich, dass ich die vorliegende Arbeit - bei einer Gruppenarbeit die von mir zu verantwortenden und entsprechend gekennzeichneten Teile - selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.
Ich trage die Verantwortung für die Qualität des Textes sowie die Auswahl aller Inhalte und habe sichergestellt, dass Informationen und Argumente mit geeigneten wissenschaftlichen Quellen belegt bzw. gestützt werden. Die aus fremden oder auch eigenen, älteren Quellen wörtlich oder sinngemäß übernommenen Textstellen, Gedankengänge, Konzepte, Grafiken etc. in meinen Ausführungen habe ich als solche eindeutig gekennzeichnet und mit vollständigen Verweisen auf die jeweilige Quelle versehen. Alle weiteren Inhalte dieser Arbeit ohne entsprechende Verweise stammen im urheberrechtlichen Sinn von mir.
2. Ich weiß, dass meine Eigenständigkeitserklärung sich auch auf nicht zitierfähige, generierende KI-Anwendungen (nachfolgend „generierende KI“) bezieht.
Mir ist bewusst, dass die Verwendung von generierender KI unzulässig ist, sofern nicht deren Nutzung von der prüfenden Person ausdrücklich freigegeben wurde (Freigabeerklärung). Sofern eine Zulassung als Hilfsmittel erfolgt ist, versichere ich, dass ich mich generierender KI lediglich als Hilfsmittel bedient habe und in der vorliegenden Arbeit mein gestalterischer Einfluss deutlich überwiegt. Ich verantworte die Übernahme der von mir verwendeten maschinell generierten Passagen in meiner Arbeit vollumfänglich selbst.
Für den Fall der Freigabe der Verwendung von generierender KI für die Erstellung der vorliegenden Arbeit wird eine Verwendung in einem gesonderten Anhang meiner Arbeit kenntlich gemacht.
Dieser Anhang enthält eine Angabe oder eine detaillierte Dokumentation über die Verwendung generierender KI gemäß den Vorgaben in der Freigabeerklärung der prüfenden Person.
Die Details zum Gebrauch generierender KI bei der Erstellung der vorliegenden Arbeit inklusive Art, Ziel und Umfang der Verwendung sowie die Art der Nachweispflicht habe ich der Freigabeerklärung der prüfenden Person entnommen.
3. Ich versichere des Weiteren, dass die vorliegende Arbeit bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt wurde oder in deutscher oder einer anderen Sprache als Veröffentlichung erschienen ist.
4. Mir ist bekannt, dass ein Verstoß gegen die vorbenannten Punkte prüfungsrechtliche Konsequenzen haben und insbesondere dazu führen kann, dass meine Prüfungsleistung als Täuschung und damit als mit „nicht bestanden“ bewertet werden kann. Bei mehrfachem oder schwerwiegendem Täuschungsversuch kann ich befristet oder sogar dauerhaft von der Erbringung weiterer Prüfungsleistungen in meinem Studiengang ausgeschlossen werden.

Ort und Datum

Unterschrift